Lecture Notes in Statistics 207
Proceedings

Emilio Porcu
José María Montero
Martin Schlather *Editors*

Advances and Challenges in Space-time Modelling of Natural Events

Springer

Lecture Notes in Statistics 207

For further volumes:
http://www.springer.com/series/694

Emilio Porcu • José María Montero
Martin Schlather
Editors

Advances and Challenges in Space-time Modelling of Natural Events

Springer

Editors
Emilio Porcu
Universität Göttingen
Institut for Mathematics Stochastics
Goldschmidtstrasse 7
37077 Göttingen
Germany
eporcu@uni-goettingen.de

Universidad de Castilla-La Mancha
Facultad de Derecho y Ciencias Sociales
Área de Estadística
Avenida Camilo Jose Cela 14
13005 Ciudad Real
España
emilio.porcu@uclm.es

José-María Montero
Universidad de Castilla-La Mancha
Facultad de Ciencias Jurídicas y Sociales
Área de Estadística
Cobertizo San Pedro Mártir s/n
45071 Toledo
España
jose.mlorenzo@uclm.es

Martin Schlather
Universität Göttingen
Institut for Mathematics Stochastics
Goldschmidtstrasse 7
37077 Göttingen
Germany
schlather@math.uni-goettingen.de

ISSN 0930-0325
ISBN 978-3-642-17085-0 e-ISBN 978-3-642-17086-7
DOI 10.1007/978-3-642-17086-7
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011943342

Ispinta da su entuésa morte rutta estśalvure
manna chi tenia, śalvure antiga de sa domo
mia, chi daiat car umbra in custa corte.
Ancora fit deretta bella e forte e de sabios
fruttos provvidia.
A tiu meu Michelli

Emilio Porcu

To Gema, Beatriz and José,
my daily support.

José-María Montero

To Monika, Achim, Sonja and Corinna.

Martin Schlather

Preface

These lecture notes arise as the continuation of the International Spring School *Advances and Challenges in Space-time modelling of Natural Events*, which took place in Toledo (Spain) in March 2010. This Spring School was addressed to young researchers (Master students, PhD students, PostDoc researchers) in academics, extrauniversitary research and industry, interested in learning about recent developments, new methods and applications in spatial and spatio-temporal statistics and related areas and to exchange their ideas and results with colleagues. At the end there were around 50 students coming from all the continents; such a success was guaranteed by a good mixture between the fascinating Toledo and the excellent lecturers being there.

There were several motivations justifying such a Spring School. Recent literature emphasize the need for comprehensive mathematical and statistical frameworks for the description of phenomena evolving over space or time or both of them.

Once established the crucial importance of simultaneously studying the spatial and temporal components in the evolution of an environmental process, it is worth mentioning that the approach to such problem can be extremely variable, depending on the researcher point of view and the discipline he comes from. A very important dichotomy regards.

The School covered the main branches of spatial and space–time statistics: Geostatistics, non-Gaussian random fields, Markov random fields, space–time point processes, large space–time dataset, spatial design, and last but not least, extreme values theory for spatial processes. Such a huge range of subjects attracted the interest of students and young researchers and we hope they appreciated the result of this organization and the time they spent in Toledo.

The Editors of these Lecture Notes are extremely grateful to the lectures for their excellent work at the Spring School, and to the students for the participation in the course and for the enthusiasm they gave to such event.

Göttingen — *Emilio Porcu & Martin Schlather*
Toledo — *José-María Montero*

Acknowledgements

The Editors are grateful to:

- The Consejería de Ordenación del Territorio y Vivienda de la Junta de Comunidades de Castilla-La Mancha, and especially to Julián Sanchez-Pingarrón and Barbara Pons, for the financial support.
- The University of Castilla-La Mancha, for logistics and financial support.
- The University of Göttingen for the important support during the preparation of the Spring School.
- Gema Fernández-Avilés, for the effort she made coordinating the Spring School and especially for taking care of everybody and everything. And for her valuable work in the preparation of these Lecture Notes.
- Jaume García Villar, President of the Spanish Statistics Institute, for both his contribution to the opening session and scientific support.
- Peter Diggle, Abdel El Shaarawi and Dietrich Stoyan for joining José-María Montero as members of the Scientific Committee.
- The Lecturers and the Contributors to these lecture notes.
- The Referees for their thorough revisions allowing for considerably improved versions, in particular Mariano Amo Salas, Jesús López Fidalgo, Michael Scheuerer, Maria Dolores Ruiz Medina, Zachar Kabluchko, Simone Padoan, Moreno Bevilacqua, Gema Fernández-Avilés, Helga Wagner, Alessandro Zini, and Antti Pentinen.

Contents

Contributors

Denis Allard Biostatistics and Spatial Processes (BioSP), INRA, Avignon, France, allard@avignon.inra.fr

Jean-Noël Bacro I3M, Université Montpellier II, 34095 Montpellier, Cedex 5, France, bacro@math.univ-montp2.fr

Gema Fernández-Avilés Facultad de Ciencias Jurídicas y Sociales, Área de Estadística, University of Castilla-La Mancha, Toledo, Spain, gema.faviles@uclm.es

Carlo Gaetan Dipartimento di Scienze Ambientali, Informatica e Statistica, Università Ca' Foscari-Venezia, I-30121 Venezia, San Giobbe, Cannaregio 873, gaetan@unive.it

Marc G. Genton Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA, genton@stat.tamu.edu

Peter Guttorp University of Washington, Seattle & Norwegian Computing Centre, Oslo, Norway, peter@stat.washington.edu

Nikolai N. Leonenko Cardiff School of Mathematics, Cardiff University, UK, LeonenkoN@cardiff.ac.uk

Bo Li Department of Statistics, Purdue University, West Lafayette, IN 47907, USA, boli@stat.purdue.edu

José-María Montero Facultad de Ciencias Jurídicas y Sociales, Área de Estadística, University of Castilla-La Mancha, Toledo, Spain, jose.mlorenzo@uclm.es

Werner G. Müller Department of Applied Statistics, Johannes-Kepler-University Linz, Altenberger Strase 69, A-4040 Linz, Austria, werner.mueller@jku.at

Emilio Porcu Universität Göttingen, Institut for Mathematics Stochastics, Goldschmidtstrasse 7, 37077 Gottingen, Germany, eporcu@uni-goettingen.de

Universidad de Castilla-La Mancha, Facultad de Derecho y Ciencias Sociales, Área de Estadística, Avenida Camilo Jose Cela 14, 13005 Ciudad Real, España, emilio.porcu@uclm.es

Luc Pronzato Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis, Les Algorithmes, 2000 route des lucioles, BP 121, 06903 Sophia Antipolis cedex, France, pronzato@i3s.unice.fr

M. Dolores Ruiz-Medina Faculty of Sciences, University of Granada, Campus Fuente Nueva s/n, 18071, Granada, mruiz@ugr.es

Martin Schlather Universität Göttingen, Institut for Mathematics Stochastics, Goldschmidtstrasse 7, 37077 Göttingen, Germany

Centre for Statistics, Georg-August-Universität Göttingen, Goldschmidtstr. 7, D – 37077 Gottingen, Germany, schlather@math.uni-goettingen.de

Michael L. Stein Department of Statistics, The University of Chicago, Chicago, USA, stein@galton.uchicago.edu

Ying Sun Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA, sunwards@stat.tamu.edu

Thordis Linda Thorarisdottir Department of Applied Mathematics, Heidelberg University, Germany, thordis@uni-heidelberg.de

Helmut Waldl Department of Applied Statistics, Johannes-Kepler-University Linz, Altenberger Strase 69, A-4040 Linz, Austria, helmut.waldl@jku.at

Hao Zhang Department of Statistics, Purdue University, USA, zhanghao@purdue.edu

Chapter 1
Space-Time Processes and Geostatistics

Gema Fernández-Avilés, José-María Montero, Emilio Porcu, and Martin Schlather

Abstract This chapter presents a broad view of space and space–time processes. Our dissertation starts with the concept of space and time from the philosophical viewpoint. Then, we relate the concept of space–time with the current practice in Geostatistics and the use of the latter as an effective framework for natural and social sciences. The rest of the chapter gives a methodological assessment of space–time geostatistics through the framework of space–time random functions, covariances and variograms.

1.1 Space, Time and Spacetime

1.1.1 A Brief Sketch of the Philosophy of Space and Time

Space and time are fundamental notions. They are so basic concepts that they were regarded as the source of the world in ancient mythological, religious, and

J.-M. Montero (✉) · G. Fernández-Avilés
Facultad de Ciencias Jurídicas y Sociales, Área de Estadística, University of Castilla-La Mancha, Toledo, Spain
e-mail: jose.mlorenzo@uclm.es; gema.faviles@uclm.es

E. Porcu
Universität Göttingen, Institut for Mathematics Stochastics, Goldschmidtstrasse 7, 37077 Göttingen, Germany
e-mail: eporcu@uni-goettingen.de

Universidad de Castilla-La Mancha, Facultad de Derecho y Ciencias Sociales, Área de Estadística, Avenida Camilo Jose Cela 14, 13005 Ciudad Real, España
e-mail: emilio.porcu@uclm.es

M. Schlather
Universität Göttingen, Institut for Mathematics Stochastics, Goldschmidtstrasse 7, 37077 Göttingen, Germany
e-mail: schlather@math.uni-goettingen.de

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_1,

philosophical systems, including *Chaos* and *Kronos* in ancient Greek mythology, *akasa* and *kala* in Indian philosophy, and *Zurvan* in early Zoroastrianism [2].

What is space? What is time? Do they exist independently of the things and processes in them? Or is their existence parasitic on these things and processes? Are they like a canvas onto which an artist paints; they exist whether or not the artist paints on them? Or are they akin to parenthood; there is no parenthood until there are parents and children? That is, is there no space and time until there are things with spatial properties and processes with temporal durations? These questions have been debated for a long time and will continue to be debated [62].

Following [68], historical views of space and time can be categorized into both continuous and discrete, and absolute or relative. In the continuous perspective the objects are contained within space and time, and as a consequence space and time are the subject matter. In contrast, in the discrete view the subject matters are the objects. The absolute perspective assumes an immutable rigid and purely geometric structure while the relative view, a subjective view, assumes a flexible structure that is more topological in nature. The relative view, defined by means of intervals between objects or locations, is bounded, while the absolute perspective is unbounded.

Therefore, focusing on the structural aspects, the following are core questions. As for space: Is space finite or infinite in extension? How many dimensions does it have? Is it Euclidean? Is it isotropic? Is it continuous or discrete? As for time: Is time finite or infinite? Does it have a beginning or an end? Is it one-dimensional? Is it linear or branching? Is it anisotropic, *i.e.* directed? Is it continuous or discrete?

To give an answer to the above questions, it is clear that the story ends with Einstein's ideas on a relative and continuous view, clearly influenced by what we now call Minkowski spacetime. However, Newton's ideas about the absolutism of space and time dominated Science until the beginning of the past century and constitute the framework of most of applications of Geostatistics nowadays. But, as stated in [40] and the references therein, Newton was responding to both Galileo and especially Descartes. But Galileo and Descartes themselves were writing in the context of late Aristotelianism, and so were trained in and critical of that rich school of thought, so if we want to fully understand their work we would need to understand scholastic views on space and motion. But late scholasticism itself is the result of a long history tracing from Plato and Aristotle through Jewish, Arabic, Islamic and European thought. And of course Plato and Aristotle are explicitly reacting to their predecessors and contemporaries. In other words, paraphrasing [40] (Ch.1 p.1) we could start the story as early in recorded thought as we like. However, to keep this sketch to manageable proportions, we will start with Greek atomists.

Greek atomists were the first that described the discrete space by reducing everything to distinct bodies adrift in space (the container of such objects: the Void). As stated in [68], ancient Greek philosophy is rooted in mythological and religious notions. The earlier notions in Western mythology on the nature of space and time can be viewed as a progression of the World from the Chaos (the boundless abysm, the infinite space) to Cosmos (the final state of order, which consists of both natural and political components). Although the final state is the state of order, the world is not viewed as a unified whole. Rather, there is a relative order within a multiplicity

of connected pieces, or territories, and discrete events. That is, space and time are discontinuous, although the story does have an overall forward-moving evolution [68]. For early Greeks time on an everyday scale consisted of the ordered rhythm of human activities [37]. This notion of cyclic time is not idiosyncratic of the Greek culture, but it appears also in the mythology and religion of other cultures (as an example, the Mayans were convinced that history would repeat every *lamat* or 260 year-period), and was the dominating though until the emergence of Christianity with only some few exceptions. The most notable exception is the Hebrews and their linear conception of time according to the eventual deliverance and salvation of Israel. The Zoroastrians and Seneca thought that time was progressive and non repeating.

In Homer's Odyssey it can be perceived an ordering of events in the sense of continuity of time, which proceeds from the past, through the present and goes to an open-ended future. Space is also seen as continuous and connected.

Although medieval scholasticism was dominated by Aristotle's views (in the Aristotelian tradition, space implies body and time implies motion; see [68] and the references therein for details), in Renaissance times Copernicus, using the notions of the ancient Greeks, constructed a uniform theory of space and time wherein the earth and the heaven operate according the same laws. This theory, against the traditional Christian distinction between Earth and Heaven, can be considered a Copernican revolution because of (1) the scientific way of expressing the notion of a continuous and non ending time, and (2) the idea of studying space and time together became to have a notable degree of consensus. However, he continued to believe in the Aristotelian idea that space is finite [2].

Galileo, strongly interested in the practical side of Science, recognized empty space. Descartes did not. In fact, Galileo claimed that all bodies fall with the same speed in empty space, and that their fall could be described by simple mathematical laws. In a note apart, as fall in empty space could be not observed accurately at that time, Galileo suggested new experiments. His new method did not aim at the description of what is visible, but rather at the design of experiments and the production of phenomena that one does not normally see and at their calculation on the basis of mathematical theory [38].

Descartes' universe was a mechanical ("wind-up") clockwork robot universe, with energy only as the property of matter being in motion and nothing other than God and human souls being non-material. His material universe was all matter with no empty space and with no separate energy besides the kinetic energy involved in body motion. His "no empty space" was in line with Aristotle and Huygens but opposed the experimental evidence offered by Gilbert, Newton and others who supported non-corporeal energies or "spirits" also existing — separately from matter and being also detectable by experimental science. For Descartes the human body and the human mind were discrete entities. The human soul, unlike the mechanical world, was something that could not be broken down.

According to Descartes the chief attribute of any object is that it occupies space and we can only assure their measurable and geometric properties. This close association of space with the existence of objects forced himself to reject empty

space, because he would otherwise have had to admit the existence of non material objects [68].

As outlined above, in his natural philosophy, all bodies are of one of two classes: those that have extension, corporeal substances, or those of thought, ideas. Corporeal substances are not the bodies of everyday experience but rather geometrical objects devoid of colour, texture, and smell, which are secondary attributes dependent upon our experience. It is only extension with its accompanying geometrical qualities that are necessarily a part of substances. Consequently, "Cartesian bodies are just the objects of geometry made real, purely geometrical objects that exist outside of the minds that conceive them" [31], p. 294.

The idea of the coordinate system was developed in 1637 in two writings (although Fermat developed the concept slightly earlier): In part two of his Discourse on Method, in an appendix, Descartes introduced the new idea of specifying the position of a point or object on a surface, using two intersecting axes as measuring guides. In *La Géométrie*, he further explored the above-mentioned concepts. Specifically, he introduced the use of coordinates for describing plane curves, the axes were omitted, and only positive values of the x- and the y-coordinates were considered, since they were defined as distances between points.

As for time, Descartes espoused the traditional notion of two times but, as in the case of space and matter, these were seen through the lens of theology. According to [68], Descartes attributed absolute time to the external world and external objects. As such, this time is measurable, but is also eternal and attributable to God. Relative time is that time experienced by man as a mode of thought (again the distinction between (external) material things and ideas, *i.e.* perceptions). Following [2], Descartes also advanced the notion that time is discrete: Due to God's omnipotence, God directly intervenes with each successive discrete instant and such interference is the cause of both the continuing existence of each individual and the cause for the entire diversity of natural objects.

The controversial on whether space and time are real objects themselves, *i.e.* absolute, or merely orderings upon real objects, *i.e.* relational, began with a debate between Isaac Newton, through his spokesman Samuel Clarke, and Gottfried Leibniz in the famous Leibniz-Clarke correspondence.

According to [68], the Newtonian notion of space arose from Greek atomists view, with space composed of points, time composed of instants, and both existing independently of the bodies that occupy this space-time. In addition, the movement of a body changes the position of that body, but space and time are viewed as a backdrop with an unchangeable structure that is absolute. Thus, the Newtonian view can be understood as both discrete and absolute (objects existing within constant space and constant time). Newton's thesis of the immobility of (absolute) space, against the backdrop of Descartes, clearly means that the parts of space, just as the parts of time, do not change their relation with respect to one another. The parts of space are their own places, and for a place to be moved out of itself is absurd. A more expansive antecedent of this argument occurs in *De Gravitatione*, applied specifically to time: *if yesterday and tomorrow were to interchange their temporal relations with respect to the remainder of time, then yesterday would become today,*

and today yesterday. Thus, Newton held an interestingly holistic identity criterion for the parts of space and time [72].

The ideas above outlined can be summarized as follows:

> *Space is eternal in duration and immutable in nature... Although space may be empty of body, nevertheless it is itself not a void: and something is there, because spaces are there, although nothing more than that.*
>
> (De Gravitatione, as quoted by [21], p. 133).

> *Absolute, true and mathematical time, of itself, and from its own nature, flows equably without relation to anything external.*
>
> (Principia, as quoted by [39], p. 118).

Leibniz disagrees with Newton that space and time are absolute. Consistent with an austere ontology according to which "there is nothing in things except simple substances, and in them perception and appetite", Leibniz denies that space and time are to be included among the worlds' most basic constituents: time and space as such are not things or substances or accidents, but merely beings of reason whose reality is grounded in the mind [30]. For him space and time are properties, which we attribute to objects in order to understand them properly and distinguish them from other objects. Therefore, space and time are relative. There is no physically spatial universe. We use the illusion of time and space in an effort to make sense of our world. As for Kant, the individual is not a passive observer of an already existing reality, but instead determines the "shape" of space and time for himself or herself ([68] in p. 24; see also [44] for details about the Leibniz's controversy with the Newtonians).

Newton's ideas about the absolutism of space and time dominated Science until the beginning of the past century. But Einstein's ideas on a relative and continuous view dominate the Science in the twentieth century. Einstein's work was influenced by Minkowsky fusion of space and time. On September 21, 1908 Hermann Minkowski began his talk at the 80th Assembly of German Natural Scientists and Physicians with the now famous introduction:

> *The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.* [60], p. 75.

The idea that our world and all objects are four-dimensional ($4D$) since he introduced the unification of space and time into an indivisible $4D$ entity (which he called "the World", $4D$ world, or what we now call *Minkowski spacetime*) can be deduced from the following paragraph:

> *A point of space at a point of time, that is, a system of values, x, y, z, t, I will call a world-point. The multiplicity of all thinkable x, y, z, t systems of values we will christen the world... Not to leave a yawning void anywhere, we will imagine that everywhere and everywhen there is something perceptible. To avoid saying*

"matter" or "electricity" I will use for this something the word "substance". We fix our attention on the substantial point which is at the world-point x, y, z, t, and imagine that we are able to recognize this substantial point at any other time. Let the variations dx, dy, dz of the space co-ordinates of this substantial point correspond to a time element dt. Then we obtain, as an image, so to speak, of the everlasting career of the substantial point, a curve in the world, a world-line, the points of which can be referred unequivocally to the parameter t from $-\infty$ *to* $+\infty$*. The whole universe is seen to resolve itself into similar world-lines, and I would fain anticipate myself by saying that in my opinion physical laws might find their most perfect expression as reciprocal relation, between these world-lines* [60], p. 76.

Since then the question of the ontological status of this union of space and time has become the subject of a continued debate, but that goes beyond the scope of this brief sketch.

1.1.2 Peuquet's Varying Views of Reality

As stated in the previous subsection, some core ideas have persisted and have been indeed refined in the history of philosophical and scientific thought on the nature of space and time and its representation. In this subsection we focus on whether space and time is (1) continuous or discrete, and (2) absolute or relative.

As can be deduced from the previous subsection and is resumed in [68]:

1. The continuous view considers that space and time are the subject matter and all objects are contained within space and time. In contrast, in the discrete perspective objects are the subject matter.
2. The absolute perspective assumes an immutable structure that is rigid and purely geometric, whereas the relative view is subjective and assumes a flexible structure that is more topological in nature.
3. Relative space and relative time are defined in terms of relationships between and among locations or in terms of relationships between and among objects, depending on whether both space and time are considered continuous or a discrete, respectively.
4. From the absolute perspective measurements are referred to some constant base, implying nonjudgmental observation. The relative view, involves explicit interpretation of process and the flux of changing pattern and process within specific phenomenological contexts.
5. The relative view is bounded (is defined via intervals between objects or locations). The absolute one is unbounded.

Following [68], the first definition of space contained in the Webster's New Nineteenth Century Dictionary (Second edition), *distance extending without limits in all directions; that which is thought of as boundless, continuous expanse extending in all directions or in the three directions, within all material things are*

contained, and the 19th definition of time, *indefinite, unlimited duration in which things are considered as happening in the past, present, or future; as 'the time of the accident'*, can be clearly interpreted as continuous space and time, respectively, although it must be noticed that continuous space appears as the first definition and continuous time is relegated to the position 19.

However, the second definition of space, *distance, interval, or area between or within things; extent; room; as 'leave a wide space between rooms'*, and the first definition of time, *the period between two events or during which something exists, happens, or acts; measured or a measurable interval*, clearly are in line with both the discrete and relative character of space and time.

As previously said, discrete space, as described by Greek atomists, implies bodies adrift in space, with space acting as a container: the Void. Newtonian notion of time and space have its roots in that notion, with space composed of points, time composed of instants, and both existing independently of the bodies that occupy this space-time. In addition, the movement of a body changes its position but space and time are viewed as a backdrop with a rigid, unchanging structure that is absolute. Thus, the Newtonian view can be considered both discrete and absolute: objects exist within constant space and constant time [68]. Leibniz, as did Kant, adopted a relative notion of space and time. But was the Newtonian perspective which dominated science until Minkowski and Einstein works. The relative view of a combined space-time continues to dominate science in general and physics in particular. There is a common modern presumption of a combined space-time matrix derived from Minkowskian ideas. But this is valid from a relative point of view rather than from an absolute perspective. Absolute space and time are highly interdependent and share many characteristics, but they are not interchangeable in the sense of a four-dimensional, mathematically defined space-time hypercube.

Which is the resolution between absolute and relative points of view? It is one superior to the other? These are two core questions for Peuquet [68]. The fundamental thesis of [68] is that both perspectives are complementary and interdependent. The same could be said for continuous and discrete views of space and time.

Let us focus on Fig. 1.1, which in fact is a space-time domain (something similar can be found in [73]) useful for showing how various views of space-time are related in a general sense. Following again [68], the absolute side assumes an unchanging structure that is independent of human perception (objective). The relative view of space and time is connected with human internality; space and time are contextual and interpretative (depending on social, religious or other contexts and prior individual experience). Thus, according to [68], at one extreme of this continuum, on the side of relative space and time, is the domain of myth and metaphor; On the side of the absolute space and time is the domain of external observation and measurement: what Peuquet names the "external truth".

Figure 1.1 can be also be related with the division between physical and social sciences. Physical sciences would be closest to the absolute side of the absolute-relative continuum due to their emphasis on the understanding of the external reality; somewhere in the middle is Geography, as both physical and social science; progressing to the relative side could be found social sciences, including

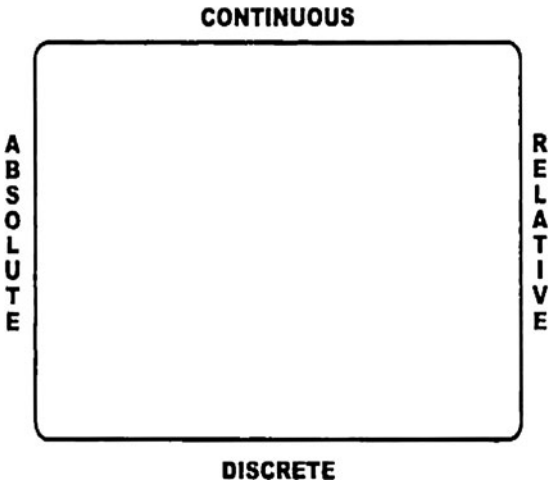

Fig. 1.1 Varying views of reality [68]

Sociology and Economics, given their emphasis on human-created environments and institutions. Psychology would be closest to the relative side (pure human interpretations and perception of reality). Where to locate Mathematics? It is the unique case that does not fall in any single area of this framework but rather seems to have aspects or subfields throughout (for example, Euclidean geometry is absolute and very continuous in nature, but topology is relative and discrete).

1.1.3 Space, Time, Spacetime and Current Practice in Geostatistics

As outlined in subsection 1.1.1, Einstein's theory of relativity established that space and time are both interdependent and inseparable. Space, consisting of three dimensions (up-down, left-right, and forward-backward) and time are all part of what's called the space-time continuum. However most of the theory and applications of statistics for spatio-temporal data, including Cressie and Wikle (2011) book, which we follow in the sequel, are related to phenomena which reside in a Newtonian framework because some modifications should be done in the classical analysis to deal with, say, astronomical data.

Next we reproduce the Cressie-Wikle version of Einstein's thought experiment: *Think of a boxcar being pulled by a train travelling at velocity v , and place a source of light at the center of the moving boxcar. An observer on the train sees twin pulses of light arrive at the front and rear end of the boxcar, simultaneously. A stationary observer standing by the train tracks sees one pulse arrives at the rear end of the boxcar before its twin arrives at the front end. That is, the reference frame of the observer is extremely important to the temporal notions of simultaneity/before/after. What ties together space and time is movement (velocity) of the boxcar'.*

Again following [20], Einsteinian physics assumes that the velocity of light is a universal constant regardless of the frame of reference. Thus, for any frame of reference, the distance traveled by a pulse of light is equal to the time taken to travel that distance multiplied by the above constant. That this relationship holds under any spatio-temporal coordinate system means that for Einsteinian physics, space and time are inextricably linked. Other physical properties are modified, too. The length of an object measured in the moving frame, moving with velocity v, is always smaller than or equal to the length of the object measured in the stationary frame, by a factor $[1-(v/c)^2]^{1/2}$. A similar factor shortens a time interval in a moving frame, leading to the famous conclusion that the crew of a spaceship flying near the speed of light would return in a few (of their) years to find that their generation on Earth had become old.

However, although Einstein's ideas are certainly relevant for some kind of phenomena, the practice of Geostatistics assumes Newtonian physical laws. That is, Geostatistics works with a coordinate system that is a Cartesian product of the space (three-dimensional) and time (one-dimensional), while respecting the directionality of the temporal coordinate, and uses models of spatio-temporal processes to attempt to capture the statistical dependencies that can arise from the evolution of phenomena at many spatial and temporal scales.

1.2 Geostatistics at Work in Natural Sciences and Social Sciences

The study of spatial and spatio-temporal dependencies is one of the hot questions in statistical and econometrical research. We live in a globalized physical world and, in addition, the join effects of space and time on variables and attributes are becoming a core question in almost all scientific disciplines.

Although the First Law of Geography, which stipulates that observations close together in space are more likely to be similar than those farther apart [82], was proposed forty years ago, only with the raise of geographic information science and technologies, sophisticated statistical and econometric spatial and spatio-temporal methods have been widely developed and applied in geographical sciences.

But, as stated in [20], this will be a century of massive (spatio-temporal) datasets collected to answer Society's dominant questions. And much of that questions are fundamental to sustaining our planet (those related with Climate and Environment) and to improve social and economic conditions (those related with Education, Health, Economics, Finance, etc.), and they involve complex spatio-temporal phenomena and are inherently statistical.

One more aspect should be considered. As outlined in [42] and stated in [58] and [28], the past two decades have witnessed space-time convergence: transportation and communication technologies have shrunk the world to an incredible degree. Locations on the Earth's surface are much closer to each other with respect to the time required for movement and interaction. Sometimes, physical distance vanishes

(this is the case in Finance). But even in such cases physical distance can be substituted by financial, economic, or social distances, and spatial and spatio-temporal strategies continue to be at the core of stochastic modeling and statistical inference.

To address the patterns and statistical variability we can opt for a data-driven orientation (Statistics) or a model-driven orientation (Econometrics) [3]. In both approaches, spatial, and especially spatio-temporal arguments, are emerging as key arguments in current research and will be paid more and more attention in the future. Irrespective of the approach, the old assumption of independent observations is not valid anymore in a geo-referenced world and this complicates statistical inferences, parameter estimation and prediction. As a consequence, more realistic techniques are needed. An interesting aspect of the path followed by the above mentioned disciplines is that the incorporation of the spatial argument and of the joined effects of space and time has lead them to work together. As an example, interpolated variables are considered as core explanatory variables in econometric models, which results in new challenges and, as a consequence, in new research and new procedures (see, for example, [4, 28, 59] for a discussion on the potential "errors in variables" aspect of interpolated variables when they are used as explanatory variables in spatial hedonic models as well as consequences and solutions).

As stated above, spatial and spatio-temporal strategies are currently at the core of a wide range of scientific disciplines. Focusing on spatial statistics, the force driving the development of Geostatistics was practical and economic. In Russia meteorologists wanted to interpolate atmospheric variables from sparse recording stations; in South Africa miners wanted to estimate the gold contents of ores locally from measurements on drill cores; elsewhere petroleum engineers wished to estimate oil reserves from logged boreholes; and all wanted their estimates to be unbiased with minimum variance. Local estimation, *i.e.* spatial prediction, was the ultimate goal of Geostatistics, and kriging was the means of achieving that goal. [45] had written out the equations for the purpose in the 1930s, but without computers no one could solve them. The advent of computers gave mining and petroleum engineers the opportunity. As a consequence, Geostatistics rapidly became in an evolving branch of applied mathematics.

The earlier applications were in mining, where the first steps were taken in South Africa, with the work of the mining engineer Krige and the statistician Sichel [46, 47, 79], followed by applications in forestry and hydrology. Typical points that were addressed included locating and quantifying minable resources, making a forest inventarisation, modelling (and predicting) changes in hydrological components, to name just a few. Much of the original impetus for the subject was driven by Geostatistics. It was in this context that the technique of kriging, optimal least squares interpolation over a random spatial field, was originally developed. Some theory was available, for example the random function theory as developed by [85] in the nineteen sixties. But that was largely insufficient to find generic solutions for the whole class of problems and, hence, applications required a new theory.

The innovative ideas of Krige and Sichel attracted the attention of French mining engineers and in particular of Georges Matheron, who put them together in a single framework: his celebrated Theory of Regionalized Variables [1, 53–57]. The new

theories were applied to new disciplinary problems leading to modifications and extensions of mathematical and statistical procedures. Well-defined problems with a common character were suddenly on the agenda and data availability and intensive discussion with practical and disciplinary researchers resulted in new theoretical developments. It is often difficult to say which came first, but different theoretical models were developed for different applications. Thus, there is no surprise that with the advent of high-speed computers in the seventies the applications of Geostatistics spread from the original metal mining topics (Matheron, who coined the term Geostatistics, used it to designate his own methodology of ore reserve evaluation) to such diverse fields as Soil Science, Oceanography, Hydrogeology, Agriculture (especially in precision farming), Environmental and Ecological Sciences (with particularly fruitful applications), and more recently Health Science, Sociology (especially social networks theory), the practice of commerce and military planning (Logistics) and the development of efficient spatial networks, Real Estate, Education, Psychology, Neighborhood research, Archeology, Financial Economics, Demographic Modelling and even Politics (see [67] for applications of Geostatistics to election forecast). The first international meeting on the subject was organized in Rome in 1975 [35]. In the eighties two further international congresses took place, namely at Lake Tahoe, U.S.A., in 1983 [83] and in Avignon, France, in 1988 [5]. It could be said that at that moment the discipline was consolidated. Of course, a continuum of conferences and congresses followed from the eighties until now.

However, when it comes to indicate the areas where Geostatistics could be applied, the most celebrated text books continuous to refer only to natural variables. Without the intention to be exhaustive, and taking [18] as a guide, [15, 22, 23, 43] focused on applications in mining industry; [65] on hydrology and water resources, [10] on soil mapping, [81] on atmospheric science, [41] on groundwater contaminant concentration. [18] provides some more applications of Gesostatistics in Chap. 4 of his celebrated text, but they refer to the same areas we have outlined above. [14] established the field of Geostatistics in the spatial or spatio-temporal study of variables such as ore grades in a mineral deposit, depth and thickness of a geological layer, porosity and permeability in a porous medium, density of trees of a certain species in a forest, soil properties in a region, rainfall over a catchment area, pressure, temperature, and wind velocity in the atmosphere, and concentrations of pollutants in a contaminated site. [20], which surveys the most recent developments in the statistical analysis of spatio-temporal data, is recommended for Climate and Environmental Sciences. Note that much more work exists where the principles of Geostatistics (second order assumptions; best linear unbiased prediction) are applied in a spatial context without referring to geostatistics itself.

As outlined above, more and more disciplines are calling for sophisticated models of spatial statistics. A particularly delicate, hence interesting area is the field of Social Sciences including Real Estate research, Neighborhood research and Medical Geography. [33] points out that the latter is a field where Geostatistics has a promising future for application (see also [48], [7]). Medical Geography or spatial epidemiology [24], is concerned with both (1) the study of spatial patterns of disease incidence and mortality and (2) the identification of potential “causes” of disease,

such as environmental exposure or socio-demographic factors [84], and has received little attention in the geostatistical literature. This lack of attention contrasts with the increasing need for methods to analyze health data following the emergence of new infectious diseases (*e.g.* West Nile Virus, bird flu), the higher occurrence of cancer mortality associated with longer life expectancy, and the burden of a widely polluted environment on human health.

Empirical and Bayesian methods have been traditionally used in the analysis of health data and putative covariates, such as environmental, socio-economic, behavioral or demographic factors. [33] argues that Geostatistics represents an attractive alternative to increasingly popular Bayesian spatial models in that: (1) it is easier to implement and less CPU intensive since it does not require lengthy and potentially non-converging iterative estimation procedures, and (2) it accounts for the size and shape of geographical units, avoiding the limitations of conditional auto-regressive (CAR) models commonly used in Bayesian algorithms while allowing for the prediction of the risk over any spatial support. However, Bayesian modelling has been successfully combined with geostatistical methods in other areas [25, 63].

Of course, as data are typically aggregated over irregular spatial supports leading to the "modifiable area unit problem" that may have a massive impact on statistical results [64]. In order to tackle such problems, tools existing in Geostatistics have to be modified or even taken from the larger tool box of spatial statistics including hierarchical models, graphical models, marked point processes and other models from stochastic geometry, such as germ-grain models and tessellations [34]. Additional problems appear as the quantity of interest are quotients of estimated quantities where, mostly, the denominator is the estimated population size. The behaviour of such quotient estimators is still not fully explored [80].

Neighborhood research assumes similarity among individuals residing in a particular neighborhood and differences between neighborhoods; thus, issues of spatial dependence are implicit in this kind of research. Community psychologists studying neighborhood effects usually turn to hierarchical linear modeling to test multilevel theories that explain neighborhood effects by examining the links between neighborhood characteristics and resident outcomes. They consider random effects models to account for the lack of independence between individuals or observations nested within neighborhoods. While it is possible to account for the clustering or grouping of individuals within neighborhoods using these methods, the grouping unit is typically a geographic region defined by administrative sources (*e.g.*, census tracts). Geostatistical methods do not require this a priori spatial definition, but instead examine spatial point locations. As a result, these methods do not rely solely on predefined boundaries that may not be consistent with residents' perceptions of neighborhood boundaries [16]. As an example, [69] shows that neighbourhood boundaries are crucial to determine the size and statistical significance of effects of neighborhood crime and neighborhood socioeconomic status on residents' perceptions of neighborhood problems.

Research on real estate relies more and more on spatial analyses, see [71] and [18], for instance, for a discussion of relevant spatial statistics methods.

The most popular application is the use of variogram to directly estimate the variance-covariance matrix of errors in a traditional or augmented econometric model and then compute an estimated generalized least squares estimate of the regression coefficients. From this or a similar view, [9] treats spatial models with both geostatistics techniques and hedonic models using a large sample of sales in New Zeland. Instead of proceeding as usual, using geographical coordinates or other spatial indicators as regressors, parametrically or even nonparametrically [11,12,29] as explanatory variables in the hedonic model to take into account spatial effects, they geostatistically model the disturbance term, and investigate whether spatial statistical models perform better than an OLS model with neighborhood dummy variables. Using a small sample from Baltimore, [26] compares exsample predictions using OLS and a geostatistical technique, and concludes that the geostatistical approach is superior even when some neighbourhood (census block group) characteristics are included as explanatory variables. Further examples of successful application of geostatistical models are given by [6, 26, 66].

[50] use geostatistics, spatial autoregressive models and geographically weighted regression in a complementary way for localizing real estate submarkets homogenous in respect of price, and direct modeling of the variance-covariance matrix later used in GLS estimation.

[13] uses kriging methods, isotopic data cokriging, and heterotopic data cokriging methods to obtain estimates of house pricing in Granada, Spain, arguing that interpolated maps of house price are interesting for appraisers, real estate companies, and bureaus because they provide an overview of location prices. Cokriging has been intended to take account for two different groups of data: sold housing samples and not-for-sale housing samples where the price has been estimated for instance for tax reasons, an information that has not received much attention in the literature on hedonic price models [49]. Other co-variates appear in spatial hedonic housing price models such environmental information (pollutants, noise, etc., measured at monitoring stations). Examples of their effect are given in [4, 28] and indirectly in [59, 61], for instance.

Finally, it is well known that understanding the spatial component within a commercial portfolio is essential since fundamental portfolio theory shows that spatial correlation is an unsystematic risk that should not be compensated by the market. [36] is a pioneer work in application of spatial statistics in the real estate field relative to spatial diversification, dividend policy and credit scoring.

1.3 Modelling Spatio-Temporal Dependencies

1.3.1 Introduction

It is well known that spatio-temporal statistics recognizes and exploits the space-time locations of data when designing for collecting, managing, analysing, and displaying such data. Spatio-temporal data are typically dependent, for which there

are classes of spatio-temporal models available that allow for process prediction and parameter estimation.

The study of spatio-temporal variability is a rather new area within Statistics. There has been a growing awareness in the last 10 years that knowing where and when data were observed could help enormously in answering the substantive questions that precipitated their collection. Of course, that realization is strongly linked to the improvements on computational power. Going a little beyond, interpolation-based kriging procedures strongly depend on the choice of the autocovariance associated to a space-time random field. Thus, the geostatistical challenge is to obtain permissible dependence structures for space and time; in other words, what is needed are adequate spatio-temporal covariance models associated to stationary or non-stationary, isotropic or anisotropic random fields. On a note apart, recent literature persistently emphasizes the use of approximation methods and new methodologies for dealing with massive spatio-temporal data sets. When dealing with spatio-temporal data, calculation of the inverse of covariance matrices becomes a crucial problem. For instance, the inverse is needed for best linear unbiased prediction, and is repeatedly calculated in maximum likelihood estimation or Bayesian inferences. Thus, large spatio-temporal sample sizes results in another big challenge from the computational point of view.

But both having a wide range of spatio-temporal models available that allow for process prediction in most of situations the researcher could find, and parameter estimation are not easy tasks, and in the two last decades much effort has been made to enlarge the set of valid covariographic models and to overcome the computational problems that usually arise when dealing with spatio-temporal data. This is why, in this section, we briefly relate the story of spatio-temporal data and some of the solutions to computational burden. Previously, we revise some basic, but core, concepts to deal with spatio-temporal covariance functions.

1.3.2 Basic Concepts in the Analysis of Spatio-Temporal Random Functions

Let $Z(\mathbf{s}, t)$, $\mathbf{s} \in D \subset \mathbf{R}^d$, $t \in T \subset \mathbf{R}$, denote a space-time process where each of Z, D and T is possibly random. Geostatistical approaches have been developed to fit random field models in continuous space and time settings, based on a limited number of spatially and/or temporally dispersed observations. These approaches model the observations as a partial realisation of a spatial-temporal, typically Gaussian, random field. Under the geostatistical framework, the domain is fixed and we assume to draw observations, in given points, from a certain distribution. In Chap. 4 the location of the sample sites moves according to a certain probability setting.

In the following we assume for simplicity that $D = \mathbb{R}^d$ and $T = \mathbb{R}$.

Definition 1.1. A spatio-temporal random field $Z(\mathbf{s},t)$ is said to be Gaussian if all the finite dimensional distributions of the random vector $\mathbf{Z} = (Z(\mathbf{s}_1, t_1), \ldots,$

$Z(\mathbf{s}_n, t_n))'$ for any set of spatio-temporal locations $(\mathbf{s}_1, t_1), \ldots, (\mathbf{s}_n, t_n) \in \mathbb{R}^{d+1}$ are multivariate Gaussian.

Definition 1.2. A spatio-temporal random field Z is strictly stationary if its probability distribution is translation invariant. In other words, if, in the case of any separation vector $(\mathbf{h}, u) \in \mathbb{R}^{d+1}$, the vectors $\mathbf{Z}$ and $\mathbf{Z}_{\mathbf{h},u} = (Z(\mathbf{s}_1 + \mathbf{h}, t_1 + u), \ldots, Z(\mathbf{s}_n + \mathbf{h}, t_n + u))'$ have the same distribution for any $\mathbf{s}_i, \mathbf{h} \in \mathbb{R}^d$, $t_i, u \in \boldsymbol{R}, i = 1, \ldots, n$, and $n \in \mathbb{N}$.

A space-time random field is called weakly stationary if it has finite first and second moment, and if the covariance between $\mathbf{Z}$ and $\mathbf{Z}_{\mathbf{h},u}$ only depends on $(\mathbf{h}, \mathbf{u})$.

Gaussianity and weak stationarity are prominent and convenient assumptions when dealing with space-time data, although often unrealistic. The Gaussian assumption is fundamental for conditional und unconditional simulation and for evaluating the impact of the misspecification of the covariance function on optimal linear prediction (for this last concept, the reader is referred to Chap. 3). Motivated by geological, meteorological and agricultural applications, Allard discusses in Chap. 7 the ways to overcome the assumption of Gaussianity and illustrates the constructions obtainable through two huge classes of processes: the transformed Gaussian and the skew-normal processes.

The assumption of weak stationarity can be weakened by assuming, for instance, instrinsic stationarity. Here, the increments of a space-time random field on two different space time locations have zero mean and their variance depends on the space-time separation, *i.e.*

$$\mathrm{Var}\left(Z(\mathbf{s}, t) - Z(\mathbf{s}', t')\right) = \gamma(\mathbf{s}' - \mathbf{s}, t' - t), \quad s, s' \in \mathbb{R}^d, \quad t, t' \in \mathbb{R}.$$

The function $\gamma(\cdot)$ is called variogram. Chap. 2 points to some of its properties and illustrates their use for unconditional simulation of weakly stationary and intrinsically stationary Gaussian random field.

Several simulation algorithms are based on Bochner's spectral representation [8] that establishes a one to one correspondence between covariance functions and the Fourier transforms of finite non-negative measures. The proof uses the fact that a covariance function is positive definite function, and vice versa. The latter is synonmic to the fact that for any finite system of locations $(\mathbf{s}_i, t_i)$, $i = 1, \ldots, n$, and any constants $a_i, \ldots, a_n$, the variance of linear combinations, which can be written as

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i Z(\mathbf{s}_i, t_i)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}\left(Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)\right),$$

must be nonnegative. A similar property holds for a variogram, which is called conditionally negative definite since the bilinear combination

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) \tag{1.1}$$

shall be nonpositive for any $n \in \mathbb{N}$, and for any $(\mathbf{s}_i, t_i) \in \mathbb{R}^d \times \mathbb{R}$ and $a_i \in \mathbb{R}$, $i = 1, \ldots, n$, with $\sum_{i=1}^{n} a_i = 0$.

The results from [78] give the relation between variogram and covariance functions.

Theorem 1.1. *Let γ a real valued function with $\gamma(0) = 0$. Then, the following statements are equivalent:*

- *γ is a variogram.*
- *$\exp(-\theta\gamma)$ is a covariance function for any $\theta > 0$.*
- *$C((\mathbf{s}_i,t_i),(\mathbf{s}_j,t_j)) = \gamma((\mathbf{s}_i,t_i),(\mathbf{0},0)) + \gamma((\mathbf{s}_j,t_j),(\mathbf{0},0)) - \gamma((\mathbf{s}_i,t_i),(\mathbf{s}_j,t_j))$ is a covariance function.*

Assuming stationarity of the increments of order k leads to building generalized covariance functions. In Chap. 8 several generalized covariances are built on the basis of stochastic differential equations, in particular through heat equations driven by a random noise or through Burger equations. In Chap. 9, a generalized covariance is discussed as a general family including as special cases some ordinary covariances that are well known in the literature.

A stationary space time covariance is called spatially isotropic if $C(\mathbf{h}, u_o) = K(\|\mathbf{h}\|, u_o)$ for $\|\cdot\|$ being usually the Euclidean norm and for any fixed temporal lag u_o. Some other metric may be used but with much care since, in general, the composition of the function K with any seminorm does not offer a positive definite function. The reader is referred to the beautiful results in [86] where covariances depending on the norm are studied. The function C is also called fully symmetric if $C(\mathbf{h}, u) = C(\mathbf{h}, -u)$ for any $\mathbf{h} \in \boldsymbol{R}^d$ and $u \in \mathbb{R}$. An example is the class of isotropic covariance functions which build a basic model component in the area of meteorology [17] when modelling rainfall. Many models proposed in the last twenty years regard such constructions as shall be explained subsequently.

A spatio-temporal random field has a separable covariance function if the spatio-temporal covariance structure factors into a purely spatial and purely temporal component, C_S and C_T, say,

$$C(\mathbf{h}, u) = C_S(\mathbf{h})C_T(u) \quad \text{or} \quad C(\mathbf{h}, u) = C_S(\mathbf{h}) + C_T(u), \quad \mathbf{h} \in \mathbb{R}^d, \quad u \in \mathbb{R}. \tag{1.2}$$

In all the other cases, the covariance shall be called nonseparable.

Equivalently, [52] speaks about reflection symmetry; complete symmetry is another usual term for this concept. It has been shown by [32] that if the covariance function is fully symmetric, so is its associated spectral density, if it exists. Another interesting aspect is that separable covariances are also fully symmetric, while reverse is not necessarily true. Hence, covariance structures that are not fully symmetric are non-separable, and tests for full symmetry ([75] and [52]) can be used to reject separability.

We shall say that a space time isotropic covariance function is space-time compactly supported if C has compact support, *i.e.*, $C(\mathbf{h}, u) = 0$ whenever

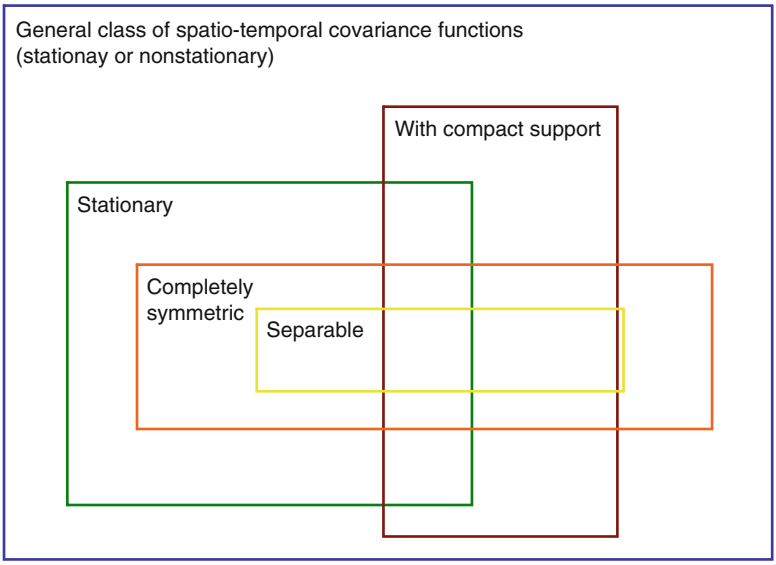

Fig. 1.2 Relations between the spatio-temporal covariance functions (Gneiting et al. 2007)

$\|(\mathbf{h}, u)\| \geq r$ for some $r \geq 0$. A less severe limitation would come in if, independently of the temporal argument, the space time covariance is identically zero whenever $\|\mathbf{h}\| \geq r$, or symmetrically, whenever $|u| \geq r$ independently of the spatial lag. Such assumption of space time compact support is for the meantime just pedagogical since, to our knowledge, there are no models with such feature available in the literature, apart of the trivial separable construction. But this definition comes as an important perspective, in view of the increasing importance of a statistical procedure celebrated in the literature under the name of tapering. For a thorough illustration of such procedure the reader is referred to Chap. 10, who illustrates the effect of tapering on asymptotic optimal prediction, and Chap. 3, where the authors apply the idea of tapering in several contexts, with the aim of overcoming the computational burdens induced by working with large spatial or space time datasets.

In [32] the relationships between the various notions are nicely summarized in terms of classes of spatio-temporal covariance functions, and an analogous scheme applies to correlation structures (see Fig. 1.2). The largest class is that of general stationary or non-stationary covariance functions. A separable covariance function can be stationary or non-stationary, and similarly for fully symmetric covariances.

Covariance functions are a convex cone closed under the topology of pointwise convergence. This act has been repeatedly used in the last years to build new models of space time covariances. Thus, linear combinations, products, limits and scale mixtures are friendly tools for the construction of such covariances. Such properties are explained in Chap. 2, where Schlather also discusses the properties

of variograms, being also a convex cone closed under the topology of pointwise convergence, but with no closure property with respect to the product, expect to some special cases as shown in [70].

1.4 Overview

The geostatistical space time framework has been very popular for over 20 years, but it cannot be the universal key for the analysis of space time datasets since the complexities of space time are extremely variable depending on the type of dataset, the purposes of the analysis, the researcher perspective, and many other factors. Adequate extensions include marked point process approaches and extreme value theory, where the usual second order assumptions are questionable [76, 77].

This book presents various aspects of the rapidly developing area space time modelling. The majority of the approaches is nonetheless based on space time covariances or variograms, for which we refer to Chap. 2, where we can find indications about the construction of covariances and variograms as well as how to simulate processes using these objects.

A big challenge is related to the analysis of massive datasets, which imply a huge computational burden for both estimation and prediction. This argument is illustrated in Chap. 3. The authors illustrate the scenarios arising recently with the aim of overcoming the computational burdens whilst preserving a certain level of statistical efficiency. Such efficiency can be measured either with respect or estimation or with respect to prediction, and this is precisely what they highlight when talking about tapering. Likelihood approximations, in the spatial or in the spectral domain, represent a nice alternative to tapering as they discuss. Recently, a nice approach based on PDE has been proposed by [51], although with severe limitations due to the fact that the solution is available in closed form for some parameters only.

As just mentioned, statistical efficiency can be evaluated with respect to prediction or to estimation. Such argument is valid also for the case of spatial or space time design, as discussed in Chap. 6. Two approaches are considered to design experiments for a correlated random field when the objective is to obtain precise predictions over the whole experimental domain. Both take the uncertainty of the estimated parameters of the correlation structure of the random field into account.

Spatial and space time extremes represent a very important and quickly growing area of space time statistics. Chapter 5 offers a review of the most important methods for the construction of spatial extremes. After a brief illustration of the extreme value theory for univariate and multivariate values, this chapter focuses on spatial max-stable processes, illustrating also the features related to statistical inference and simulation for these processes. Max-stable processes are also contrasted with spatial hierarchical models.

The problem of large and complex datasets is thoroughly discussed in Chap. 4, but from the perspective of Bayesian inference for space time point processes,

which have become more and more popular, especially thanks to the development of Markov chain Monte Carlo methods, which expand the scope of application of Bayesian methods considerably. In particular, the authors mainly focus on Bayesian contributions to inference for point processes, with emphasis to non-Markovian processes, specifically Poisson and related models, doubly stochastic models, and cluster models. They also discuss Bayesian model selection for these models and give examples in which Bayes factors are applied both directly and indirectly through a reversible jump algorithm.

The assumption of normality if fundamental for, *e.g.*, the simulation of random fields with a certain dependence structure (see Chap. 2), and justified by the stability of multivariate normal distribution under summation and conditioning which offers tractability and simplicity. On the other hand, Gaussian spatial processes are well modeled and understood by the statistical and scientific communities, but for a wide range of environmental applications Gaussian spatial or spatio-temporal models cannot reasonably be fitted to the observations. Motivated by these facts, Allard offers in Chap. 7 several ways to avoid the assumption of Gaussianity starting from a Gaussian random field and transforming according to some rule. He discusses the effects of such transformations on the moments of the resulting random field and illustrates the issues related to estimation.

Space time covariances are nice tools since they allow for estimation and prediction starting from an object available in closed form. But in many situations the use of a covariance model respect to another is not justified by the laws of physics or by the natural characteristics of the phenomenon of interest. If one wants to consider such a perspective, then Chap. 8 will be a fundamental reference. It provides a general overview on the main results derived by the authors in relation to limit theory for the solution of linear and non-linear random evolution equations. Additionally, they show the local regularity properties of the solution to fractional pseudodifferential equations driven by random innovations. Specifically, limit results derived for the heat and Burgers equations with linear and quadratic external potentials are described in the first part of the chapter. In the second part, the local quadratic variation properties of the solution to fractional pseudodifferential equations on regular and fractal domains are formulated.

Two critical aspects of random fields are their behavior over short scales (or equivalently, at high frequencies) and over long scales (or low frequencies). Some standard models have flexibility at one but not both of these scales. Recent years have seen a number of proposed models that have at least some flexibility at both scales. In Chap. 9, Porcu and Stein list the most prominent examples and analyze their local and global behaviour, either in terms of covariance or associated spectra.

What about asymptotics? What kind of asymptotics should we use for spatial processes? And, is there a way to use a kind of asymptotics conciling the existing ones? Chap. 10 starts asserting that *useful asymptotics are those that can help with the statistical inferences.* For example, asymptotics can help identify statistically and computationally efficient estimators. The chapter illustrates how the study of asymptotics in spatial statistics is complicated by the fact there are more than one asymptotic frameworks in spatial statistics and the asymptotic results are very

different under the different asymptotic frameworks. This chapter also reviews some results under these asymptotic frameworks and shows how the asymptotic results can help alleviate the computational challenges in the analysis of massive spatial data.

Reference

1. Agterberg, F.P.: Appreciation of contributions by Georges Matheron and John Tukey to a mineral-resources research project. Nat. Resour. Res. **10**, 287–295 (2001)
2. Akhundov, M.D.: Conceptions of Space and Time: Sources, Evolutions, Directions. MIT Press, Cambridge, Mass (1986)
3. Anselin, L.: Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Boston, MA. (1988)
4. Anselin, L., Lozano-Gracia, N.: Errors in variables and spatial effects in hedonic house price models of ambient air quality. Empir. Econom. **34**, 5–34 (2008)
5. Armstrong, M. (Ed.): Geostatistics. Kluwer, Amsterdam (1989)
6. Basu, A., Thibodeau, T.G.: Analysis of spatial autocorrelation in house prices. J. Real Estate Finance **17**(1), 61–85 (1998)
7. Berke O.: Exploratory disease mapping: kriging the spatial risk function from regional count data. Int. J. Health Geographics. **3**, 18 (2004)
8. Bochner, S.: Vorlesungen über Fouriersche Integrale. Akademische Verlagsgesellschaft, Leipzig (1932)
9. Bourassa, S. C., Cantoni, E., Hoesli, M.: Spatial dependence, housing submarkets, and house price prediction. J. Real Estate Finance **35**(1), 142–160 (2007)
10. Burgess, T.M., Webster, R.: Optimal interpolation and isarithmic mapping of soil properties, I. The semivariogram and punctual kriging. J. Soil Sci. **31**, 315–331 (1980)
11. Clapp, J.M.: A semiparametric method for valuing residential locations: application to automated valuation. J. Real Estate Finance **27**(3), 303–320 (2003)
12. Colwell, P.F.: A primer on piecewise parabolic multiple regression analysis via estimations of Chicago CBD land prices. J. Real Estate Finance **17**(1), 87–97 (1998)
13. Chica-Olmo, J.: Prediction of housing location price by a multivariate spatial method: cokriging. J. Real Estate Res. **29**(1), 91–114 (2007)
14. Chilès, J.P., Delfiner, P.: Geostatistics. Modeling Spatial Uncertainty. John Wiley & Sons, New York, Chichester (1999)
15. Clark, I.: Does geostatistics work? In Proceedings of the 16th APCOM International Synzposiuni, T.J. O'Neil, ed. Society of Mining Engineers of the AIME, New York, 213–225 (1979)
16. Coulton, C. J., Korbin, J. E., Chan, T., Su, M.: Mapping residents' perceptions of neighborhood boundaries: A methodological note. Am. J. Commun. Psychol. **29**(2), 371–383 (2001)
17. Cox, D.R., Isham, V.S.: A simple spatial-temporal model of rainfall. Proc. R. Soc. Lond. Ser. A, Math. Phys. Eng. Sci. **415**, 317–328 (1988)
18. Cressie, N.A.C.: Statistics for Spatial Data. John Wiley & Sons, New York (1993)
19. Cressie, N.A.C., Huang, C.: Classes of nonseparable, spatiotemporal stationary covariance functions. J. Amer. Statist. Assoc. **94**, 1330–1340 (1999)
20. Cressie, N., Wikle, C.K.: Statistics for Spatio-Temporal Data. Wiley, Hoboken, New Jersey (2011)
21. Dainton, B.: Time and Space. McGill-Queen's University Press, Montreal (2001)
22. David, M.: Geostatistical Ore Reserve Estimation. Elsevier, Amsterdam (1977)
23. David, M.: Handbook of Applied Advanced Geostatistical Ore Reserve Evaluation. Elsevier, Amsterdam (1988).

24. Diggle, P., Moyeed, R., Rowlingson, B, Thomson, M.: Childhood malaria in the Gambia: a case-study in model-based geostatistics J. R. Statist. Soc. Ser. C (Applied Statistics), **51**, 493–506 (2002)
25. Diggle, P.J., Ribeiro, P.J. Jr.: Model-Based Geostatistics. Springer, New York (2007)
26. Dubin, R.A.: Estimation of regression coefficients in the presence of spatially autocorrelated error terms. Rev. Econ. Stat. **70**(3), 466–474 (1988)
27. Dubin, R., Pace, R.K., Thibodeau, T.G.: Spatial autoregression techniques for real estate data. J. Real Estate Literature **7**(1), 79–95 (1999)
28. Fernández-Avilés, G., Minguez, R., Montero, J.M.: Geostatistical air pollution indexes in spatial hedonic models. The case of Madrid, Spain. J. Real Estate Res., Accepted article yet un assigned to a volume. http://aux.zicklin.baruch.cuny.edu/jrer/papers/abstract/forth/accepted/jrer_175(f110101r3).htm
29. Fik, T.J., Ling, D.C., Mulligan, G. F.: Modeling spatial variation in housing prices: a variable interaction approach. Real Estate Econ. **31**(4), 623–646 (2003)
30. Futch, M.J.: Leibniz's Metaphysics of Time and Space. Boston Studies in the Philosophy of Science 258, Springer Science (2008)
31. Garber, D.: Descartes' physics. In J. Cottingham (Ed.) The Cambridge Companion to Descartes, 286-334, Cambridge University Press, Cambridge (1992)
32. Gneiting, T., Genton, M. G. and Guttorp, P. Geostatistical space-time models, stationarity, separability and full symmetry. In Finkenstaedt, B., Held, L. and Isham, V. (eds.), Statistics of Spatio-Temporal Systems, Chapman & Hall/CRC Press, Monographs in Statistics and Applied Probability, 151–175 (2007)
33. Goovaerts, P.: Medical geography: a promising field of application for geostatistics. Math. Geol. **41**, 243–264 (2009)
34. Gotway, C.A., Young, L.J.: Combining incompatible spatial data. J. Amer. Statist. Assoc. **197**, 632–648 (2002)
35. Guasarascio, M., David, M., Huijbregts, C. (Eds.): Advances Geostatistics in the Mining Industry. NATO ASI Series C-24. Reidel, Dordrecht (1976)
36. Hayunga, D.R., Pace, R.K.: Spatial aspects of commercial real estate. J. Real Estate Finan. Econ. **41**(2), 103–125 (2010)
37. Hesiod (West, M.L. (Ed.)): Theogony and Works and Days. Oxford World's Classics Series, USA (1999)
38. Heisemberg, W.: Tradition in Science. Science and Public Affairs, XXIX **10**, 4–10 (1973)
39. Huggett, N.: Space from Zeno to Einstein. MIT Press, Cambridge MA (1999)
40. Huggett, N.: Motion and relativity before Newton. PhilPapers: online research in philosophy (2008)
41. Istock, J.D.: Geostatistics applied to groundwater pollution. III Global estimates. J. Environ. Eng. **114**, 915–928 (1988)
42. Janelle, D.G.: Spatial organization: A model and concept. Ann. Assoc. Am. Geogr. **59**, 348–364 (1969)
43. Journel, A. G., Huijbregts, C. J.: Mining Geostatistics. Academic Press, London (1978)
44. Kamara, E. J.: Space, time, and theology in the Leibniz-Newton controversy. Ontos Verlag, New York (2006)
45. Kolmogorov, A.N.: Sur l'interpolation et l'extrapolation des suites stationnaires. C. R. Acad.Sci. **208**, 2043–2045 (1939)
46. Krige, D.G.: A statistical approach to some mine valuation and allied problems on the Witwatersrand. Master's thesis, University of Witwatersrand (1951)
47. Krige, D.G.: A statistical analysis of some of the borehole values in the Orange Free State gold field. J. Chem. Metall. Soc. S. Afr. **53**, 47–64 (1952)
48. Lawson A.B.: Statistical Methods in Spatial Epidemiology. John Wiley & Sons, New York (2001)
49. LeSage, J.P., Pace, R.K.: Spatial and spatiotemporal econometrics. In J.P. LeSage, and R.K. Pace (Eds.), Advances in Econometrics Elsevier, Oxford, 1–32 (2004)

50. Ligas, M., Kulczycki, M.: Spatial statistics for real estate data. Strategic Integration of Surveying Services FIG Working Week, Hong Kong SAR, China (2007)
51. Lindgren, F, Rue, H., Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, J. R. Statist. Soc., Ser. B, **73**(4), 423–498 (2011)
52. Lu, N., Zimmerman, D.L.: Testing for directional symmetry in spatial dependence using the periodogram. J. Statist. Plann. Inference **129**, 369–385 (2005)
53. Matheron, G.: Application des méthodes statistiques l'évaluation des gisements. Annales des Mines **144**(12), 50–75 (1962a)
54. Matheron, G.: Traité de Géostatistique Appliquée, Vol. 1. Technip, Paris (1962b)
55. Matheron, G.: Principles of geostatistics. Econ. Geol. **58**, 1246–1266 (1963)
56. Matheron, G.: Les Variables Régionalisées et leur Estimation, Mason, Paris (1965)
57. Matheron, G.: The Theory of Regionalized Variables and its Applications. Les Cahiers du Centre de Morphologie Mathématique 5. Ecole des Mines de Paris, Fontainebleau (1969)
58. Miller, H.J.: Tobler's first law and spatial analysis. Ann. Assoc. Am. Geogr. **94**(2), 284–289 (2004)
59. Minguez, R., Montero, J.M., Fernández-Avilés, G.: Measuring the impact of pollution on property prices: objective vs. subjective pollution indicators in spatial models. J. Geograph. Systems, to appear (2012)
60. Minkowski, H.: Space and time. In H.E. Lorentz, H.A., H. Weyl, H. Minkowski (Eds.). Einstein. The Principle of Relativity: A Collection of Original Papers on the Special and General Theory of Relativity, 73–91. Dover, New York (1952)
61. Montero, J.M., Larraz, B.: Space-time approach to commercial property prices valuation. Appl. Econ. **44**(28), 3705–3715 (2012)
62. Norton, J.D.: The hole argument. The Stanford Encyclopaedia of Philosophy (Winter 2008 Edition), In E.N. Zalta (ed.). http://plato.stanford.edu/archives/win2008/entries/spacetime-holearg
63. Ober, U., Erbe, M., Long, N., Porcu, E., Schlather, M., Simianer, H.: Predicting genetic values: a kernel-based best linear unbiased prediction with genomic data. Genetics. Published online (2011)
64. Openshaw, S., Taylor, P.: A million or so correlation coefficients: three experiments of the modifiable area unit problem. In Wrigley, N. (ed.), Statistical Applications in the Spatial Science. Pion, London, 217–244 (1979)
65. Ord, J.K., Rees, M.: Spatial processes: recent developments with applications to hydrology. In E.H. Lloyd, T. O'Donnell, and J.C. Wilkinson (eds), The Mathematics of Hydrology and Water Resources. Academic Press, London, 95–118 (1979)
66. Pace, R. K., Barry, R.: Quick computation of regressions with a spatially autoregressive Dependent Variable. Geogr. Anal. **29**(3), 232–247 (1997)
67. Pavía, J.M., Larraz, B., Montero, J.M.: Election forecasts using spatio-temporal models. J. Amer. Statist. Assoc. **103**(483), 1050–1059 (2008)
68. Peuquet, D.J.: Representations of space and time. The Guildford Press, New York (2002)
69. Pierce, S.J.: Using geostatistical models to study neighborhood effects: an alternative to using hierarchical linear models. UMI Dissertation Publishing, Ann Arbor (2009)
70. Porcu, E., Schilling, R.: From Schoenberg to Pick-Nevanlinna: towards a complete picture of the variogram class. Bernoulli, published online (2011)
71. Ripley, B.: Spatial Statistics. Wiley, New York (1981)
72. Rynasiewicz, R.: Newton's views on space, time, and motion. In E.N. Zalba (Ed.) The Stanford Encyclopaedia of Philosophy (fall 2008 Edition) http://plato.stanford.edu/archives/fall2008/entries/newton-stm
73. Sack, R.: Conceptions of space in social thought: A geographic perspective. University of Minnesota Press, Minneapolis (1980)
74. Sahu, S. K. and Mardia, K.V.: Recent trends in modeling spatio-temporal data. Meeting of the Italian Statistical Society on Statistics and the Environment. Messina, Italy, 69–83 (2005)

75. Scaccia, L., Martin, R.J.: Testing axial symmetry and separability of lattice processes. J. Stat. Plan. Infer. **131**, 19–39 (2005)
76. Schlather, M., Ribeiro, P.J., Diggle, P.J.: Detecting dependence between marks and locations of marked point processes. J. R. Statist. Soc., Ser. B **66**, 79–83 (2004)
77. Schlather, M., Tawn,, J.A.: A dependence measure for multivariate and spatial extreme values: properties and inference. Biometrika **90**, 139–156 (2003)
78. Schoenberg, I.J.: Metric spaces and completely monotone functions. Ann. Math. **39**, 811–841 (1938)
79. Sichel, H.S.: New methods in the statistical evaluation of mine sampling data. Trans. Inst. Mining and Metall. **61**, 261–288 (1952)
80. Stoyan, D., Stoyan, H.: Improving ratio estimators of second order point process characteristics. Scand. J. Stat. **27**, 641–656 (2000)
81. Tiebaux, H.J., Pedder, M.A.: Spatial Objective Analysis with Applications in Atmospheric Science. Academic Press, London (1987)
82. Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. Econ. Geogr. **46**, 234–240 (1970)
83. Verly, G., David, M. Journel, A.G. (Eds.): Geostatistics for Natural Resources Characterization. NATO ASI Series C-122, Reidel, Dordrech (1984)
84. Waller, L.A., Gotway, C.A.: Applied Spatial Statistics for Public Health Data. John Wiley & Sons, New Jersey (2004)
85. Yaglom, A.M.: An Introduction to the Theory of Stationary Random Functions. Prentice-Hall, Englewood Cliffs, N.J. (1962)
86. Zastavnyi, V.P.: Positive definite functions depending on the norm. Russian J. Math. Physics **1**(4), 511–522 (1993)

Chapter 2
Construction of Covariance Functions and Unconditional Simulation of Random Fields

Martin Schlather

Abstract Covariance functions and variograms are the most important ingredients in the classical approaches to geostatistics. We give an overview over the approaches how models can be obtained. Variant types of scale mixtures turn out to be the most important way of construction. Some of the approaches are closely related to simulation methods of unconditional Gaussian random field, for instance the turning bands and the random coins. We discuss these methods and complement them by an overview over further methods.

2.1 Introduction

Random fields are used to model regionalized variables [65] such as temperature, humidity, soil moisture, wave heights or metal concentrations of reservoirs, to mention a few. A random field, Z say, can be seen as a random real function on $\mathbb{R}^d$, or as a bundle of dependent random variables $Z(x)$, indexed by $x \in \mathbb{R}^d$. Assuming that the variances exist, such a random field can be characterized by its expectation and its covariance function

$$C(x, y) = \operatorname{cov}(Z(x), Z(y)), \qquad x, y \in \mathbb{R}^d.$$

These two characteristics determine the random field uniquely if the field is Gaussian, i.e. if $(Z(x_1), \ldots, Z(x_n))$ has a multivariate Gaussian distribution for any $x_i \in \mathbb{R}^d$ and $n \in \mathbb{N}$. Considering the variances of linear combinations $\sum_{k=1}^{n} a_k Z(x_k)$ with $a_k \in \mathbb{R}$ and $x_k \in \mathbb{R}^d$, we get that

M. Schlather (✉)
Centre for Statistics, Georg-August-Universität Göttingen, Goldschmidtstr. 7,
D – 37077 Gottingen, Germany
e-mail: schlather@math.uni-goettingen.de

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_2,

$$\sum_{k=1}^{n}\sum_{j=1}^{n} a_k\, C(x_k, x_j)\, a_j \geq 0 \qquad \text{for all } x_k \in \mathbb{R}^d,\ a_k \in \mathbb{R}, \text{ and for all } n \in \mathbb{N}. \tag{2.1}$$

Hence, $C(x, y)$ cannot be any arbitrary function. On the other hand, Kolmogorov's existence theorem (cf. [9], for instance) shows that if a symmetric, real-valued function C satisfies (2.1) then at least a Gaussian random field exists that has C as covariance function. Note that not all covariance functions are compatible with a given marginal distribution. For instance, a log-Gaussian process on the real axis cannot have the cosine as covariance function [1, 68].

Beyond characterizing (Gaussian) random fields from both a practical and a theoretical point of view, covariance functions are the key elements to determine likelihoods, to perform simulations and to spatially interpolate data (kriging).

In this chapter, we concentrate on the construction of covariance functions. In Sect. 2.2, we give methods that are as elementary as important. Sections 2.3–2.5 introduce the spectral approach, the convolutions, and the power series. The approaches in Sects. 2.2–2.4 are closely related to simulation methods for unconditional Gaussian random fields. Hence, they are presented on the way. Unconditional simulations are the key ingredients for conditional simulations [55] and are used for simulation studies. Scale mixtures, discussed in Sect. 2.6, allow for an elegant way to construct models. In particular, scale mixtures of the "Gaussian" covariance model, $C(x, y) = \exp(-\|x - y\|)^2)$, play an exceptional role. The turning bands method, presented in Sect. 2.7, is primarily a simulation method, but also defines a way to construct covariance models. In Sect. 2.8, the *montée* is presented. Section 2.9 gives an overview over simulation methods that are not related to the construction of covariance functions. Sections 2.10 and 2.11 deal with the advanced topics of space-time covariance functions and multivariate covariance models. Some excercises are given in section 2.12.

Henceforth, we will always assume that the expectation of the random field is zero. Translation invariant covariance functions, i.e. covariance functions C with $C(x, y) = \varphi(x - y)$ for some function $\varphi : \mathbb{R}^d \to \mathbb{R}$, play a dominant role when modelling spatial data. In this case, the function φ is called a positive definite function. A corresponding random field is called (weakly) stationary. If, furthermore, the covariance function is motion invariant, i.e. $C(x, y) = \tilde{\varphi}(\|x - y\|)$ for some function $\tilde{\varphi} : [0, \infty) \to \mathbb{R}$, then the corresponding random field is called (weakly) stationary and isotropic. Henceforth, $\|\cdot\|$ will always denote the Eucledian distance.

If $Z(x + h) - Z(x)$ is weakly stationary for all $h \in \mathbb{R}^d$, then the random field Z is called intrinsically stationary and the (uncentred) (semi-)variogram γ is used to characterize the random field:

$$\gamma(h) = \frac{1}{2}\,\mathbb{E}(Z(h) - Z(0))^2.$$

Matheron [66] shows that a function $\gamma : \mathbb{R}^d \to [0, \infty)$ is a variogram if and only if $\gamma(0) = 0$ and γ is conditionally negative definite, i.e., γ is symmetric and

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(x_i - x_j) \leq 0 \quad \text{for all } x_i \in \mathbb{R}^d, a_i \in \mathbb{R} \text{ with } \sum_{i=1}^{n} a_i = 0, n \in \mathbb{N}. \tag{2.2}$$

If Z is even weakly stationary then $\gamma(h) = \varphi(0) - \varphi(h)$.

For theoretical considerations, we will also consider complex valued random fields and hence complex-valued covariance functions C, i.e., functions that satisfy

$$\sum_{k=1}^{n}\sum_{j=1}^{n} a_k \, C(x_k, x_j) \, \overline{a_j} \geq 0 \qquad \text{for all } x_k \in \mathbb{R}^d, a_k \in \mathbb{C}, \text{ and for all } n \in \mathbb{N}. \tag{2.3}$$

[78] shows that any complex valued function satisfying (2.3) is Hermitian.

Complementaries and applications are given, for instance, in the books of [15], [18], and [55]. Related review papers are given by [36] and [60], for example. See also the technical report by [84].

Most of the models, many construction principles and nearly all simulation methods given here are available within the R package RandomFields of [87].

2.2 Basic Constructions of Positive Definite Functions

A simple, but also important example of a covariance function is the scalar product $C(x, y) = \langle x, y \rangle$. Most generally, let $H : \mathbb{R}^d \to \mathscr{H}$ be a mapping into a Hilbert space $\mathscr{H}$. Then

$$C(x, y) = \langle H(x), H(y) \rangle_{\mathscr{H}} \tag{2.4}$$

is a covariance function. This representation of covariance functions is used particularly in machine learning, see [27] and [95], for instance. As a consequence, the function

$$C(x, y) = e^{i\langle t, x-y \rangle} \tag{2.5}$$

is a covariance function for any fixed $t \in \mathbb{R}^d$. Here, i is the imaginary number.

Further, if C is a covariance function on $\mathbb{R}^d$ and A is a linear mapping from $\mathbb{R}^m$ into $\mathbb{R}^d$, then $C(A\cdot, A\cdot)$ is a covariance function on $\mathbb{R}^m$. In particular, rescaling $C(s\cdot, s\cdot)$, $s > 0$, does not change the property (2.1).

Remark 2.1. If A has full rank then the corresponding random field is called geometrically anisotropic, otherwise zonally anisotropic. Such kind of anisotropies are frequently assumed due to preferential directions of underlying processes. Note that the zonal anisotropy implies that if $\tilde{\varphi}(\|\cdot\|)$ is a positive definite function in $\mathbb{R}^d$ so is $\tilde{\varphi}(\|\cdot\|)$ in $\mathbb{R}^k$ with $k < d$.

Also sums and products of covariance functions are again covariance functions [15, 18]. This can easily be seen by considering sums and products of respective

independent random fields. In particular, vC is a covariance function for any constant $v \geq 0$ since a non-negative constant function is positive definite.

Assume C_n is a sequence of covariance functions that converges pointwise to some function C,

$$C(x, y) = \lim_{n\to\infty} C_n(x, y), \qquad x, y \in \mathbb{R}^d. \tag{2.6}$$

Then, it can be easily seen that condition (2.1) holds also for C if $C(x, x)$ is finite for all $x \in \mathbb{R}^d$.

These basic construction principles for covariance functions already allow us to create many classes of covariance functions.

2.3 Spectral Representation

Equations (2.5) and (2.6), or (2.4) with a suitably defined scalar product, yield that

$$\varphi(h) = \int_{\mathbb{R}^d} e^{i\langle\omega,h\rangle}\mu(d\omega) \tag{2.7}$$

is a positive definite, complex valued function for any finite, non-negative measure μ on $\mathbb{R}^d$. For real-valued random fields we have

$$\varphi(h) = \int_{\mathbb{R}^d} \cos(\langle\omega, h\rangle)\mu(d\omega).$$

It is easy to see that φ is uniformly continuous. Bochner's celebrated theorem [10, 11] gives the reverse statement, namely that all continuous positive definite functions have a unique representation (2.7).

The representation (2.7) allows for an immediate simulation procedure. Let $Z(x) = \sqrt{\mu(\mathbb{R}^d)}e^{i(\langle R,x\rangle+\Phi)}$ or $Z(x) = \sqrt{\mu(\mathbb{R}^d)}\cos(\langle R, x\rangle + \Phi)$ where $\Phi \sim U[0, 2\pi)$ and $R \sim \mu/\mu(\mathbb{R}^d)$ are independent. Then Z is (strongly) stationary, i.e., the finite dimensional distributions of $(Z(x))_{x\in\mathbb{R}^d}$ and $(Z(x+h))_{x\in\mathbb{R}^d}$ are the same for any $h \in \mathbb{R}^d$. The marginal distributions are not multivariate Gaussian. However, an approximation Z' to a Gaussian random field is obtained if $Z_i, i = 1, \ldots, n$, are independent and identically distributed according to Z and $Z' = n^{-1/2}\sum_{i=1}^n Z_i$ for some n large enough.

Example 2.1. The important Whittle-Matérn model [44, 62, 90],

$$W_\nu(h) = 2^{1-\nu}(\Gamma(\nu))^{-1}\|h\|^\nu K_\nu(\|h\|), \qquad \nu > 0, \tag{2.8}$$

has spectral density

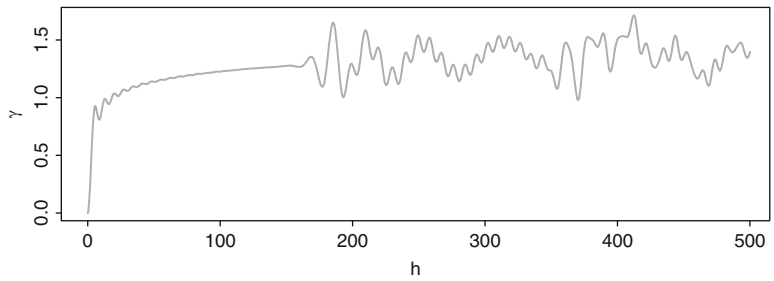

Fig. 2.1 A variogram in $\mathbb{R}$ constructed by an infinite sum over cosine functions, see example 2.2

$$\frac{d\mu(\omega)}{d\omega} = \frac{\Gamma(\nu + d/2)}{\Gamma(\nu)\pi^{d/2}(1 + \|\omega\|^2)^{\nu+d/2}}.$$

Here, Γ is the Gamma function and K_ν is a modified Bessel function of the second kind.

Example 2.2. Variograms with exceptional properties can be obtained by sums of cosine functions. However, they do not have any practical relevance. Let

$$\gamma(h) = \sum_{k=1}^{\infty} a_k(1 - \cos(h/b_k)).$$

If $a_k = 1$ and $b_k = k!$ then $\liminf_{h\to\infty} \gamma(h) = 0$ and $\limsup_{h\to\infty} \gamma(h) = \infty$ [4]. Independently, [50] showed that these two properties hold also for $b_k = 2^k$. Figure 2.1 illustrates γ for $a_k = 10^{-3} \cdot k^{1.1}$ and $b_k = 1.1^{1.1^k}$.

2.3.1 Spectral Turning Bands

An important special case appears when μ is rotation invariant and thus can be represented by spherical coordinates α and a radial coordinate r, i.e.,

$$\mu(d\omega) = s_{d-1}^{-1} d\alpha F(dr) \tag{2.9}$$

for some finite non-negative measure F on $[0, \infty)$. Here, s_d denotes the surface area of the d-dimensional sphere. Integrating over α in (2.7) we get

$$\tilde{\varphi}(r) = \int_{[0,\infty)} B_{(d-2)/2}(rs) F(ds) \quad \text{for all } r \in [0, \infty), \tag{2.10}$$

where

$$B_\nu(r) = \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(\nu+1)}{k!\Gamma(\nu+k+1)} \left(\frac{r}{2}\right)^{2k} = \Gamma(\nu+1)\left(\frac{2}{r}\right)^{\nu} J_\nu(r), \quad \nu > -\frac{1}{2}, \tag{2.11}$$

and J_ν is the Bessel function. Hence, the function $\varphi(h) = B_{(d-2)/2}(\|h\|)$ is *the* elementary rotation invariant, continuous positive definite function in $\mathbb{R}^d$. For $d = 1, 2$, and 3, the function $J_{(d-2)/2}(h)$ equals

$$\sqrt{2}\cos(h)/\sqrt{\pi h}, \qquad 2\pi^{-1}\int_0^\infty \sin(h\cosh t)dt, \quad \text{and} \quad \sqrt{2}\sin(h)/\sqrt{\pi h},$$

respectively [46]. In analogy to Bochner's theorem, [88] stated that a rotation invariant function $h \mapsto \tilde{\varphi}(\|h\|)$, $h \in \mathbb{R}^d$, is real, continuous and positive definite if and only if $\tilde{\varphi}$ is the Hankel transform (2.10) of a non-negative finite measure F on the half-line $[0, \infty)$. Note that equation 6.567.1 in [42] ensures that $B_\nu(\|h\|)$ is a positive definite function on $\mathbb{R}^d$ for any $\nu \geq (d-2)/2$.

Remark 2.2. In three dimensions we have

$$\tilde{\varphi}(r) = \int_{[0,\infty)} \frac{\sin rs}{rs} F(ds),$$

i.e., the elementary rotation invariant positive definite function in $\mathbb{R}^3$ is $\varphi(h) = \sin(\|h\|)/\|h\|$, the so-called *hole effect model*.

Example 2.3. Equations 6.649.2, 6.618.1, and 6.623.3 in [42] consider functions $\tilde{\varphi}$ of the form (2.10), and hence yield that $\tilde{\varphi}(\|h\|)$ is a positive definite function on $\mathbb{R}^d$ if

1. $\nu \geq (d-2)/4$ and $\tilde{\varphi}(r) = 2\nu I_\nu(r) K_\nu(r)$,
2. $\nu \geq (d-2)/4$ and $\tilde{\varphi}(r) = \begin{cases} 2^\nu \Gamma(\nu+1) r^{-2\nu} e^{-r^2} I_\nu(r^2), & r \neq 0 \\ 1, & r = 0 \end{cases}$,
3. $\nu \geq \max\{0, (d-2)/2\}$ and $\tilde{\varphi}(r) = \begin{cases} \frac{2^\nu \left(\sqrt{1+r^2}-1\right)^\nu}{r^{2\nu}}, & r > 0 \\ 1, & r = 0 \end{cases}$,

respectively. Here, I_ν denotes the Bessel I-function. For instance, the first model is $\lfloor \nu \rfloor$ times differentiable where $\lfloor \nu \rfloor$ denotes the largest integer less than or equal to ν; it decays at rate h^{-1} to infinity.

Remark 2.3. The function $B_\nu(2\sqrt{\nu}r)$ converges to the function $r \mapsto \exp(-r^2)$ as $\nu \to \infty$. Since $B_\nu(\|\cdot\|)$ is a positive definite function in $\mathbb{R}^d$ for $d < 2\nu + 2$, the "Gaussian" covariance function $C(x, y) = \exp(-\|x-y\|^2)$ is the candidate for a fundamental motion invariant covariance function that is valid in all dimensions d. This is indeed true, see Sect. 2.6.2.

The simulation method that uses decomposition (2.9) is called spectral turning bands method in geostatistics, see [61]. A random field Z with a motion invariant covariance function $C(x, y) = \tilde{\varphi}(\|x - y\|)$ is obtained if

$$Z(x) = \sqrt{2F([0,\infty))}\cos(R\langle S, x\rangle + \Phi)$$

and $R \sim F/F([0,\infty))$ is given by (2.9), $\Phi \sim U[0, 2\pi)$, and $S \sim U\mathscr{S}_{d-1}$ is uniformly distributed on the $(d-1)$-dimensional sphere $\mathscr{S}_{d-1}$. All random variables are independent. Again, $Z' = n^{-1/2}\sum_{i=1}^{n} Z_i$ yields an approximation to a Gaussian random field for Z_i, $i = 1, \ldots, n$, that are independent and identically distributed according to Z. The value of n should be of order 500 to get good results. Figure 2.2 shows the performances of the method for the "Gaussian" covariance function.

Remark 2.4. The spectral representation by Bochner and Schoenberg leaves the question open, which discontinuous positive definite functions exist and which are of practical interest. In practice, only one discontinuous model exists that is regularly used as a summand in additive covariance models, the so-called nugget effect $\varphi(h) = \mathbf{1}_{\{0\}}(h)$ ([15], for instance). Here, $\mathbf{1}_A$ denotes the indicator function for a set A, i.e. $\mathbf{1}_A(h)$ equals 1 if $h \in A$ and 0 otherwise. It is easily seen that

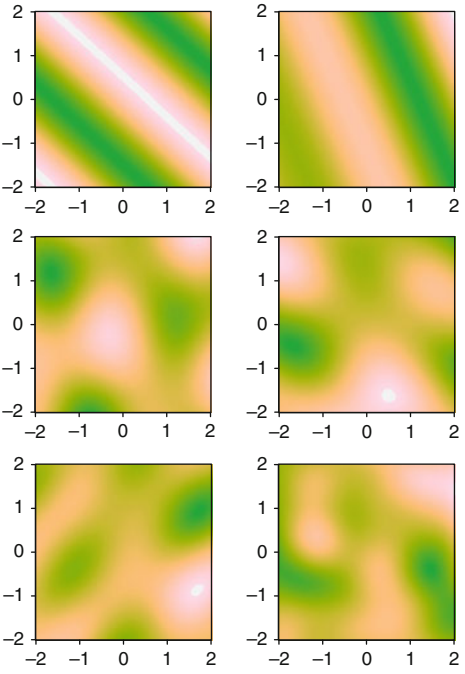

Fig. 2.2 Simulation of the spectral turning bands method with 1, 2, 3, 4, 10, and 1,000 lines (*top left to bottom right*); the random field has the "Gaussian" covariance function

the nugget effect is a positive definite function in any dimension. More generally, for any subgroup Q of $\mathbb{R}^d$, the validity of inequality (2.1) is readily checked for $C(x, y) = \mathbf{1}_Q(x - y)$.

Any measurable positive definite function φ is a sum of a continuous positive definite function and a positive definite function that vanishes almost everywhere [78]. If, additionally, φ is rotation invariant and $d \geq 2$, then φ must be the sum of a continuous positive definite function and a nugget effect [38]. However, covariance functions do not need to be measurable [78].

2.4 Convolutions and Random Coin Method

Another immediate consequence of equation (2.4) is that

$$\varphi(h) = \int_{\mathbb{R}^d} f(x) f(x + h) dx, \qquad h \in \mathbb{R}^d, \tag{2.12}$$

is a positive definite function for any real-valued L_2-function f on $\mathbb{R}^d$. The function φ is called a covariogram. If f is an indicator function, then φ is also called a set covariance function.

Whilst in $\mathbb{R}^1$ many functions f lead to analytic formulae for φ, the situations where the explicit calculation of φ is feasable are limited in higher dimensions. Examples are $f(x) = (\pi/4)^{d/4} \exp(-2\|x\|^2)$ leading to the "Gaussian" model $\varphi(h) = \exp(-\|h\|^2)$, and the indicator functions of the d-dimensional balls of radius $1/2$, up to a multiplicative constant, yielding covariance functions with finite range 1, i.e. compact support. Examples are the hat function $\varphi(h) = (1 - |h|)_+$ for $d = 1$, the circulant model $\varphi(h) = 1 - 2\pi^{-1}(\|h\|\sqrt{1 - \|h\|^2} + \arcsin(\|h\|))$, $\|h\| \leq 1$, in $\mathbb{R}^2$ and the spherical model $\varphi(h) = 1 - \frac{3}{2}\|h\| + \frac{1}{2}\|h\|^3$, $\|h\| \leq 1$, in $\mathbb{R}^3$. See [33] for further properties of these functions, and sufficient conditions for positive definiteness based on these properties.

A random field that corresponds to (2.12) can be defined as

$$Z(x) = \sum_{y \in \Pi} f(x - y)$$

where Π is a stationary Poisson point process on $\mathbb{R}^d$ with intensity $\lambda = 1$. The random field Z has a direct interpretation as the sum of effects of certain events $y \in \Pi$ and is therefore a convenient model for a non-Gaussian random field in many applications. It possesses a lot of names, for instance, dilution random field [15], random coin model, random token model [55], shot noise process [17, 69], moving average model [62], and trigger process [20]. Of course, an approximation to a Gaussian random field can be obtained through the central limit theorem.

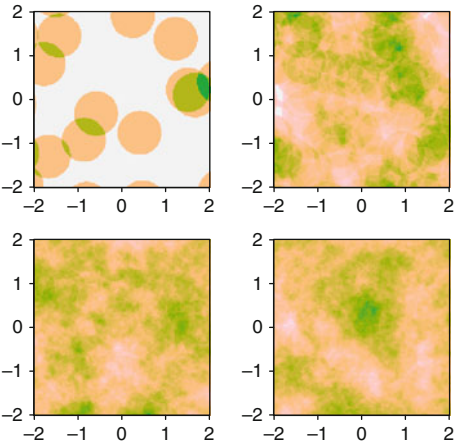

Fig. 2.3 Recentred and renormalised superpositions of 1, 10, 100, and 1,000 simulations of an additive Boolean model with radius $r = 1/2$ of the disks (*top left, top right, bottom left, bottom right*)

Figure 2.3 shows Z for $\lambda = 4/\pi$ and f the indicator function of a disk with radius 0.5, i.e., the covariance function of Z is the circulant model. A satisfying approximation to the Gaussian distribution is obtained if $n \approx 500$ independent realizations are superposed.

Remark 2.5. A related method to obtain positive definite functions and corresponding random fields replaces the product in the integrand of (2.12) by a maximum:

$$\psi(h) = \int_{\mathbb{R}^d} \max\{f(x), f(x+h)\}dx, \qquad h \in \mathbb{R}^d,$$

Here, f is a non-negative, integrable function. Then, ψ is a conditionally negative definite function and the function $\varphi(h) = 2\int f(x)dx - \psi(h)$ is positive definite. These functions appear in extreme value theory and are called extremal coefficient functions ψ or extremal correlation functions φ [25]. A random field that has $C(x, y) = \varphi(x - y)$ as its covariance function appears as a thresholded max-stable random field [85], a special class of Boolean random functions [47].

Both, the spectral representation and the convolution representation are special cases of the Karhunen orthogonal representation [52]. We refer here to the version of [8] who give a more rigorous proof and more general results.

Theorem 2.1. *Let Z be a second order random field on $V \subset \mathbb{R}^d$, i.e.* $\operatorname{Var} Z(x)$ *exists for all $x \in \mathbb{R}^d$. Assume that for some measurable space $(W, \mathscr{W})$, the covariance function C allows for a representation*

$$C(x, y) = \int_W g(x, s)\overline{g(y, s)}F(ds), \qquad x, y \in V,$$

where

1. *F is a positive, σ-finite measure on W;*
2. *L is the L_2 space of functions that are square integrable with respect to F;*
3. *$g : V \times W \to \mathbb{C}$ is such that $g(x,\cdot) \in L$ for all $x \in V$;*
4. *$\dim(\mathrm{span}\{g(x,\cdot) : x \in V\}^{\perp}) \leq \dim(\overline{\mathrm{span}}\{Z(x) : x \in V\})$ where the complement is taken with respect to L.*

Then Z can be represented as

$$Z(x) = \int_W g(x,s)\,d\zeta(s), \qquad x \in V,$$

where ζ is a uniquely determined random orthogonal measure on $\mathscr{W}_0 = \{A \in \mathscr{W} : F(A) < \infty\}$ with $\zeta(A \cup B) = \zeta(A) + \zeta(B)$ for all disjoint $A, B \in \mathscr{W}_0$ and $\mathbb{E}\,\zeta(A)\zeta(B) = F(A \cap B)$ for any $A, B \in \mathscr{W}_0$.

This theorem complements Mercer's theorem [7] which implies that any continuous covariance function $C(x,y)$ on a compact set can be decomposed into eigenfunctions. In case the eigenvalues drop quickly towards zero, fast simulation algorithms for excellent approximations can be obtained by neglecting eigenfunctions that have small eigenvalues.

2.5 Power Series

Since products and pointwise limits of covariance functions are covariance functions, power series of covariance functions with summable, non-negative coefficients yield further models.

For instance, consider the Taylor development of $(1+x)^q$ ([42], formula 1.10), i.e.,

$$(1+x)^q = 1 + qx + \frac{q(q-1)}{2!}x^2 + \cdots + \frac{q(q-1)\ldots(q-k+1)}{k!}x^k + \ldots$$

Then we get that

$$C_1(x,y) = (M - C(x,y))^q - M^q, \qquad q < 0, \quad M > \sup C, \tag{2.13}$$

$$C_2(x,y) = \sum_{j=0}^{2k+1} \frac{q(q-1)\ldots(q-j+1)}{j!}[-C(h)]^j M^{q-j} - (M - C(h))^q,$$

$$q \in (2k, 2k+1), \quad k \in \mathbb{N}_0, \quad M \geq \sup C,$$

and

$$C_3(x,y) = (M - C(h))^q - \sum_{j=0}^{2k} \frac{q(q-1)\dots(q-j+1)}{j!}[-C(h)]^j M^{q-j}, \quad (2.14)$$

$$q \in (2k-1, 2k), \quad k \in \mathbb{N}_0, \quad M \geq \sup C,$$

are covariance functions for any covariance function C. In particular,

$$\varphi(0)^q - (\varphi(0) - \varphi(\cdot))^q, \quad (2.15)$$

is a positive definite function for $q \in (0,1]$ and any positive definite function φ. The function $1 - \frac{1}{\sqrt{2}}(1-\rho)^{1/2}$ has the form (2.15) up to an additive constant, and appears as the covariance function of a thresholded extremal Gaussian random field [85].

Further examples of functions that have power series with non-negative coefficients are exp, sinh and cosh. Hence, if C is a covariance function, so are $\exp(C)$, $\sinh C$ and $\cosh C$. See also [86].

2.5.1 Application to Variograms

If γ is a variogram then

$$C(x,y) = \gamma(x) + \gamma(y) - \gamma(x-y)$$

is a covariance function [66]. This is readily seen if an intrinsically stationary random field Y with variogram γ is considered and the covariance function of $Z(x) = Y(x) - Y(0)$ is calculated. As $e^{-s(\gamma(x)+\gamma(y))}$ is a covariance function by (2.4), cf. [66], it follows that

$$h \mapsto \exp(-s\gamma(h)) \quad (2.16)$$

is a positive definite function for all $s > 0$ and any conditionally negative definite function γ. Since $\gamma(h) = \lim_{s\to 0} s^{-1}(1 - e^{-s\gamma(h)})$, the reverse holds as well, i.e., if the function given by (2.16) is positive definite for all $s > 0$, then γ is a conditionally negative definite function.

Remark 2.6. Equations (2.16) and (2.15) yield that for any conditionally negative definite function γ and any $q \in (0,1]$ the function

$$\gamma_q(h) = \lim_{s\to 0}(s^{-1}(1 - e^{-s\gamma(h)}))^q = \gamma^q(h) \quad (2.17)$$

is non-negative and conditionally negative definite. As additing a constant does not change the property of a function being conditionally negative definite, $(\gamma + a)^q - a^q$ is a variogram for any variogram γ and $a \geq 0$.

Note that for $q > 1$, the function γ^q may not be a variogram anymore. In general, products of variograms are not variograms. See [100] for a discussion and classes of examples. In contrast, any convex combination of variograms is a variogram.

Example 2.4. It is immediately seen from inequality (2.2) that $\gamma(h) = \langle h, h\rangle$ is a variogram for any scalar product $\langle\cdot,\cdot\rangle$. Equation (2.17) yields that $h \mapsto \|h\|^\alpha$, $\alpha \in (0, 2]$ is a variogram model for any dimension d. If $d = 1$, the corresponding random field is called fractional Brownian motion, and Brownian motion if $\alpha = 1$. Equation (2.13) yields that $(1+\gamma)^{-\beta} = \lim_{s\to 0}(1 + s^{-1} - s^{-1}\exp(-s\gamma))^{-\beta}$ is a positive definite function for any variogram γ and $\beta > 0$. Hence, the generalized Cauchy model [28],

$$\varphi(h) = (1 + \|h\|^\alpha)^{-\beta/\alpha} \tag{2.18}$$

and, by (2.16), the powered exponential model $\varphi(h) = \exp(-\|h\|^\alpha)$ are positive definite functions on $\mathbb{R}^d$ for any $d \in \mathbb{N}$, $\beta > 0$ and $\alpha \in (0, 2]$.

Although power series are useful for constructing covariance functions, they have not been of direct use for simulating random fields.

2.6 Mixtures

Equation (2.6) yields that $C = \int C_v \mu(dv)$ is a covariance function if μ is a non-negative finite measure and C_v are covariance functions such that C is finite everywhere. In this case, C is called a mixture of the models C_v.

Example 2.5. Integrating (2.16) over the interval $[0, 1]$ with respect to s yields that

$$\varphi(h) = \begin{cases} \dfrac{1 - e^{-\gamma(h)}}{\gamma(h)}, & \gamma(h) \neq 0 \\ 1, & \gamma(h) = 0 \end{cases}$$

is a positive definite function for any variogram γ.

2.6.1 Scale Mixtures

The most important class of mixtures are the scale mixtures. Let φ and φ_0 be complex-valued functions on $\mathbb{R}^d$. The function φ is called a *scale mixture* of φ_0 if there exists a non-negative measure F on $[0, \infty)$, such that

$$\varphi(h) = \int_{[0,\infty)} \varphi_0(sh) F(ds) \qquad \text{for all } h \in \mathbb{R}^d \tag{2.19}$$

or, more generally,

$$C((x_1,\dots,x_d),(y_1,\dots,y_d)) =$$
$$\int_{[0,\infty)^d} C_0((s_1x_1,\dots,s_dx_d),(s_1y_1,\dots,s_dy_d))\,F(d(s_1,\dots,s_d)), \qquad x_i, y_i \in \mathbb{R},$$

for some non-negative measure F on $[0,\infty)^d$. For instance, all continuous, isotropic covariance functions are scale mixtures of Bessel functions, see Sect. 2.3.1.

Example 2.6. The scale mixture of the "Gaussian" model with mixing density

$$f(s) = \frac{(\kappa/\delta)^\lambda}{2K_\lambda(\delta\kappa)} s^{\lambda-1} \exp\big(-(\kappa^2 s + \delta^2/s)/2\big)$$

yields the generalized hyperbolic model [3, 28, 89],

$$\tilde{\varphi}(r) = \frac{\delta^{-\lambda}}{K_\lambda(\kappa\delta)} (\delta^2 + r^2)^{\lambda/2} K_\lambda(\kappa(\delta^2 + r^2)^{1/2}), \qquad r \geq 0.$$

Here, the parameters λ, κ, and δ satisfy:

$$\delta \geq 0,\ \kappa > 0 \text{ for } \lambda > 0,$$
$$\delta > 0,\ \kappa > 0 \text{ for } \lambda = 0,$$
$$\delta > 0,\ \kappa \geq 0 \text{ for } \lambda < 0.$$

It includes, as special cases, the Cauchy model (2.18) with $\alpha = 2$ and the Whittle-Matérn model in example 2.1.

2.6.2 Completely Monotone Functions

A continuous function ψ on $[0,\infty)$ with $\psi(0) \in \mathbb{R} \cup \{\infty\}$ is called completely monotone function if it is infinitely often differentiable and $(-1)^n \psi^{(n)}(r) \geq 0$ for any $r \in (0,\infty)$ and $n \in \mathbb{N}$. It is well-known [98] that ψ is completely montone if and only if it is a scale mixture of the exponential function, i.e.,

$$\psi(r) = \int_0^\infty e^{-sr} F(ds), \qquad r > 0, \tag{2.20}$$

for some non-negative measure F such that ψ is finite on $(0,\infty)$. A function ψ is called absolutely monotone if all derivatives are positive.

Since $\exp(-s\gamma)$ is a positive definite function for any $s > 0$ and any variogram γ, the function $\psi(\gamma)$ is positive definite on $\mathbb{R}^d$ for any bounded, completely monotone

function ψ and any variogram γ on $\mathbb{R}^d$. As $h \mapsto \|h\|^2$ is variogram for any dimension d, we get that $\psi(\|h\|^2)$ is a covariance function for any dimension d and any bounded, completely monotone function ψ. [88] proved that the reverse also holds. Namely, if $\psi(\|h\|^2)$, $h \in \mathbb{R}^d$, is a continuous and isotropic positive definite function in all dimensions $d \in \mathbb{N}$, then ψ is a bounded, completely monotone function.

Since $1 - e^{-s\gamma} = s\int_0^\gamma e^{-ts}dt$ is a variogram for any variogram γ we get that $\int_0^\gamma \psi(u)du$ is a variogram for any completely monotone, integrable function ψ. A non-negative function on $(0,\infty)$ that is infinitely often differentiable and whose first derivative is completely monotone is called a Bernstein function. For particular properties and a considerable amount of examples, see [74] and [83].

Example 2.7. The conditional negative definiteness of γ^α, $\alpha \in (0,1)$, see equation (2.17), also follows immediately from the fact that $r \mapsto r^\alpha$ is a Bernstein function.

Example 2.8. A completely monotone function is $r \mapsto (1+r)^{-1}$, cf. (2.18), which implies that

$$h \mapsto \log(\gamma(h) + 1)$$

is a variogram for any variogram γ. If $\gamma(h) = \|h\|^\alpha$, $\alpha \in (0,2]$, then the model $h \mapsto \log(\|h\|^\alpha + 1)$ is called de Wijsian model [96].

Example 2.9. The concatenation of two Bernstein functions is a Bernstein function [5]. This is a consequence of the product rule for the nth derivative, which implies that the product of two completely monotone functions is completely monotone. Hence,

$$r \mapsto \int_0^{f(r)} g(f^{\leftarrow}(s))ds$$

is a Bernstein function for any completely monotone function g and any Bernstein function f. For instance, choosing $g(r) = \exp(-r)$ and $f(r) = r^{1/2}$ shows that $\operatorname{erfc}(\sqrt{\gamma})$ is a covariance function for any variogram γ. The latter function appears as the covariance function of a thresholded Brown-Resnick process [51].

Remark 2.7. If ψ is a bounded and absolutely monotone function and C is a covariance function then $\psi(C)$ is a covariance function, see Sect. 2.5. Let $0 < M < \pi/2$, $\alpha \in (0,1)$ and ρ be a covariance function with $|\rho| \leq 1$. Then the following functions are also covariance functions

$$\rho/(1-e^{-4M\rho}), \quad \arcsin\rho, \quad \tan M\rho, \quad \operatorname{cosec}(\rho) - \rho^{-1}, \quad (2M\rho)^{-1} - \cot(2M\rho),$$

$$\sec\rho, \quad -\log(1-\alpha\rho), \quad \log|\rho/\sin\rho|, \quad -\log\cos(M\rho), \quad \log|\tan(M\rho)/(M\rho)|.$$

The function $\arcsin\rho$ appears as the covariance function of a thresholded Gaussian random field, see [2] for instance.

Remark 2.8. If ψ is a bounded, absolutely monotone function, the function $\psi(\cdot) - \psi(0)$ is also absolutely monotone and should be considered preferentially, since the

covariance function $\psi(C)$ with $\psi(0) \neq 0$ always refers to a non-ergodic random field.

2.6.3 More Mixture Models

Mixtures of $\exp(-\lambda\gamma)$ yield several mappings from the set of variograms to the set of positive definite functions.

For instance, equation 6.521.3 in [42] and equation (2.8) yield that

$$\varphi(h) = \begin{cases} \dfrac{\gamma_1^\nu(h) - \gamma_2^\nu(h)}{\gamma_1(h) - \gamma_2(h)}, & \gamma_1(h) \neq \gamma_2(h) \\ \nu\gamma_1^{\nu-1}(h), & \text{otherwise} \end{cases},$$

is a positive definite function for $\nu \in (0, 1]$ and any two non-negative, conditionally negative definite functions γ_1 and γ_2 where at least one of them is strictly positive. Equation 9.111 in [42] and example 2.4 yield that

$$F(\alpha; \beta; \delta; -\gamma) \tag{2.21}$$

is a positive definite function for $\alpha > 0, \delta > \beta > 0$ and any variogram γ. Here, F is the hypergeometric function, see Sect. 9.1 in [42]. Similarly, $F(\alpha; \beta; \delta; C(x, y))$ is a covariance function for any covariance function C with $C(x, y) < 1$ for all $x, y \in \mathbb{R}^d$, if $(\alpha + k)(\beta + k)/(\delta + k) \geq 0$ for all $k \in \mathbb{N}$.

Furthermore,

$$\varphi(h) = \begin{cases} \dfrac{f(\gamma(h))}{\gamma(h)}, & \gamma(h) \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

is a positive definite function for $f(z) = \log(1+z)$, $f(z) = \arctan(z)$, $f = \log(z + \sqrt{z^2+1})$ and any variogram γ, see equations 9.121.6, 9.121.27 and 9.121.28 in [42], respectively. See also example 2.5.

One more example is

$$\varphi(h) = \int_0^\infty e^{-s^2\gamma_1(h)} e^{-\gamma_2(h)/s^2} e^{-\beta s^2} ds = \frac{\sqrt{\pi}}{2\sqrt{\beta+\gamma_1(h)}} \exp\{-2\sqrt{(\beta+\gamma_1(h))\gamma_2(h)}\},$$

where γ_1 and γ_2 are variograms and $\beta > 0$, cf. equation 3.325 in [42] and [22]. Hence,

$$\varphi(h) = (\beta + \gamma_1(h))^{-1/2} \tilde{\varphi}\left(\sqrt{(\beta + \gamma_1(h))\gamma_2(h)}\right)$$

is a positive definite function for any bounded, completely monotone function $\tilde{\varphi}$.

2.7 Turning Bands Operator

The turning bands method, introduced by [67], see also [49], allows for the simulation of a stationary random field using a projection technique onto one-dimensional spaces. In almost all applications, the field is assumed to be isotropic and the dimension d is less than or equal to 3.

The turning bands method is based on the following idea. Let s be an arbitrary fixed orientation in $\mathbb{R}^d$ and Z_s a random field in $\mathbb{R}^d$ that is constant on hyperplanes perpendicular to s. Assume that the random process Y along direction s is stationary. Then Z_s is stationary, but not isotropic, except for the trivial case that Y is constant for any realization. An isotropic random field is obtained if we replace s by a random unit vector S that is uniformly distributed on the $(n-1)$-dimensional sphere $\mathscr{S}_{n-1}$ and that is independent of Y,

$$Z_S(x) = Y(\langle x, S\rangle), \qquad h \in \mathbb{R}^d.$$

Let $C_1(x, y) = \varphi_1(x - y)$ be the covariance function of Y. Then, the covariance function $C(x, y) = \varphi(x - y)$ of Z_S is given by

$$\varphi(h) = \mathbb{E}\,\varphi_1(\langle h, S\rangle) = \int_{\mathscr{S}_{n-1}} \varphi_1(\langle h, s\rangle)\pi(ds) = \int_{\mathscr{S}_{n-1}} \varphi_1(\|h\|\langle e, s\rangle)\pi(ds)$$

where π is the uniform probability measure on $\mathscr{S}_{n-1}$ and $e \in \mathbb{R}^d$ denotes any fixed unit vector. Hence, C is rotation invariant, i.e., $C(x, y) = \tilde{\varphi}(\|x - y\|)$ for some function $\tilde{\varphi} : [0, \infty) \to \mathbb{R}$. [67] showed the following relation between $\tilde{\varphi}$ and φ_1:

$$\varphi_1(r) = \begin{cases} \dfrac{d}{dr}[r\tilde{\varphi}(r)], & d = 3 \\ \dfrac{d}{dr}\displaystyle\int_0^r \frac{s\tilde{\varphi}(s)}{\sqrt{r^2 - s^2}} ds, & d = 2 \end{cases}, \qquad r > 0. \tag{2.22}$$

In fact, relation (2.22) holds reversely for any continuous positive definite function $\tilde{\varphi}(\|\cdot\|)$ on $\mathbb{R}^d$, $d = 2$ and 3, respectively [32]. The mapping which assigns $\tilde{\varphi}_1$ to $\tilde{\varphi}$ is called the *turning bands operator*. See Fig. 2.4 for an illustration of the turning bands method. In Sect. 7.4.2 of [15] the case of a general dimension $d \in \mathbb{N}$ is considered.

Note that the continuity assumption is equivalent to the assumption that C has no nugget effect [38] and that C is at least m times differentiable away from the origin for m the largest integer less than or equal to $(d-1)/2$ [32].

An approximation to a Gaussian random field is again obtained through the central limit theorem:

$$Z(x) = n^{-1/2} \sum_{i=1}^{n} Y_i(\langle x, S_i\rangle).$$

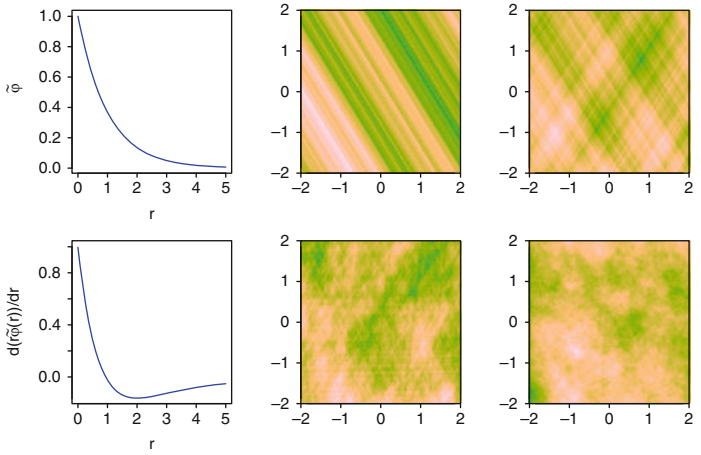

Fig. 2.4 Recentred and renormalised superpositions of 1, 10, 100, and 1000 simulations of an additive Boolean model with radius $r = 1/2$ of the disks (top left, top right, bottom left, bottom right)

Here, $Y_i \sim Y$, $i = 1, \ldots, n$, and $S_i \sim S$, $i = 1, \ldots, n$, are all independent. The number of independent copies k that are needed is about 60 for $d = 2$ and 500 for $d = 3$ [30], see also [55]. The simulation of the random field Y is performed on a grid for example by methods described in Sect. 2.9, and the closest grid point to the left, say, is taken as an approximation for $\langle x, S \rangle$.

Remark 2.9. Closed solutions for the Abel integral (2.22) in the case $d = 2$ are rare [29]. Hence, the covariance function on the line must be evaluated numerically, using the following more convenient form if $r\tilde{\varphi}(r)$ is differentiable:

$$\tilde{\varphi}_1(r) = \frac{d}{dr} \int_0^1 r\tilde{\varphi}(r\sqrt{1-s^2})ds = \int_0^1 \frac{d}{dr} r\tilde{\varphi}(r\sqrt{1-s^2})ds. \qquad (2.23)$$

Alternatively, if $\tilde{\varphi}(\|\cdot\|)$ is a positive definite function also in $\mathbb{R}^3$, the space $\mathbb{R}^2$ can be considered as a hyperplane in $\mathbb{R}^3$ and the simulation is performed in $\mathbb{R}^3$.

Remark 2.10. In practice, one should not use random directions S_i in the two-dimensional turning bands method. Instead, equal angles between the lines should be taken. By choosing the direction of the very first line purely random, isotropy is still guaranteed from a theoretical point of view.

In dimension 3 or higher, a deterministic point pattern of equally spaced locations does not exist for an arbitrary number of points. Therefore, the directions are usually chosen randomly. A random direction S in $\mathbb{R}^3$ is obtained by

$$(\sqrt{1-V^2}\cos U, \sqrt{1-V^2}\sin U, V),$$

where $U \sim U[0, 2\pi]$ is independent of $V \sim U[0, 1]$, see [26], for instance.

Remark 2.11. [32] generalizes the turning bands operator in the following way. Let $\tilde{\varphi}(\|\cdot\|)$ be a positive definite function on $\mathbb{R}^d$ and

$$\tilde{\varphi}_{d-2}(r) = \tilde{\varphi}(r) - \frac{r}{d-2}\tilde{\varphi}'(r), \qquad d-2 \geq 1.$$

Then $\tilde{\varphi}_{d-2}(\|\cdot\|)$ is a positive definite function in $\mathbb{R}^{d-2}$, and vice versa.

2.8 *Montée*

Apart from the turning bands operator, further operators transform between sets of positive definite functions by means of derivations or integrals.

For instance, the ith second partial derivative $\partial^2\varphi(h)/(\partial h_i)^2$ of a positive definite function φ is positive definite, provided it exists (e.g., [81]). This is proved by considering the covariance function of the ith partial derivative of a random field corresponding to φ.

[71] show that, if $\varphi(h) = \tilde{\varphi}(\|h\|)$ is a positive definite function in $\mathbb{R}^d$, then $\varphi_1(h) = \tilde{\varphi}_1(\|h\|)$ with $\tilde{\varphi}_1(r) = d\tilde{\varphi}(\sqrt{r})/dr$ is a positive definite function in $\mathbb{R}^{d-2}$.

Here, the *montée*, and its inverse, the *descente*, are considered. See [101] for a unified approach to the turning bands operator and the *montée*.

Let $Z(x_1, x_2)$ be a random field on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with covariance function C and $C((x_1, x_2), (y_1, y_2)) = C((x_1, x_2 - y_2), (y_1, 0))$. Let

$$Y_M(x_1) = \frac{1}{(2M)^{d_2/2}} \int_{[-M,M]^{d_2}} Z(x_1, x_2)dx_2, \qquad x_1 \in \mathbb{R}^{d_1}.$$

Then, the covariance function C_M of the random field Y_M yields

$$\begin{aligned} C_M(x_1, y_1) &= \frac{1}{(2M)^{d_2}} \int_{[-M,M]^{d_2}} \int_{[-M,M]^{d_2}} C((x_1, x_2), (y_1, y_2))dx_2 dy_2 \\ &\to \int_{\mathbb{R}^{d_2}} C((x_1, h), (y_1, 0))dh \qquad (M \to \infty). \end{aligned}$$

This transformation of the covariance functions is called *montée* [64]. If $C(x, y) = \tilde{\varphi}(\|x - y\|)$ is motion invariant, then $C_M(x, y) \to \tilde{\varphi}_{d_1}(\|x - y\|)$ with

$$\tilde{\varphi}_{d_1}(r) = \begin{cases} 2\displaystyle\int_0^\infty \tilde{\varphi}(\sqrt{r^2 + s^2})ds, & d_2 = 1 \\ 2\pi \displaystyle\int_r^\infty s\tilde{\varphi}(s)ds, & d_2 = 2 \end{cases}.$$

That is, $\tilde{\varphi}_{d_1}(\|x-y\|)$ is a positive definite function in $\mathbb{R}^{d_1}$. Reversely, let $\varphi(h) = \tilde{\varphi}(\|h\|)$ be a positive definite function in $\mathbb{R}^d$ and assume that $\tilde{\varphi}''(0)$ exists. Then the *descente* is given by

$$D\tilde{\varphi}(r) = \begin{cases} 1, & r = 0 \\ \tilde{\varphi}'(r)/(r\tilde{\varphi}''(0)), & r > 0 \end{cases},$$

and $\varphi(h) = D\tilde{\varphi}(\|h\|)$ is a positive definite function in $\mathbb{R}^{d+2}$ [34].

[31], see also [97] and [12], uses the *montée* to construct classes of differentiable covariance functions with compact support from the function

$$\tilde{\varphi}(r) = (1-r^b)^a 1_{[0,1]}(r).$$

If $b = 1$, then the function $\tilde{\varphi}(\|h\|)$ is positive definite if and only if $a \geq (d+1)/2$ [40]. For instance, $\varphi(h) = \tilde{\varphi}(\|h\|)$ is a positive definite function in $\mathbb{R}^d$ for

$$\tilde{\varphi}(r) = (1 + (\nu+2)r + 3^{-1}[(\nu+2)^2 - 1]r^2)(1-r)^{\nu+2} 1_{[0,1]}(r)$$

and $\nu \geq (d+5)/2$.

[80] and [79] extend the *montée* by considering integrations of real-valued order. See [45] for a further extension of the Wendland-Gneiting functions. [70] derive vector-valued covariance functions with comport support.

2.9 General Simulation Methods

In the following, widely used simulation methods are presented that are not immediately related to construction methods of covariance functions and variograms.

2.9.1 Simulation of a Multivariate Gaussian Vector

Let Y be an n-vector of independent Gaussian random variables with zero expectation and unit variance, and $D = D_0 D_0^\top$ be any positive semi-definite $n \times n$-matrix. Let

$$X \sim D_0 Y. \tag{2.24}$$

Then X has a multivariate, centred Gaussian distribution with covariance matrix D. Of course, this basic fact can also be used to simulate from stationary or non-stationary random fields, defining $D = (C(x_i, x_j))_{i,j=1,\ldots,n}$. The method has its numerical limitation at about $n = 10^4$ for general matrices.

2.9.2 Circulant Embedding

The circulant embedding method allows to simulate a stationary random field on a grid which is equally spaced in each direction. The idea is to expand the covariance matrix to a circulant matrix, i.e. to simulate from a torus. If this is feasible, the square root of the expanded matrix can be calculated using the Fast Fourier Transform. This approach was independently published by [23] and [13, 99]. [99] show that such an expansion is always possible if the covariance function has compact support. The algorithm is then exact in principle. In case negative eigenvalues appear in the expanded matrix, [99] suggest an approximation by putting them to zero. However, this can lead to deficient simulation results.

If n is the number of grid points and d the dimension, the number of flops is of order $2^d n \log(2^d n)$, hence the simulation method is very fast unless the dimension d is high.

Extensions to conditional simulation, to arbitrary locations [24], and to multivariate random fields [14] exist.

Further extensions are the intrinsic circulant embedding and the cut-off circulant embedding [39, 91]. The idea is to replace a given covariance function by a covariance function that equals or essentially equals the required covariance on the given finite grid, but has finite range.

2.9.3 Approximations Through Markovian Fields

In a space-time setup, a field might be simulated on a few spatial points at arbitrary locations, but at many instances in time on a grid. Instead of simulating all variables at once, (approximating) Markov fields can be used in the temporal direction, using a temporal neighbourhood of k instances. Namely, for each instance, Gaussian variables are simulated simultaneously for all locations, conditioned on the previous k instances and all locations.

[77] rigorously suggest to approximate Gaussian random fields through Markov fields with a huge increase in speed for the simulations. In a recent paper, [56] relate the Markov random fields to partial differential equations.

2.10 Space-Time Models

A current, important task is to find covariance functions that are useful for modelling space-time data. In the following, let d be the dimension in space. Mathematically, the set of space-time covariance functions cannot be distinguished from the set of covariance functions in $\mathbb{R}^{d+1}$. However, the sets of those covariance functions that are of interest in practice differ. In the purely spatial context, an isotropic random

field constitutes the standard model. In contrast, the temporal development of a process differs in most cases from the spatial development, leading to anisotropies between space and time. For example, geometrical anisotropy matrices A, see remark 2.1, that have the form

$$A = \begin{pmatrix} A_0 & -v \\ 0 & s \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}$$

connect space and time through the vector $v \in \mathbb{R}^d$. The latter can be interpreted, for instance, as wind speed in a meteorological context [43]. The matrix $A_0 \in \mathbb{R}^{d\times d}$ gives the purely spatial anisotropy and $s > 0$ is a scaling factor for the temporal axis.

To simulate space-time random fields, all the approaches presented in the previous sections can be used if they are appropriate. For example, circulant embedding will be useful if the space-time data lie on a grid. In the following, some additional, specific methods are presented.

2.10.1 Separable Models

The simplest class of anisotropic space-time models are separable models. By definition, a separable model has one of the following two forms

$$C((x,t),(y,s)) = C_S(x,y)+C_T(t,s) \quad \text{or} \quad C((x,t),(y,s)) = C_S(x,y)C_T(t,s),$$

where C_S is a covariance function in $\mathbb{R}^d$ and C_T is a covariance function in $\mathbb{R}$ [75]. All other models are called non-separable. It is easy to see from the results in Sect. 2.2 that separable models are covariance functions. A variogram is called separable if

$$\gamma(h,u) = \gamma_S(h) + \gamma_T(u)$$

for two variograms $\gamma_S(h)$ and $\gamma_T(u)$ in $\mathbb{R}^d$ and $\mathbb{R}$, respectively. Products of variograms should not be considered, cf. remark 2.6. Random fields with separable covariance function can easily be simulated. Namely, a spatial random field with covariance C_S that is constant in time is added (or multiplied, respectively) to a temporal random process with auto-covariance C_T that is independent of the former and is constant in space. The obtained field is not Gaussian and an approximation can be obtained through the central limit theorem. Although separable models are quite appealing, they have practical disadvantages [19, 54, 76] and theoretical disadvantages [92].

Many non-separable models given in the literature are based on separable models and general transformations of covariance functions and variograms as presented in the preceding sections. An example that refers to the models discussed in Sect. 2.5

is $\varphi(h,u) = (1 + |h|^\nu + |u|^\lambda)^{-\delta}$, cf. [21]. The function φ is positive definite on $\mathbb{R}^d \times \mathbb{R}^{d'}$ for any dimensions d and d', if $\nu, \lambda \in [0,2]$ and $\delta > 0$. Another example is $\varphi(h,u) = [1 - \varphi_S(h)\varphi_T(u)]^{-\alpha}$, where $\alpha > 0$ and $\varphi_S(0)\varphi_T(0) < 1$ [57].

Further models are obtained by means of scale mixtures of separable models. For instance,

$$C(h,u) = \int_0^\infty e^{-s\gamma(h)} \cos(s|u|) ds = \frac{\gamma(h)}{\gamma(h)^2 + |u|^2}$$

is a covariance model in $\mathbb{R}^d \times \mathbb{R}$ for any strictly positive, conditionally negative definite function γ on $\mathbb{R}^d$, cf. [22].

2.10.2 Gneiting's Class

[35] has introduced an important class of space-time covariance functions generalizing the findings in [19]. Let $\tilde{\varphi}_S(r)$, $r \geq 0$, be a bounded, completely monotone function and

$$\varphi(h,u) = \gamma(u)^{-d/2}\, \tilde{\varphi}_S\left(\|h\|^2/\gamma(u)\right), \qquad (h,u) \in \mathbb{R}^d \times \mathbb{R}.$$

[35] shows that φ is a positive definite function if $\gamma(u) = \psi(|u|^2)$ for some strictly positive Bernstein function ψ. [100] show that φ is a positive definite function if and only if γ is a strictly positive, conditionally negative definite function. Note that Gneiting's model is fully symmetric [35], i.e. $C((x,t),(y,s)) = C((x,-t),(y,-s))$, restricting its ability to model correlations between space and time.

[86] generalizes Gneiting's model towards models that are not fully symmetric, using the fact that $\exp(-u^2\gamma(h))$ is, for fixed u, a positive definite function in h, and, for fixed h, the spectral density of the "Gaussian" model.

Remark 2.12. The ambivalency that a function is a positive definite function in one argument and a spectral density in the other has been used previously by [94] considering the function $[c_1(a_1^2+\|h\|^2)^{\alpha_1} + c_2(a_2^2+\|u\|^2)^{\alpha_2}]^{-\nu}$. Here, $h \in \mathbb{R}^{d_1}$, $u \in \mathbb{R}^{d_2}$, $a_1^2+a_2^2 > 0$ and $c_1, c_2, \nu_1, \nu_2, \alpha_1, \alpha_2 > 0$, such that $d_1/(\alpha_1\nu) + d_2/(\alpha_2\nu) < 2$. If $\alpha_1 = 1$, then the corresponding positive definite function is given by

$$\varphi(h,u) = \frac{\pi^{d_2/2} W_{\nu - d/2}(f(\|u\|^2)\|h\|)}{2^{\nu - d_2/2 - 1} c_2^\nu \Gamma(\nu) f(\|u\|^2)^{2\nu - d_2}}.$$

The function f equals $f(s) = (a_2^2 + c_1 c_2^{-1}(a_1^2+s)^{\alpha_1})^{1/2}$, $s \geq 0$, and W_ν denotes the Whittle-Matérn model with parameter ν. See [58] and [94] for further, sophisticated models, and [93] for non-stationary covariance functions.

2.10.3 Turning Layers

Space-time data typically consist of longer, regularly measured time series given at several arbitrary locations in $\mathbb{R}^d$. The turning layers method respects this fact and is applicable for fully symmetric models that are isotropic in space. As for the turning bands method, a non-ergodic random field that is isotropic in space is obtained if a random field with translation invariant covariance function C_1 is simulated on a plane where one axis has a random direction in space and the other axis equals the time axis. The random field in $\mathbb{R}^d \times \mathbb{R}$ is constant in perpendicular direction to the plane, cf. [53]. Denote the covariance function of the latter by C. Similar to the derivation of the turning bands relation we obtain a reverse formula for $C_1((x_1, t), (y_1, s)) = \tilde{\varphi}_1(|x_1 - y_1|, |t - s|)$ given $C((x, t), (y, s)) = \tilde{\varphi}(\|x - y\|, |t - s|)$:

$$\tilde{\varphi}_1(r,t) = \begin{cases} \dfrac{\partial}{\partial r}[r\tilde{\varphi}(r,t)], & d = 3 \\ \dfrac{\partial}{\partial r}\displaystyle\int_0^r \frac{s\tilde{\varphi}(s,t)}{\sqrt{r^2 - s^2}} ds, & d = 2 \end{cases}. \tag{2.25}$$

An approximation to a Gaussian random field is obtained through the central limit theorem as in the case of the turning bands method. A realization on the plane might be obtained by using circulant embedding, see Sect. 2.9.2. The turning layers have the advantages of being an exact method in the temporal direction at any fixed location. However, it exhibits the usual approximation error of the turning bands method in space.

Remark 2.13. Assume that, for some functions $f, g : \mathbb{R} \to \mathbb{R}$ and $\tilde{\varphi}_S : [0, \infty) \to \mathbb{R}$, the function $\varphi(r, t)$ is of the form $g(t)\tilde{\varphi}_S(rf(t))$, as it is the case for the Gneiting class. Let $\tilde{\varphi}_{1,S}$ be the function obtained for $\tilde{\varphi}_S$ through (2.22) for $d = 2, 3$. Then we get

$$\tilde{\varphi}_1(r,t) = g(t)\varphi_{1,S}(rf(t)),$$

assuming that equality (2.23) holds if $d = 2$.

Naturally, the turning layers can be generalized to simulate random fields on $\mathbb{R}^d \times \mathbb{R}^n$, $n \geq 1$, that are isotropic in both components. Namely, the two-dimensional random field Y can be replaced by a higher dimensional one, or the turning bands principle can be applied also to the second component of Y.

2.10.4 Spectral Turning Layers

A variant of the turning layers that corresponds to the spectral turning bands is useful for covariance functions of the form

$$C(h,t) = \mathbb{E}\,\tilde{\varphi}(\|x - Vt\|^2). \tag{2.26}$$

Here, $\tilde{\varphi}$ is a bounded, completely monotone function and V is a d-dimensional random vector that might be interpreted as wind speed in a meteorological context [16]. A corresponding random field Z is obtained as

$$Z(x,t) = \sqrt{2F([0,\infty))}\cos(A\langle S, x - Vt\rangle + \Phi)$$

where $A \sim F/F([0,\infty))$, F is the radial spectral measure for $\tilde{\varphi}$ given by (2.9), $\Phi \sim U[0,2\pi)$, and $S \sim U\mathscr{S}_{d-1}$ is uniformly distributed on the $(d-1)$-dimensional sphere $\mathscr{S}_{d-1}$. All the random variables are independent. We call the method spectral turning layers.

Let $\tilde{\varphi}$ be a bounded, completely monotone function and $V \sim \mathscr{N}(\mu, M/2)$ for some covariance matrix M. [86] shows that C has a closed form,

$$C(h,t) = \frac{1}{\sqrt{\mathbf{1} + t^2 M}}\tilde{\varphi}\left((h - t\mu)^\top(\mathbf{1} + t^2 M)^{-1}(h - t\mu)\right)$$

which is fully symmetric if and only if $\mu = 0$.

2.10.5 Models Related to PDEs

A challenging problem is to find closed-form covariance models that refer to solutions of physical equations. Let B be the random orthogonal measure on $\mathbb{R}^2$ such that $B(I \times J) \sim \mathscr{N}(0, |I||J|)$ for any bounded intervals $I, J \subset \mathbb{R}$. [48] show that the solution of

$$\left(\frac{\partial^2}{\partial t^2} - a\frac{\partial}{\partial x} - b^2\right) Y(x,t)dx = B(d(x,t)), \qquad x,t \in \mathbb{R},$$

has covariance function

$$C(h,u) = \frac{1}{2}\left\{e^{-b|u|}\mathrm{erfc}\left(\frac{2b|h| - c|u|}{2\sqrt{c|h|}}\right) + e^{b|u|}\mathrm{erfc}\left(\frac{2b|h| + c|u|}{2\sqrt{c|h|}}\right)\right\}, \quad h,u \geq 0,$$

see also [53] and the references therein. [58] generalizes this covariance function by showing that $|h|$ on the right hand side can be replaced by $\sqrt{\gamma(h)}$ for any variogram γ.

2.11 Multivariate Models

A commonly used model for a multivariate process $Z = (Z_1, \ldots, Z_n)$ is the so-called linear model of coregionalization [41], where each component Z_j is a

linear combination $\sum_{i=1}^{K} a_{ji} Y_i$ of independent, latent processes Y_i. Assume Y_i has covariance function C_i. Then the matrix valued covariance function of Z,

$$C_{ij}(x, y) = \operatorname{cov}(Z_i(x), Z_j(y)), \qquad i, j = 1, \ldots, n, \quad x, y \in \mathbb{R}^d,$$

equals $ACA^\top$ with $A = (a_{ij})_{j=1,\ldots,n;i=1,\ldots,K}$ and $C = \operatorname{diag}(C_1, \ldots, C_K)$.

Except for some further special constructions, see [96] and [86] for instance, parametrized classes have been rare.

Recently, [37] introduced an extension of the Whittle-Matérn model W_ν to the multivariate case. In the bivariate case, they show that $C_{ij}(h) = (b_{ij} W_{\nu_{ij}}(a_{ij} h))_{i,j=1,2}$ with $C_{ij} = C_{ji}$, $\nu_{ij} > 0$, $i, j = 1, 2$ and $b_{ii} \geq 0$, $i = 1, 2$, is a matrix-valued positiv definite function if and only if

$$b_{ij}^2 \leq b_{11} b_{22} \frac{\Gamma(\nu_{11} + \frac{d}{2}) \Gamma(\nu_{22} + \frac{d}{2}) \Gamma(\nu_{12})^2}{\Gamma(\nu_{11}) \Gamma(\nu_{22}) \Gamma(\nu_{12} + \frac{d}{2})^2} \frac{a_{11}^{2\nu_{11}} a_2^{2\nu_{22}}}{a_{12}^{4\nu_{12}}} \inf_{t \geq 0} \frac{(a_{12}^2 + t^2)^{2\nu_{12} + d}}{\prod_{i=1}^{2} (a_{ii}^2 + t^2)^{\nu_{ii} + d/2}}.$$

[70] derive multivariate models with compact support. [82] give both necessary and sufficient conditions such that a matrix-valued covariance function is divergence free or curl free. They also show that this property is inherited by the corresponding Gaussian random field.

[14] present a multivariate version of the circulant embedding method, Sect. 2.9.2. Nonetheless, further methods for simulating multivariate models need to be developed.

2.12 Exercises

In the following, we give examples of covariance functions given in the literature that can be derived from the assertions presented in Sects. 2.2–2.8.

Exercise 2.1. [72] show that certain quasi-arithmetic means of completely monotone functions lead to positive definite functions. They give three examples for classes of positive definite functions. Show the positive definiteness for two of their examples:

1. Gumbel-Hougard family

$$\varphi(h_1, h_2) = \exp(-(\|h_1\|^{\rho_1} + \|h_2\|^{\rho_2})^\beta)$$

for any $\beta \in [0, 1]$, $\rho_i \in [0, 2]$, $h_i \in \mathbb{R}^{d_i}$, $d_i \in \mathbb{N}$, $i = 1, 2$.

2. Clayton family

$$\varphi(h_1, h_2) = [(1 + \|h_1\|)^{\rho_1} + (1 + \|h_2\|)^{\rho_2}]^{-\beta}$$

for any $\beta > 0$, $\rho_i \in [0, 1]$, $h_i \in \mathbb{R}^{d_i}$, $d_i \in \mathbb{N}$, $i = 1, 2$.

Hint: show that γ^{-1} is a covariance function for any strictly positive, conditionally negative definite function γ, considering a suitable mixture of the functions $\exp(-s\gamma)$, $s \geq 0$.

Exercise 2.2. [73] introduce the Dagum family

$$\gamma(h) = (1 + \|h\|^{-\beta})^{-\alpha}$$

and show that γ is a variogram in $\mathbb{R}^3$ if $\beta < (7-\alpha)/(1+5\alpha)$ and $\alpha < 7$. [6] present conditions so that the function $r \mapsto (1 + r^{-\beta})^{-\alpha}$ is completely monotone. Show that the Dagum family yields a variogram on $\mathbb{R}^d$ for $d \in \mathbb{N}$, $\alpha \in (0, 1]$ and $\beta \in (0, 2]$.

Hint: show that

$$\frac{\psi(0) - \psi(h)}{1 + \psi(0) - \psi(h)}$$

is a variogram for any positive definite function ψ, and conclude that $\gamma_0/(1+\gamma_0)$ is a variogram for any variogram γ_0. See [63] for an alternative proof.

Exercise 2.3. Let Z be an intrinsically stationary random field on $\mathbb{R}^d$ with variogram γ and z be fixed. Show that the covariance function of Y with $Y(x) = Z(x+z) - Z(x)$ equals

$$C(x, y) = \gamma(x - y + z) + \gamma(x - y - z) - 2\gamma(x - y), \qquad x, y \in \mathbb{R}^d$$

and conclude that

1. $f(h, z) = 2\gamma(z) + 2\gamma(h) - \gamma(h + z) - \gamma(h - z)$ is a variogram for any fixed z. See, for instance, Lemma 17 in [74] and Lemma 1 in [59] for proofs given in the literature.

   Show further that, although $f(z, h) = f(h, z)$, the function f is not a variogram in (h, z), in general. To this end, consider $\gamma(h) = |h|$ on $\mathbb{R}^1$ and verify that (2.2) is not satisfied.
2. The function $\varphi(h) = 0.5(\|h + 1\|^\alpha - 2\|h\|^\alpha + \|h - 1\|^\alpha)$ is positive definite for $\alpha \in (0, 2]$. The corresponding random field is called fractional Gaussian noise if $d = 1$.

Acknowledgements The author is grateful to Sebastian Engelke, Alexander Malinowski, Marco Oesting, Robert Schaback, and Kirstin Strokorb for valuable hints and comments.

References

1. Armstrong, M.: Positive definiteness is not enough. Math. Geol. **24**, 135–143 (1992)
2. Ballani, F., Kabluchko, S., Schlather, M.: Random marked sets. ArXiv **0903.2388** (2009)
3. Barndorff-Nielsen, O.: Hyperbolic distributions and distributions on hyperbolae. Scand. J. Statist. **5**, 151–157 (1978)

4. Berg, C.: Hunt convolution kernels which are continuous singular with respect to Haar measure. In Probability Measures on Groups. **706** of **Lecture Notes in Mathematics**, 10–21. Springer, Berlin (1979)
5. Berg, C.: Stieltjes-Pick-Bernstein-Schoenberg and their connection to complete monotonicity. In E. Porcu and J. Mateu (Eds.), Positive Definite Functions: From Schoenberg to Space-Time Challenges. Castellon de la Plana: Citeseer (2008)
6. Berg, C., Mateu, J., Porcu, E.: The Dagum family of isotropic correlation functions. Bernoulli **14**(4), 1134–1149 (2008)
7. Berlinet, A., Thomas-Agnan, C.: Reproducing Kernel Hilbert Spaces in Probability and Statistics. Kluwer, Boston (2004)
8. Berschneider, G., Sasvári, Z.: On a theorem of Karhunen and related moment problems and quadrature formulae. Proceedings IWOTA 2010. To appear
9. Billingsley, P.: Probability and Measure (third ed.). John Wiley & Sons, New York, Chichester, Brisbane (1995)
10. Bochner, S.: Vorlesungen über Fouriersche Integrale. Akademische Verlagsgesellschaft, Leipzig (1932)
11. Bochner, S.: Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse. Math. Ann. **108**, 378–410 (1933)
12. Buhmann, M.: Radial basis functions: theory and implementations. University Press, Cambridge (2003)
13. Chan, G., Wood, A.: An algorithm for simulating stationary Gaussian random fields. J. R. Statist. Soc., Ser. C **46**, 171–181 (1997)
14. Chan, G. A. Wood.: Simulation of stationary gaussian vector fields. Statistics and Computing **9**, 265 – 268 (1999)
15. Chilès, J.-P., Delfiner, P.: Geostatistics. Modeling Spatial Uncertainty. John Wiley & Sons, New York, Chichester (1999)
16. Cox, D., Isham, V.: A simple spatial-temporal model of rainfall. Proc. R. Soc. Lond. Ser. A, Math. Phys. Eng. Sci. **415**, 317–328 (1988)
17. Cox, D., Miller, H.: The Theory of Stochastic Processes. Chapman & Hall, London (1965)
18. Cressie, N.: Statistics for Spatial Data. John Wiley & Sons, New York, Chichester (1993)
19. Cressie, N., Huang, H.C.: Classes of nonseparable, spatio-temporal stationary covariance functions. J. Am. Stat. Assoc. **94**, 1330–1340 (1999)
20. Daley, D. Vere-Jones, D.: An Introduction to the Theory of Point Processes. Springer, New York, Berlin, Heidelberg (1988)
21. de Cesare, L., Myers, D., Posa, D.: Fortran programs for space-time modeling. Comput. Geosci. **28**, 205–212 (2002)
22. de Iaco, S., D. Myers, and D. Posa (2002). Nonseparable space-time covariance models: some parameteric families. Math. Geol. **34**, 23–42.
23. Dietrich, C., Newsam, G.: A fast and exact method for multidimensional Gaussian stochastic simulations. Water Resour. Res **29**, 2861–2869 (1993)
24. Dietrich, C., Newsam, G.: A fast and exact method for multidimensional Gaussian stochastic simulations: Extensions to realizations conditioned on direct and indirect measurement. Water Resour. Res **32**, 1643–1652 (1996)
25. Fasen, V., Klüppelberg, C., Schlather, M.: High-level dependence in time series models. Extremes **13**, 1–33 (2010)
26. Fouquet, C. d.: Simulation conditionnelle de fonctions aléatoires: cas gaussien stationnaire et schéma linéaire. Technical Report Cours C-151, Ecole des Mines de Paris (1993)
27. Genton, M.: Classes of kernels for machine learning: a statistics perspective. J. Mach. Learn. Res. **2**, 299–312 (2001)
28. Gneiting, T.: Normal scale mixtures and dual probability densities. J. Stat. Comput. Simul. **59**, 375–384 (1997)
29. Gneiting, T.: Closed form solutions of the two-dimensional turning bands equation. Math. Geol. **30**, 379–390 (1998a)

30. Gneiting, T.: The correlation bias for two-dimensional simulations by turning bands. Math. Geol. **31**, 195–211 (1998b).
31. Gneiting, T.: Correlation functions for atmospheric data analysis. Q. J. Roy. Meteor. Soc. **125**, 2449–2464 (1999a)
32. Gneiting, T.: On the derivatives of radial positive definite functions. J. Math. Anal. Appl. **236**, 86–93 (1999b)
33. Gneiting, T.: Radial positive definite functions generated by Euclid's hat. J. Multivariate Anal. **69**, 88–119 (1999c)
34. Gneiting, T.: Compactly supported correlation functions. J. Multivariate Anal. **83**, 493–508 (2002a)
35. Gneiting, T.: Nonseparable, stationary covariance functions for space-time data. J. Amer. Statist. Assoc. **97**, 590–600 (2002b)
36. Gneiting, T., Genton, M., Guttorp, P.: Geostatistical space-time models, stationarity, separability and full symmetry. In B. Finkenstadt, L. Held, and V. Isham (Eds.), Statistical Methods for Spatio-Temporal Systems, Chapter 4. Chapman & Hall/CRC, Boca Raton (2007)
37. Gneiting, T., Kleiber, W., Schlather, S.: Matérn cross-covariance functions for multivariate random fields. J. Amer. Statist. Assoc. **105**, 1167–1177 (2011)
38. Gneiting, T., Sasvári, Z.: The characterization problem for isotropic covariance functions. Math. Geol. **31**, 105–111 (1999)
39. Gneiting, T., Ševčíková, H., Percival, D., Schlather, M., Jiang, Y.: Fast and exact simulation of large Gaussian lattice systems in R^2: Exploring the limits. J. Comput. Graph. Stat. **15**, 483–501 (2006)
40. Golubov, B.: On Abel-Poisson type and Riesz mean. Anal. Math. **7**, 161–184 (1981)
41. Goulard, M., Voltz, M.: Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. Math. Geol. **24**, 269–282 (1992)
42. Gradshteyn, I., Ryzhik, I.: Table of Integrals, Series, and Products (6th ed.). Academic Press, London, (2000)
43. Gupta, V., Waymire, E.: On taylor's hypothesis and dissipation in rainfall. J. Geophys. Res. **92**, 9657–9660 (1987)
44. Guttorp, P., Gneiting, T.: On the Matern correlation family. Biometrika **93**(4), 989 (2006)
45. Hubbert, S.: Closed form representations for a class of compactly supported radial basis functions. Adv. Comput. Math.. DOI: 10.1007/s10444-011-9184-5, Online (2011)
46. Itô, K. (Ed.) Encyclopedic Dictionary of Mathematics (third ed.). MIT Press Cambridge, MA (1996)
47. Jeulin, D., Jeulin. P.: Synthesis of rough surfaces of random morphological functions. Stereo. Iugosl. **3**, 239–246 (1981)
48. Jones, R., Zhang, Y.: Models for continuous stationary space-time processes. In T. Gregoire, D. Brillinger, P.J.Diggle, E. Russek-Cohen, W. Warren, and R. Wolfinger (Eds.), Modelling Longitudinal and Spatially Correlated Data, 289–298. Springer, New York (1997)
49. Journel, A., Huijbregts, C.: Mining Geostatistics. Academic Press, London: (1978)
50. Kabluchko, Z., Schlather, M.: Ergodic properties of max-infinitely divisible processes. Stochastic Process. Appl. **120**, 281–295 (2010)
51. Kabluchko, Z., Schlather, M., de Haan, L.: Stationary max-stable fields associated to negative definite functions. Ann. Probab. **37**, 2042–2065 (2009)
52. Karhunen, K.: Zur Spektraltheorie stochastischer Prozesse. Ann. Acad. Sci. Fenn., Ser. I. A. Math. **34**, 1–7 (1946)
53. Kolovos, A., Christakos, G., Hristopulos, D., Serre, M.: Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. Adv. Water. Res. **27**, 815–830 (2004)
54. Kyriakidis, P., Journel, A.: Geostatistical space-time models: A review. Math. Geol. **31**, 651–684 (1999)
55. Lantuéjoul, C.: Geostatistical Simulation. Springer, Berlin (2002)
56. Lindgren, F., Rue, H., Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Statist. Soc., Ser. B **73**, 423–498 (2011)

57. Ma, C.: Spatio-temporal covariance functions generated by mixtures. Math. Geol. **34**, 965–975 (2002)
58. Ma, C.: Families of spatio-temporal stationary covariance models. J. Stat. Plann. Inference **116**, 489–501 (2003)
59. Ma, C.: Spatio-temporal variograms and covariance models. Adv. Appl. Probab. **37**, 706–725 (2005)
60. Ma, C.: Recent developments on the construction of spatio-temporal covariance models. Stoch. Environ. Res. Risk. Asses. **22**, 39–47 (2008)
61. Mantoglou, A.: Digital-simulation of multivariate two-dimensional and three-dimensional stochstic processes with a spectral turning bands method. Math. Geol. **19**, 129–149 (1987)
62. Matérn, B.: Spatial Variation (second ed.). Springer, Berlin, New York (1986)
63. Mateu, J., E. Porcu, Nicolis, O.: A note on decoupling of local and global behaviours for the Dagum random field. Probab. Engineer. Mech. **22**, 320–329 (2007)
64. Matheron, G.: Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux Sciences de la Nature. Masson, Paris (1965)
65. Matheron, G.: La théorie des variables régionalisées, et ses applications. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fascicule 5, Ecole des Mines de Paris (1970)
66. Matheron, G.: Leçon sur les fonctions aléatoire d'ordre 2. Technical Report C-53, Ecole des Mines de Paris (1972)
67. Matheron, G.: The intrinsic random functions and their applications. Adv. Appl. Probab. **5**, 439–468 (1973)
68. Matheron, G.: Suffit-il, pour une covariance, d'être de type positif? Sci. de la Terre, Sér. Informatique Géologique **26**, 51–66 (1987)
69. Papoulis, A.: Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York, St. Louis (1965)
70. Porcu, E., Daley, D., Buhmann,B., Bevilacqua, M.: Classes of compactly supported correlation functions for vector-valued random fields. In preparation (2011)
71. Porcu, E., Gregori, P., Mateu, J.: La descente et la montée étendues: the spatially d-anisotropic and the spatio-temporal case. Stoch. Environm. Res. Risk Assess. **21**, 683–693 (2007)
72. Porcu, E., Mateu, J., Christakos, G.: Quasi-arithmetic means of covariance functions with potential applications to space-time data. J. Multivariate Anal. **100**, 1830–1844 (2009)
73. Porcu, E., Mateu, J., Zini, A., Pini, R.: Modelling spatio-temporal data: A new variogram and covariance structure proposal. Statist. Probab. Lett. **77**(1), 83–89 (2007)
74. Porcu, E., Schilling, R.: From Schoenberg to Pick-Nevanlinna: towards a complete picture of the variogram class. Bernoulli. **17**(1), 441–455 (2011)
75. Rodríguez-Iturbe, I., Mejía, J.: The design of rainfall networks in space and time. Water Resour. Res. **10**, 713–728 (1974)
76. Rouhani, S., Myers, D.: Problems in space-time kriging of geohydrological data. Math. Geol. **22**, 611–623 (1990)
77. Rue, L., Held, H.: Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC, Boca Raton, London (2005)
78. Sasvári, Z.: Positive Definite and Definitizable Functions. Akademie Verlag, Berlin (1994)
79. Schaback, R.: The missing Wendland functions. Adv. Computat. Math. **34**, 67–81 (2011)
80. Schaback, R., Wu, Z.: Operators on radial functions. J. Computat. Appl. Math. **73**, 257–270 (1996)
81. Scheuerer, M.: A Comparison of Models and Methods for Spatial Interpolation in Statistics and Numerical Analysis. Ph. D. thesis, Univ. Göttingen (2009)
82. Scheuerer, M., Schlather, M.: Covariance models for random vector fields. Submitted (2011)
83. Schilling, R., Song, R., Vondracek, Z.: Bernstein Functions. de Gruyter, Berlin, New York (2010)
84. Schlather, M.: Introduction to positive definite functions and to unconditional simulation of random fields. Technical Report ST-99-10, Lancaster University, Lancaster (1999)
85. Schlather, M.: Models for stationary max-stable random fields. Extremes **5**, 33–44 (2002)

86. Schlather, M.: Some covariance models based on normal scale mixtures. Bernoulli **16**, 780–797 (2010)
87. Schlather, M.: RandomFields V 2.0: contributed extension package to R for the simulation of Gaussian and max-stable random fields. URL cran.r-project.org (2011)
88. Schoenberg, I.: Metric spaces and completely monotone functions. Ann. Math. **39**, 811–841 (1938)
89. Shkarofsky, I.P.: Generalized turbulence space-correlation and wave-number spectrum-function pairs. Can. J. Phys. **46**, 2133–2153 (1968)
90. Stein, M.: Interpolation of Spatial Data. Springer, Heidelberg, New York, (1999)
91. Stein, M.: Fast and exact simulation of fractional Brownian surfaces. J. Comput. Graph. Stat. **11**, 587–599 (2002a)
92. Stein, M.: The screening effect in kriging. Annals of Statistics **30**, 298–323 (2002b)
93. Stein, M.: Nonstationary spatial covariance functions. Technical report, University of Chicago, URL www.stat.uchicago.edu/cises/research/cises-tr21.pdf (2005a)
94. Stein, M.: Space-time covariance functions. J. Amer. Statist. Assoc. **100**, 310–321 (2005b)
95. Vert, J., Tsuda, K., Schölkopf, B.: A primer on kernel methods. In B. Schölkopf, K. Tsuda, and J. Vert (Eds.), Kernel Methods in Computational Biology, pp. 35–70. The MIT Press, Cambridge, MA (2004)
96. Wackernagel, H.: Multivariate Geostatistics (3rd ed.). Springer, Berlin, Heidelberg (2003)
97. Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Adv. Comput. Math. **4**, 389–396 (1995)
98. Widder, D.: What is the Laplace Transform? Amer. Math. Monthly **52**, 419–425 (1945)
99. Wood, A., Chan, G.: Simulation of stationary Gaussian processes in $[0, 1]^d$. J. Comput. Graph. Stat. **3**, 409–432 (1994)
100. Zastavnyi, V. Porcu, E. Characterization theorems for the Gneiting class space-time covariances. Bernoulli **17**(1), 456–465 (2011)
101. zu Castell, W.: Recurrence relations for radial positive definite functions. J. Math. Anal. Appl. **271**, 108–123 (2002)

Chapter 3
Geostatistics for Large Datasets

Ying Sun, Bo Li, and Marc G. Genton

Abstract We review various approaches for the geostatistical analysis of large datasets. First, we consider covariance structures that yield computational simplifications in geostatistics and briefly discuss how to test the suitability of such structures. Second, we describe the use of covariance tapering for both estimation and kriging purposes. Third, we consider likelihood approximations in both spatial and spectral domains. Fourth, we explore methods based on latent processes, such as Gaussian predictive processes and fixed rank kriging. Fifth, we describe methods based on Gaussian Markov random field approximations. Finally, we discuss multivariate extensions and open problems in this area.

3.1 Introduction

Due to the advancement of technology, massive amounts of data are often observed at a large number of spatial locations in geophysical and environmental sciences. There are many interesting aspects to discuss for the geostatistical analysis of large spatial datasets. Here we focus on computational issues, that is, how to make the geostatistical analysis of large datasets feasible or how to improve computational efficiency. This is crucial because spatial problems with modern data often overwhelm traditional implementations of spatial statistics, such as maximum likelihood estimation, Bayesian methods, and best linear unbiased prediction (kriging). In particular, large dimensional covariance matrices must be inverted. Moreover, many

Y. Sun (✉) · M.G. Genton
Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA
e-mail: sunwards@stat.tamu.edu; genton@stat.tamu.edu

B. Li
Department of Statistics, Purdue University, West Lafayette, IN 47907, USA
e-mail: boli@stat.purdue.edu

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_3,

geophysical processes are observed on a globe and it is common to have spatial data covering a large portion of the Earth. This requires special techniques to deal with large datasets observed over a sphere [27, 28]. Finally, the computational burden is aggravated in spatio-temporal settings and in multivariate situations where multiple observations occur at each location.

For instance, the exact computation of the likelihood of a Gaussian spatial random field observed at n irregularly sited locations generally requires $O(n^3)$ operations and $O(n^2)$ memory [43]. Therefore while sample sizes of $n = 1{,}000$ are no longer a challenge, $n = 1{,}000{,}000$ remains out of reach with classical procedures for even large clusters of processors. In a Bayesian framework, hierarchical models implemented through Markov Chain Monte Carlo (MCMC) methods have become especially popular for spatial modeling, given their flexibility and power to fit models that would be infeasible with classical methods. However, fitting hierarchical spatial models also involves expensive matrix operations whose computational complexity increases in a cubic order of the number of spatial locations n at every iteration of the MCMC algorithm [3], and thus here as well the computations can become problematic for large spatial datasets. Kriging, or spatial best linear unbiased prediction (BLUP), is an optimal interpolation in geostatistics. Solving the kriging equations directly requires the solution of a large linear system and involves inversion of an $n \times n$ covariance matrix $\mathbf{C}$, where $O(n^3)$ computations are required to obtain $\mathbf{C}^{-1}$ [7, 18].

Because large dataset issues often arise from the difficulty of dealing with large covariance matrices, understanding and modeling covariance structures is the key to tackle this problem. Let $\{Z(\mathbf{x}); \mathbf{x} \in \mathscr{D} \subset \mathbb{R}^d\}$, $d \geq 1$, be a random field and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be the sampling points in $\mathscr{D}$. For a second-order stationary random field, the covariance function, $C(\mathbf{k}) = \text{cov}\{Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{k})\}$, is determined only by the lag $\mathbf{k}$ but not the location $\mathbf{x}$. Here $\mathbf{x}$ denotes the location and $\mathbf{k}$ denotes the lag and they can be defined in either a purely spatial domain $\mathscr{D} = \mathscr{S}$ or a spatio-temporal domain $\mathscr{D} = \mathscr{S} \times \mathscr{T}$ depending on the nature of the data. Let $\mathbf{h}$ denote the spatial lag and u the temporal lag, that is, $\mathbf{k} = \mathbf{h}$ for spatial data indexed by $\mathbf{x} = \mathbf{s} \in \mathscr{S}$ while $\mathbf{k} = (\mathbf{h}, u)$ for spatio-temporal data indexed by $\mathbf{x} = (\mathbf{s}, t) \in \mathscr{S} \times \mathscr{T}$. Further, we can define the second-order stationary covariance function for a multivariate random field $\{\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), \ldots, Z_p(\mathbf{x}))^{\mathrm{T}}; \mathbf{x} \in \mathscr{D} \subset \mathbb{R}^d\}$ by $C_{ij}(\mathbf{k}) = \text{cov}\{Z_i(\mathbf{x}), Z_j(\mathbf{x} + \mathbf{k})\}$, for $i, j = 1, \ldots, p$, where p is the number of variables at each location. Compared to the $n \times n$ univariate covariance matrix $\mathbf{C}$ induced by n spatial locations, the size of the multivariate cross-covariance matrix is inflated to $np \times np$.

There have been several approaches to overcome this large matrix problem, such as imposing separability on covariance functions, tapering the covariance matrix, using composite likelihoods, truncating the spectral representation of a random field, modeling the realizations by a latent process with reduced dimension, and approximating the random field with a Gaussian Markov random field. One common feature implied by all these methods is to sacrifice some unimportant information in the data in order to gain computational efficiency. How to define "unimportant" distinguishes these methods. Separability ignores the interaction between different

types of covariances so that the dimension of the covariance matrices to be inverted is reduced dramatically and thus facilitates computational procedures. Covariance tapering makes use of the computational advantage of sparse matrices obtained by zeroing out the "small" values in the covariance matrix. Composite likelihood methods throw away weak correlations between observations that are far apart, and spectral methods estimate parameters by truncating spectral representations. Latent process approaches keep only the most fundamental structure and wash out the details in the random field via adopting a low rank structure on the covariance matrix, which enables to simply deal with a matrix of low dimension rather than a large covariance matrix. While sharing this dimension reduction idea, many low rank models in different forms have been developed. Besides, Gaussian Markov random fields, for which the conditional distributions only depend on nearby neighbors, lead to sparseness of the precision matrix, the inverse of the covariance matrix.

All these methods can be regarded as approximations of the underlying random field. In [48], an efficient computing algorithm was thus developed for calculating different predictive scores which are measures of the predictive performance and assess how well an approximation works. Estimators are then constructed to minimize some prediction scores.

The remainder of this chapter is organized as follows. In Sect. 3.2, we review separable covariance structures and explains how they can facilitate computational procedures for large spatial datasets. Then, in Sect. 3.3, we describe the use of covariance tapering for both maximum likelihood estimation and kriging purposes. We provide some other practical ways to evaluate likelihood functions in both spatial and spectral domains in Sect. 3.4. Next, in Sect. 3.5, we introduce different forms of low rank models for latent process approaches, including Gaussian predictive process models and fixed rank kriging, and in Sect. 3.6 we discuss approximations using Gaussian Markov random fields. We review some existing methods and extensions for multivariate spatial datasets in Sect. 3.7, and the chapter ends with a discussion in Sect. 3.8.

3.2 Separable Covariance Structures

One way to deal with the computational issue of large covariance matrices is to take advantage of some special covariance structures. A class of such covariance structures that has been used widely is that of separable covariance functions. Separability is defined with different notions depending on the context. For example, a space-time separable covariance model is defined as $C(\mathbf{h}, u) = C(\mathbf{h}, 0)C(\mathbf{0}, u)/C(\mathbf{0}, 0)$. That is, the space-time covariance function can be factored into a product of a purely spatial covariance and a purely temporal covariance. Another notion of separability, also called intrinsic model and defined only for multivariate random fields, is that $\mathbf{C}(\mathbf{k}) = \rho(\mathbf{k})\mathbf{A}$ for a spatial or spatio-temporal correlation function $\rho(\mathbf{k})$ and a $p \times p$ positive definite matrix $\mathbf{A}$. This type of separability indicates that the covariance between variables is independent of the covariance induced by spatial locations.

The aforementioned separable covariance functions can significantly reduce the dimension of the covariance matrices that need to be inverted, hence they can alleviate the computational demand. This is because separability enables the large covariance matrix to be a Kronecker product of smaller matrices and then only the inversions of those smaller matrices are required due to nice properties of Kronecker products. For instance, a spatio-temporal dataset observed at 100 spatial locations and 100 time points leads to a covariance matrix of size $10{,}000 \times 10{,}000$, for which the inversion is difficult to compute. However, employing a space-time separable covariance function decomposes this large covariance matrix into a Kronecker product of two square matrices each of size 100×100 with one being the spatial covariance matrix and the other the temporal covariance matrix. Then the inversion of the matrix of size $10{,}000 \times 10{,}000$ is reduced to the inversion of two matrices of size 100×100. Similar gains can be achieved when separable models are used in multivariate spatial or spatio-temporal data analysis.

Despite their attractive properties, separable models are not always appropriate for real data due to their lack of flexibility to allow for interactions between different types of correlations. [21] illustrated the lack of space-time separability for an Irish wind data and [6] suggested a nonseparable space-time covariance underlying a tropical wind dataset in the Pacific Ocean. Further, [32] also demonstrated the lack of multivariate separability for a trivariate pollution data over California. A variety of tests have been developed to assess the appropriateness of space-time separability. Among those, [31] proposed a unified nonparametric framework to test many different assumptions, including separability, made on the covariance structure. Their approach is based on the asymptotic normality of covariance estimators and can be easily implemented without assuming any specific marginal or joint distribution of the data other than some mild moment and mixing conditions. Later, this test was extended by [32] to assess separability for multivariate covariance functions, for which no effective and formal methods had been developed. Based on the testing framework in [31, 32], [40] proposed a self-normalization idea in place of subsampling methods to estimate the covariance of empirical covariance estimators. This new estimating method avoids the choice of the optimal block size required by the subsampling method.

In the case of lack of separability, [20] described a nearest Kronecker product approach to find separable approximations of nonseparable space-time covariance matrices. His main idea was to identify two small matrices that minimize the Frobenius norm of the difference between the original covariance matrix and the Kronecker product of those two matrices. His data example with Irish wind speeds showed that the prediction deteriorated only slightly whereas large computational savings were obtained.

Other structures of the covariance function can lead to simplifications too. For instance, a spatial random field in $\mathbb{R}^2$ with an isotropic stationary covariance function yields a symmetric block Toeplitz covariance matrix with Toeplitz blocks. Recall that a matrix is said to be of Toeplitz form if its entries are constant on each diagonal. Kriging can then be performed more efficiently with such a structured covariance matrix [50]. Stationarity of the random field can be tested with the procedure of [26] and then isotropy with the method of [22].

3.3 Tapering

The idea of tapering is to set the covariances at large distances to zero but still keep the original covariances for proximate locations. This is done in a way such that the new matrix retains the property of positive definiteness while efficient sparse matrix algorithms can be used. However, since tapering strictly zeroes out weak correlations, a natural question is whether statistical inference based on the tapered version shares the same desirable properties with the untapered exact solution. This question is answered separately in two sections: Sect. 3.3.1 focuses on properties of the maximum likelihood estimator (MLE) of the covariance parameters, and Sect. 3.3.2 discusses spatial interpolation using kriging with known covariance functions.

3.3.1 Tapering for Estimation

We consider data drawn from a zero-mean, stationary and isotropic Gaussian random field Z. Let $C(h;\boldsymbol{\theta})$ be the parametric covariance function between any two observations whose locations are apart by a distance h. The parameter vector $\boldsymbol{\theta} \in \mathbb{R}^b$ needs to be estimated from a finite number of observations, $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^{\mathrm{T}}$. Since the vector $\mathbf{Z}$ follows a multivariate normal distribution, we have the log-likelihood function for $\boldsymbol{\theta}$

$$l(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{C}(\boldsymbol{\theta})| - \frac{1}{2}\mathbf{Z}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\theta})^{-1}\mathbf{Z}, \tag{3.1}$$

where $\mathbf{C}(\boldsymbol{\theta})$ is a $n \times n$ covariance matrix with the (i,j)th element equal to $C(\|\mathbf{s}_i - \mathbf{s}_j\|;\boldsymbol{\theta})$, $i,j = 1,\ldots,n$. [30] proposed a method of covariance tapering to approximate the log-likelihood (3.1). Then, focusing on the particular case of the Matérn class of covariance functions, they illustrated the behavior of the MLE.

The tapering idea is particularly suitable if the correlations between observations far apart are negligible. This type of structure can then be modeled by a compactly supported covariance function. Let a tapering function $C_{tap}(h;\gamma)$ be an isotropic correlation function of compact support, that is, $C_{tap}(h;\gamma) = 0$ whenever $h \geq \gamma$ for some $\gamma > 0$. Denote a tapered covariance function by

$$\widetilde{C}(h;\boldsymbol{\theta},\gamma) = C(h;\boldsymbol{\theta})C_{tap}(h;\gamma), \quad h > 0. \tag{3.2}$$

The tapered covariance matrix defined by $\widetilde{\mathbf{C}}$ is a Schur (or Hadamard) product $\widetilde{\mathbf{C}}(\boldsymbol{\theta}) = \mathbf{C}(\boldsymbol{\theta}) \circ \mathbf{T}(\gamma)$, where $\mathbf{T}(\gamma)_{ij} = C_{tap}(\|\mathbf{s}_i - \mathbf{s}_j\|;\gamma)$, or $\widetilde{\mathbf{C}}_{ij} = \widetilde{C}(\|\mathbf{s}_i - \mathbf{s}_j\|;\boldsymbol{\theta},\gamma)$. The tapered covariance matrix is positive definite, since the elementwise matrix product of two covariance matrices is again a valid covariance matrix. In addition, it has high proportion of zero elements when γ is small and is, therefore, a sparse matrix. Hence it can be inverted much more efficiently than inverting a full matrix of the same dimension when evaluating the log-likelihood.

Using covariance tapering, [30] proposed two approximations of the log-likelihood in (3.1). The first approximation simply replaces the model covariance matrix $\mathbf{C}(\boldsymbol{\theta})$ by $\mathbf{C}(\boldsymbol{\theta}) \circ \mathbf{T}(\gamma)$, yielding

$$l_{1tap}(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{C}(\boldsymbol{\theta}) \circ \mathbf{T}(\gamma)| - \frac{1}{2}\mathbf{Z}^{\mathrm{T}}[\mathbf{C}(\boldsymbol{\theta}) \circ \mathbf{T}(\gamma)]^{-1}\mathbf{Z} \tag{3.3}$$

with biased score function, that is, $\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}} l_{1tap}(\boldsymbol{\theta})] \neq \mathbf{0}$. This means that there is no guarantee that the estimator that maximizes (3.3) is asymptotically unbiased. To correct the bias, the second approximation takes an estimating equations approach. First, rewrite $\mathbf{Z}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\theta})^{-1}\mathbf{Z} = \mathrm{tr}\{\widehat{\mathbf{C}}\mathbf{C}(\boldsymbol{\theta})^{-1}\}$, where $\widehat{\mathbf{C}} = \mathbf{Z}\mathbf{Z}^{\mathrm{T}}$ is the sample covariance matrix. Then replace both the model and sample covariance matrices with tapered versions yielding

$$\begin{aligned} l_{2tap}(\boldsymbol{\theta}) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{C}(\boldsymbol{\theta}) \circ \mathbf{T}(\gamma)| - \frac{1}{2}\mathrm{tr}\left\{[\widehat{\mathbf{C}} \circ \mathbf{T}(\gamma)][\mathbf{C}(\boldsymbol{\theta}) \circ \mathbf{T}(\gamma)]^{-1}\right\} \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{C}(\boldsymbol{\theta}) \circ \mathbf{T}(\gamma)| - \frac{1}{2}\mathbf{Z}^{\mathrm{T}}\left\{[\mathbf{C}(\boldsymbol{\theta}) \circ \mathbf{T}(\gamma)]^{-1} \circ \mathbf{T}(\gamma)\right\}\mathbf{Z}. \end{aligned}$$

Maximizing $l_{2tap}(\boldsymbol{\theta})$ now corresponds to solving an unbiased estimating equation for $\boldsymbol{\theta}$, that is, $\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}} l_{2tap}(\boldsymbol{\theta})] = \mathbf{0}$.

In both approximations, small values of γ correspond to more severe tapering. When $\gamma = 0$, observations are treated as independent, and as $\gamma \to \infty$, we recover the original likelihood. For the particular case of the Matérn class of covariance functions, it has been shown that the estimators maximizing the tapering approximations, such as the MLE, are strongly consistent under certain conditions.

[11] then investigated how the tapering affects the asymptotic efficiency of the MLE for parameters in the Matérn covariance function under the assumption that data are collected along a line in a bounded region. Their results imply that, under some conditions on the taper, the tapered MLE is asymptotically as efficient as the true MLE. Recently, [39] showed that under suitable asymptotics, maximum tapered likelihood estimators are consistent and asymptotically normal for a wide range of covariance models.

[19] proposed a combination of tapering and backfitting to estimate the fixed and random spatial component parameters in a very general type of mixed model. They were able to model and analyze spatial datasets several orders of magnitude larger than those analyzed with classical approaches. Tapering techniques in Kalman filter updates were studied in [16].

3.3.2 Tapering for Kriging

Instead of parameter estimation, [18] addressed the problem of covariance tapering for interpolation of large spatial datasets. In geostatistics the standard approach, termed kriging, is based on the principle of minimum mean squared prediction

error. Starting with the simplest spatial model, we assume that $Z(\mathbf{s})$ is observed without any measurement error. Then the best linear unbiased prediction (BLUP) at an unobserved location $\mathbf{s}_0 \in \mathscr{S}$ is

$$\widehat{Z}(\mathbf{s}_0) = \mathbf{c}_0\mathbf{C}^{-1}\mathbf{Z}, \tag{3.4}$$

where $\mathbf{C}_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$ and $\mathbf{c}_{0_i} = C(\mathbf{s}_i, \mathbf{s}_0)$ are based on a possibly nonstationary covariance function C. The mean squared prediction error $\text{MSPE}(\mathbf{s}_0, C)$ has the form

$$\text{MSPE}(\mathbf{s}_0, C) = C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}_0^{\text{T}}\mathbf{C}^{-1}\mathbf{c}_0. \tag{3.5}$$

Similar to (3.2), let $\widetilde{C}(\mathbf{s}, \mathbf{s}_0) = C(\mathbf{s}, \mathbf{s}_0)C_{tap}(\mathbf{s}, \mathbf{s}_0; \gamma)$. By replacing the covariance matrix $\mathbf{C}$ by the tapered version defined by $\widetilde{C}$, the linear system defining the weights in (3.4) can be solved efficiently. The implication is that we limit the covariance function to a local neighborhood. In general we expect the weights $\mathbf{c}_0\mathbf{C}^{-1}$ in (3.4) to be close to zero for observations whose locations are far from $\mathbf{s}_0$. The localization of the weights in the prediction equation motivates kriging using only a neighborhood of locations.

However, if the BLUP (3.4) is calculated under the covariance function $\widetilde{C}$, the mean squared prediction error is of the form

$$\text{MSPE}(\mathbf{s}_0, \widetilde{C}) = C(\mathbf{s}_0, \mathbf{s}_0) - 2\tilde{\mathbf{c}}_0^{\text{T}}\widetilde{\mathbf{C}}^{-1}\mathbf{c}_0 + \tilde{\mathbf{c}}_0^{\text{T}}\widetilde{\mathbf{C}}^{-1}\mathbf{C}\widetilde{\mathbf{C}}^{-1}\tilde{\mathbf{c}}_0, \tag{3.6}$$

where the tilde terms are based on $\widetilde{C}$. For the Matérn covariance family, [18] showed that under specific conditions the asymptotic mean squared error of the predictions in (3.6) using the tapered covariance converges to the minimal error in (3.5). It was also shown that the naive prediction error $\text{MSPE}(\mathbf{s}_0, \widetilde{C})$, assuming that $\widetilde{C}$ is the true covariance function, has the correct convergence rate as well. As can be seen, covariance tapering for kriging purpose is an approximation to the standard linear spatial predictor which is justified to be both asymptotically accurate and computationally efficient.

3.4 Likelihood Approximations

Likelihood approaches for large irregularly spaced spatial datasets are often very difficult, if not infeasible, to implement due to computational limitations. Tapering methods in Sect. 3.3 approximate the Gaussian likelihood through sparse covariance matrices. In this section, we review some other practical ways to evaluate likelihood functions in both spatial and spectral domains.

3.4.1 Likelihood Approximations in the Spatial Domain

In a spatial setting, [46] suggested a simple likelihood approximation. The idea is that any joint density can be written as a product of conditional densities based on some ordering of the observations. Then, one way to lessen the computations is to condition on only some of the "past" observations when computing the conditional densities. Specifically, suppose that $\mathbf{Z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ has joint density $p(\mathbf{z}; \boldsymbol{\phi})$, where $\boldsymbol{\phi}$ is a vector of unknown parameters. By partitioning $\mathbf{Z}$ into subvectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_b$ of possibly different lengths and defining $\mathbf{Z}_{(j)}^{\mathrm{T}} = (\mathbf{Z}_1^{\mathrm{T}}, \ldots, \mathbf{Z}_j^{\mathrm{T}})$, we always have

$$p(\mathbf{z}; \boldsymbol{\phi}) = p(\mathbf{z}_1; \boldsymbol{\phi}) \prod_{j=2}^{b} p(\mathbf{z}_j | \mathbf{z}_{(j-1)}; \boldsymbol{\phi}). \tag{3.7}$$

To calculate the conditional densities $p(\mathbf{z}_j | \mathbf{z}_{(j-1)}; \boldsymbol{\phi})$, it may not be crucial to condition on all components of $\mathbf{z}_{(j-1)}$ for the purpose of reducing the computational effort. In particular, if, for $j = 1, \ldots, b-1$, $\mathbf{V}_{(j)}$ is some subvector of $\mathbf{Z}_{(j)}$, then we have the approximation:

$$p(\mathbf{z}; \boldsymbol{\phi}) \approx p(\mathbf{z}_1; \boldsymbol{\phi}) \prod_{j=2}^{b} p(\mathbf{z}_j | \mathbf{v}_{(j-1)}; \boldsymbol{\phi})$$

which is the general form of Vecchia's approximation to the likelihood. For Gaussian $\mathbf{Z}$, the best linear predictor (BLP) of $\mathbf{Z}_j$ given $\mathbf{Z}_{(j-1)}$ is the conditional expectation $\mathrm{E}(\mathbf{Z}_j | \mathbf{Z}_{(j-1)}; \boldsymbol{\phi})$ as a function of $\boldsymbol{\phi}$, and therefore, $p(\mathbf{z}_j | \mathbf{z}_{(j-1)}; \boldsymbol{\phi})$ is the density of the error of the BLP of $\mathbf{Z}_j$ given $\mathbf{Z}_{(j-1)}$. Vecchia's approximation is accomplished by replacing this density with the one for errors of the BLP of $\mathbf{Z}_j$ given $\mathbf{V}_{(j-1)}$.

[44] adapted Vecchia's approach for the full likelihood to approximate the restricted likelihood of a Gaussian process and showed that the approximation gives unbiased estimating equations. Suppose that $\mathbf{Z} \sim_n (\mathbf{X}\boldsymbol{\beta}, \mathbf{C}(\boldsymbol{\theta}))$, where $\mathbf{X}$ is a known $n \times q$ matrix of rank q, $\boldsymbol{\beta} \in \mathbb{R}^q$ is a vector of unknown coefficients and $\boldsymbol{\theta} \in \Theta$ is a length r vector of unknown parameters for the covariance matrix of $\mathbf{Z}$, then $\boldsymbol{\phi} = (\boldsymbol{\beta}, \boldsymbol{\theta})$. For estimating $\boldsymbol{\theta}$, the maximum likelihood acts as if $\boldsymbol{\beta}$ were known and, hence, tends to underestimate the variation of the spatial process. Restricted maximum likelihood (REML) avoids this problem by considering $\boldsymbol{\beta}$ as nuisance and estimating $\boldsymbol{\theta}$ by using only contrasts, or linear combinations of the observations whose means do not depend on $\boldsymbol{\beta}$.

Just as the full likelihood can be written in terms of the densities of errors of BLPs, the restricted likelihood can also be written in terms of the densities of errors of best linear unbiased predictors (BLUPs) similar to equation (3.7). Specifically, let $\mathbf{Z}_i$ have length n_i and take $\mathbf{X}_i$ to be the corresponding n_i rows of $\mathbf{X}$ assuming that $\mathrm{rank}(\mathbf{X}_1) = q$. For $j > 1$, let $\mathbf{B}_j(\boldsymbol{\theta})$ be the $n_j \times n$ matrix such that

$\mathbf{W}_j(\boldsymbol{\theta}) = \mathbf{B}_j(\boldsymbol{\theta})\mathbf{Z}$ is the vector of errors of the BLUP of $\mathbf{Z}_j$ based on $\mathbf{Z}_{(j-1)}$. For $j = 1$, take $\mathbf{B}_1(\boldsymbol{\theta})$ to be a matrix independent of $\boldsymbol{\theta}$ of size $(n_1 - q) \times n$ such that $\mathbf{W}_1(\boldsymbol{\theta}) = \mathbf{B}_1(\boldsymbol{\theta})\mathbf{Z}$ is a set of contrasts of $\mathbf{Z}_1$. Then $\mathbf{W}_j(\boldsymbol{\theta}) \sim N_{n_j}(\mathbf{0}, \mathbf{A}_j(\boldsymbol{\theta}))$ where $\mathbf{A}_j(\boldsymbol{\theta}) = \mathbf{B}_j(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})\mathbf{B}_j^{\mathrm{T}}(\boldsymbol{\theta})$. We then could obtain the restricted likelihood, which only depends on $\boldsymbol{\phi}$ through $\boldsymbol{\theta}$:

$$rl(\boldsymbol{\theta}; \mathbf{Z}) = \frac{n-q}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{b}\left[\log\{\det(\mathbf{A}_j)\} + \mathbf{W}_j^{\mathrm{T}}\mathbf{A}_j^{-1}\mathbf{W}_j\right].$$

Now consider approximating the restricted likelihood by computing the BLUP of $\mathbf{Z}_j$ in terms of some subvector $\mathbf{V}_{(j-1)}$ of $\mathbf{Z}_{(j-1)}$. The BLUP of, say, Z_ℓ given some subvector $\mathbf{S}$ of $\mathbf{Z}$ that does not contain Z_ℓ is just the linear combination $\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{S}$ that minimizes $\mathrm{var}(Z_\ell - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{S})$ subject to $\mathrm{E}(Z_\ell - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{S}) = 0$ for all values of $\boldsymbol{\beta}$. Let $\mathbf{V}$ be the collection of subvectors $\mathbf{V}_{(1)}, \ldots, \mathbf{V}_{(b-1)}$. Define $\mathbf{W}_1(\mathbf{V}) = \mathbf{W}_1$ and, for $j > 1$, $\mathbf{W}_j(\mathbf{V})$ is the error of the BLUP of $\mathbf{Z}_j$ based on $\mathbf{V}_{(j-1)}$. Let $\mathbf{A}_j(\mathbf{V})$ be the covariance matrix of $\mathbf{W}_j(\mathbf{V})$. Then the approximation to $rl(\boldsymbol{\theta}; \mathbf{Z})$ is of the form

$$rl(\boldsymbol{\theta}; \mathbf{V}) = \frac{n-q}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{b}\left[\log\{\det(\mathbf{A}_j(\mathbf{V}))\} + \mathbf{W}_j^{\mathrm{T}}(\mathbf{V})\mathbf{A}_j(\mathbf{V})^{-1}\mathbf{W}_j(\mathbf{V})\right]. \tag{3.8}$$

Having this restricted likelihood approximation, [44] showed that equation (3.8) gives a set of unbiased estimating equations for $\boldsymbol{\theta}$. The properties of its solutions were studied using the well-developed theory of estimating equations, and the effectiveness of various choices for $\mathbf{V}$ was also investigated. [46] only considered prediction vectors of length 1 such that $\mathbf{Z}_j = Z_j$, whereas [44] considered prediction vectors $\mathbf{Z}_j$ of length greater than 1 and added more observations in the conditioning set rather than just the nearest neighbors in order to further reduce the computational effort and to improve the efficiency of the estimated parameters. However, difficulties with the composite likelihoods of [46] and [44] arise with the selection of the observation order and of the conditioning sets as pointed out by [45], who reviewed recent developments of composite likelihood. To overcome such complications, three different likelihood approximations together with their statistical properties all based on splitting the data into blocks were proposed and investigated by [4, 5].

3.4.2 Likelihood Approximations in the Spectral Domain

The method proposed by [44] is a spatial domain approach. There are also some spectral methods which give another way to approximate the likelihood without involving the calculation of determinants, and to obtain the MLEs of the covariance

parameters $\boldsymbol{\theta}$. These methods are based on [47]'s approximation to the Gaussian negative log-likelihood, which can only be used for datasets observed on a regular complete lattice. In this situation, fewer calculations are required. For irregularly spaced data, [15] presented a version of Whittle's approximation to the Gaussian negative log-likelihood by introducing a lattice process. Additional computational savings were obtained by truncating the spectral representation of the lattice process.

Suppose Z is a continuous Gaussian spatial process of interest with a covariance function C, observed at m irregularly spaced locations in $\mathbb{R}^2$. Let f_Z be the stationary spectral density of Z, which is the Fourier transform of the covariance function:

$$f_Z(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \exp(-i\mathbf{h}^{\mathrm{T}}\boldsymbol{\omega}) C(\mathbf{h}) d\mathbf{h}.$$

We define a process Y at location $\mathbf{s}$ as the integral of Z in a block of area Δ^2 centered at $\mathbf{s}$,

$$Y(\mathbf{s}) = \Delta^{-2} \int h(\mathbf{s} - \tilde{\mathbf{s}}) Z(\tilde{\mathbf{s}}) d\tilde{\mathbf{s}}, \tag{3.9}$$

where for $\mathbf{u} = (u_1, u_2)$ we have

$$h(\mathbf{u}) = \begin{cases} 1, & \text{if } |u_1| < \Delta/2, \ \ |u_2| < \Delta/2, \\ 0, & \text{otherwise.} \end{cases}$$

Then Y is also a stationary process with spectral density f_Y given by

$$f_Y(\boldsymbol{\omega}) = \Delta^{-2} |\Gamma(\boldsymbol{\omega})|^2 f_Z(\boldsymbol{\omega}),$$

where $\Gamma(\boldsymbol{\omega}) = \int h(\mathbf{u}) e^{-i\boldsymbol{\omega}^{\mathrm{T}}\mathbf{u}} d\mathbf{u} = [2\sin(\Delta\omega_1/2)/\omega_1][2\sin(\Delta\omega_2/2)/\omega_2]$ and $\boldsymbol{\omega} = (\omega_1, \omega_2)^{\mathrm{T}}$.

For small values of Δ, $f_Y(\boldsymbol{\omega})$ is approximately $f_Z(\boldsymbol{\omega})$, because we have

$$\lim_{\Delta \to \mathbf{0}} \Delta^{-2} |\Gamma(\boldsymbol{\omega})|^2 = 1.$$

By (3.9), $Y(\mathbf{s})$ can be treated as a continuous spatial process defined for all $\mathbf{s} \in \mathscr{S}$, but here we consider the process Y only on a lattice $(n_1 \times n_2)$ of sample size $m = n_1 n_2$. That is, the values of $\mathbf{s}$ in (3.9) are the centroids of the m grid cells in the lattice, where the spacing between neighboring sites is Δ. Then the spectrum of observations of the sample sequence $Y(\Delta\mathbf{s})$, for $\mathbf{s} \in \mathbb{Z}^2$, is concentrated within the finite-frequency band $-\pi/\Delta \le \boldsymbol{\omega} < \pi/\Delta$ (aliasing phenomenon). The spectral density $f_{\Delta,Y}$ of the process on the lattice can be written in terms of the spectral density f_Y of the process Y as

$$f_{\Delta,Y}(\boldsymbol{\omega}) = \sum_{\mathbf{Q}\in\mathbb{Z}^2} f_Y(\boldsymbol{\omega} + 2\pi\mathbf{Q}/\Delta). \tag{3.10}$$

[15] justified that in practice the sum in (3.10) can be truncated after $2m$ terms.

Whittle's approximation to the Gaussian negative log-likelihood is of the form

$$\frac{m}{(2\pi)^2} \sum_{\boldsymbol{\omega}} \left\{ \log f_\Delta(\boldsymbol{\omega}) + I_m(\boldsymbol{\omega}) f_\Delta(\boldsymbol{\omega})^{-1} \right\}, \tag{3.11}$$

where the sum is evaluated at the Fourier frequencies, I_m is the periodogram, and f_Δ is the spectral density of the lattice process. Now for the lattice process of Y, by computing the periodogram, Whittle's approximate likelihood (3.11) can be applied to $f_{\Delta,Y}$, written in terms of f_Y, then f_Z. Therefore, we can obtain the MLE for the covariances/spectral density parameters of Z. [15] also showed that this version of Whittle's approximation converges to the exact negative log-likelihood for Y, and if n is the total number of observations of the process Z, the calculation requires $O(m \log_2 m + n)$ operations rather than $O(n^3)$ for the exact likelihood of Z.

Another spectral method proposed by [34] extends the definition of a periodogram for time series to the situation where the sampling locations are irregularly spaced. They showed that the well-known property for time series that the periodogram at different frequencies are asymptotically independent still holds for irregularly spaced data. Therefore, it allows for nonparametric and parametric spatial spectral estimators similar to the classical time series analysis setting.

For a stationary random field Z observed at irregularly spaced locations in $\mathbb{R}^d$, by assuming some distribution of the sampling locations, [34] defined the periodogram based on a finite Fourier transform of $Z(\mathbf{s})$ as well as a tapered version of the periodogram. Just as the methods of estimating spectral densities in time series analysis, both nonparametric and parametric estimators were then proposed. The nonparametric method is the spectral window estimator and the parametric approach is based on Whittle's likelihood approximation using the proposed periodogram. Their asymptotic properties were studied and comparisons with [44] and [15] were reported on numerical examples. [44] focused on high frequency components to estimate parameters which better capture the behavior at very short distances, while [34] focused on low frequency components and [15] did on both high and low frequency components. In terms of computational considerations, the latter two spectral methods have a clear advantage.

3.5 Latent Processes

Statistics for spatial data also faces the problem of dealing with noisy data when the interest is in inference on unobserved latent processes. For large spatial datasets, one way to speed up computation is from the perspective of data dimension reduction. [3] developed a spatial model, called a Gaussian predictive process model, which we

introduce in Sect. 3.5.1. There, an approximation of the latent process is proposed to achieve dimension reduction. In Sect. 3.5.2, another solution is given by [7] who defined a spatial mixed effects model for the latent process and proposed fixed rank kriging within a flexible family of nonstationary covariance functions.

First, we define a latent process. Let $\{Y(\mathbf{s}); \mathbf{s} \in \mathscr{S} \subset \mathbb{R}^d\}$ be a real-valued spatial process. We are interested in making inference on the latent process Y on the basis of data that have measurement error. Consider the process Z defined by

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}), \quad \mathbf{s} \in \mathscr{S},$$

where $\{\epsilon(\mathbf{s}); \mathbf{s} \in \mathscr{S}\}$ is a spatial white noise process with mean 0, $\text{var}\{\epsilon(\mathbf{s})\} = \tau^2 \upsilon(\mathbf{s}) \in (0, \infty)$ for $\tau^2 > 0$ and a known $\upsilon(\cdot)$. In fact, the process Z is observed only at a finite number of spatial locations. Define the vector of available data to be $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^{\mathrm{T}}$.

The hidden process Y is assumed to have a linear mean structure,

$$Y(\mathbf{s}) = \mathbf{x}^{\mathrm{T}}(\mathbf{s})\boldsymbol{\beta} + \omega(\mathbf{s}), \quad \mathbf{s} \in \mathscr{S},$$

where $\mathbf{x}(\mathbf{s}) = (x_1(\mathbf{s}), \ldots, x_q(\mathbf{s}))^{\mathrm{T}}$ represents a $q \times 1$ vector of known covariates; the coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^{\mathrm{T}}$ are unknown, and the process ω has zero mean, $0 < \text{var}\{\omega(\mathbf{s})\} < \infty$, for all $\mathbf{s} \in \mathscr{S}$, and a generally nonstationary spatial covariance function:

$$\text{cov}\{\omega(\mathbf{s}), \omega(\mathbf{s}')\} = C(\mathbf{s}, \mathbf{s}'), \quad \mathbf{s}, \mathbf{s}' \in \mathscr{S}.$$

We discuss techniques to reduce computational burden for large spatial datasets under the model

$$Z(\mathbf{s}) = \mathbf{x}^{\mathrm{T}}(\mathbf{s})\boldsymbol{\beta} + \omega(\mathbf{s}) + \epsilon(\mathbf{s}). \tag{3.12}$$

That is, $\mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} = \mathbf{C} + \tau^2\mathbf{V}$, where $\mathbf{X} = [\mathbf{x}^{\mathrm{T}}(\mathbf{s}_i)]_{i=1}^n$ is a matrix of regressors, $\mathbf{C} = [C(\mathbf{s}_i, \mathbf{s}_j)]_{i,j=1}^n$ and $\mathbf{V} = \text{diag}\{\upsilon(\mathbf{s}_1), \ldots, \upsilon(\mathbf{s}_n)\}$.

3.5.1 Gaussian Predictive Processes

With regard to the challenge of computational cost on covariance matrices, [3] proposed a class of models based on the idea of a spatial predictive process which is motivated by kriging ideas. The predictive process projects the original process onto a subspace generated by realizations of the original process at a specified set of locations (or knots). The approach is in the same spirit as process modeling approaches using basis functions and kernel convolutions, that is, specifications which attempt to facilitate computations through lower dimensional process representations.

Assume at locations $\mathbf{s} \in \mathscr{S} \subset \mathbb{R}^2$, a response variable $Z(\mathbf{s})$ is observed from model (3.12), where $\omega(\mathbf{s})$ is a zero-centered Gaussian Process (GP) with covariance function $C(\mathbf{s}, \mathbf{s}')$ capturing the effect of unobserved covariates with spatial pattern, and $\epsilon(\mathbf{s}) \sim N(0, \tau^2)$ is an independent process, often called the nugget, that models the measurement error. In applications, we usually specify $C(\mathbf{s}, \mathbf{s}') = \sigma^2 \rho(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta})$ where $\rho(\cdot; \boldsymbol{\theta})$ is a correlation function and $\boldsymbol{\theta}$ includes parameters in the covariance function. The likelihood for n observations is based on $\mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_\mathbf{Z})$, with $\boldsymbol{\Sigma}_\mathbf{Z} = \mathbf{C}(\boldsymbol{\theta}) + \tau^2 \mathbf{I}_n$, where $\mathbf{C}(\boldsymbol{\theta}) = [C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})]_{i,j=1}^n$. To project the original or parent process onto a subspace, consider the lower-dimensional subspace chosen by the user by selecting a set of knots, $\mathscr{S}^* = \{\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*\}$, which may or may not form a subset of the entire collection of observed locations $\mathscr{S}$. The predictive process $\tilde{\omega}(\mathbf{s})$ is defined as the kriging interpolator

$$\tilde{\omega}(\mathbf{s}) = \mathrm{E}[\omega(\mathbf{s})|\boldsymbol{\omega}^*] = \mathbf{c}^\mathrm{T}(\mathbf{s}; \boldsymbol{\theta})\mathbf{C}^{*-1}(\boldsymbol{\theta})\boldsymbol{\omega}^*, \tag{3.13}$$

where $\boldsymbol{\omega}^* = [\omega(\mathbf{s}_i^*)]_{i=1}^m \sim N_m(\mathbf{0}, \mathbf{C}^*(\boldsymbol{\theta}))$ is derived from the parent process realization over the knots in $\mathscr{S}^*$, $\mathbf{C}^*(\boldsymbol{\theta}) = [C(\mathbf{s}_i^*, \mathbf{s}_j^*; \boldsymbol{\theta})]_{i,j=1}^m$ is the corresponding $m \times m$ covariance matrix, and $\mathbf{c}(\mathbf{s}; \boldsymbol{\theta}) = [C(\mathbf{s}, \mathbf{s}_j^*; \boldsymbol{\theta})]_{j=1}^m$.

The predictive process $\tilde{\omega}(\mathbf{s}) \sim GP(0, \widetilde{C}(\cdot))$ defined in (3.13) has nonstationary covariance function,

$$\widetilde{C}(\mathbf{s}, \mathbf{s}'; \boldsymbol{\theta}) = \mathbf{c}^\mathrm{T}(\mathbf{s}; \boldsymbol{\theta})\mathbf{C}^{*-1}(\boldsymbol{\theta})\mathbf{c}(\mathbf{s}'; \boldsymbol{\theta}),$$

and is completely specified by the parent covariance function. Realizations associated with $\mathbf{Z}$ are given by $\tilde{\boldsymbol{\omega}} = [\tilde{\omega}(\mathbf{s}_i)]_{i=1}^n \sim N_n(\mathbf{0}, \mathbf{c}^\mathrm{T}(\boldsymbol{\theta})\mathbf{C}^{*-1}(\boldsymbol{\theta})\mathbf{c}(\boldsymbol{\theta}))$, where $\mathbf{c}^\mathrm{T}(\boldsymbol{\theta})$ is the $n \times m$ matrix whose ith row is given by $\mathbf{c}^\mathrm{T}(\mathbf{s}_i; \boldsymbol{\theta})$. The theoretical properties of the predictive process including its role as an optimal projection have been discussed in [3].

Replacing $\omega(\mathbf{s})$ in model (3.12) with $\tilde{\omega}(\mathbf{s})$, we obtain the predictive process model

$$Z(\mathbf{s}) = \mathbf{x}^\mathrm{T}(\mathbf{s})\boldsymbol{\beta} + \tilde{\omega}(\mathbf{s}) + \epsilon(\mathbf{s}). \tag{3.14}$$

Since in (3.13), $\tilde{\omega}(\mathbf{s})$ is a spatially varying linear transformation of $\boldsymbol{\omega}^*$, the dimension reduction is seen immediately. In fitting model (3.14), the n random effects $\{\omega(\mathbf{s}_i), i = 1, \ldots, n\}$ are replaced with only m random effects in $\boldsymbol{\omega}^*$. So we can work with an m-dimensional joint distribution involving only $m \times m$ matrices. Although we introduced the same set of parameters in both models (3.12) and (3.14), they are not identical.

The predictive process systematically underestimates the variance of the parent process $\omega(\mathbf{s})$ at any location $\mathbf{s}$, since we have $\mathrm{var}\{\tilde{\omega}(\mathbf{s})\} = \mathbf{c}^\mathrm{T}(\mathbf{s}; \boldsymbol{\theta})\mathbf{C}^{*-1}(\boldsymbol{\theta})\mathbf{c}(\mathbf{s}; \boldsymbol{\theta})$, $\mathrm{var}\{\omega(\mathbf{s})\} = C(\mathbf{s}, \mathbf{s})$ and $0 \leq \mathrm{var}\{\omega(\mathbf{s})|\boldsymbol{\omega}^*\} = C(\mathbf{s}, \mathbf{s}) - \mathbf{c}^\mathrm{T}(\mathbf{s}; \boldsymbol{\theta})\mathbf{C}^{*-1}(\boldsymbol{\theta})\mathbf{c}(\mathbf{s}; \boldsymbol{\theta})$. In practical implementations, this often reveals itself by overestimating the nugget variance in model (3.12), where the estimated τ^2 roughly captures $\tau^2 + \mathrm{E}\{C(\mathbf{s}, \mathbf{s}) - \mathbf{c}^\mathrm{T}(\mathbf{s}; \boldsymbol{\theta})\mathbf{C}^{*-1}(\boldsymbol{\theta})\mathbf{c}(\mathbf{s}; \boldsymbol{\theta})\}$. Here $\mathrm{E}\{C(\mathbf{s}, \mathbf{s}) - \mathbf{c}^\mathrm{T}(\mathbf{s}, \boldsymbol{\theta})\mathbf{C}^{*-1}(\boldsymbol{\theta})\mathbf{c}(\mathbf{s}, \boldsymbol{\theta})\}$ denotes the

averaged bias underestimation over the observed locations. Indeed, [3] observed that while predictive process models employing a few hundred knots excelled in estimating most parameters in several complex high-dimensional models for datasets involving thousands of data points, reducing this upward bias in τ^2 was problematic.

To remedy this problem, [14] proposed a modified predictive process, defined as $\ddot{\omega}(\mathbf{s}) = \tilde{\omega}(\mathbf{s}) + \tilde{\epsilon}(\mathbf{s})$, where now $\tilde{\epsilon}(\mathbf{s}) \sim N(0, C(\mathbf{s},\mathbf{s}) - \mathbf{c}^{\mathrm{T}}(\mathbf{s};\boldsymbol{\theta})\mathbf{C}^{*-1}(\boldsymbol{\theta})\mathbf{c}(\mathbf{s};\boldsymbol{\theta}))$ is a process of independent variables but with spatially adaptive variances. It is then easy to see that $\mathrm{var}\{\ddot{\omega}(\mathbf{s})\} = C(\mathbf{s},\mathbf{s}) = \mathrm{var}\{\omega(\mathbf{s})\}$, as desired. Furthermore, $\mathrm{E}\{\ddot{\omega}(\mathbf{s})|\boldsymbol{\omega}^*\} = \tilde{\omega}(\mathbf{s})$ which ensures that $\ddot{\omega}(\mathbf{s})$ inherits the attractive properties of $\tilde{\omega}$ discussed by [3]; see also [2]. A recent application of predictive process models to a complex dataset from forestry can be found in [13].

3.5.2 *Fixed Rank Kriging*

Kriging, the spatial optimal interpolation, involves the inversion of covariance matrices. Straightforward kriging of massive datasets is not possible and ad hoc local kriging neighborhoods are typically used. One remedy is to approximate the kriging equations, for example by means of covariance tapering as we discussed in Sect. 3.3.2. Another approach is to choose classes of covariance functions for which kriging can be done exactly, even though the spatial datasets are large. One advantage of having a spatial model that allows exact computations is that there is no concern about how close the approximate kriging predictors and approximate mean squared prediction errors are to the corresponding theoretical values. For exact methods, two important questions arise: how flexible are the spatial covariance functions that are used for kriging and how are they fitted. [7] constructed a multiresolution spatial process and showed that there is a very rich class of spatial covariances for which kriging of large spatial datasets can be carried out both exactly and extremely rapidly, with computational complexity linear in the size of the data. They showed how to specify the $n \times n$ covariance matrix $\boldsymbol{\Sigma}$ so that $\boldsymbol{\Sigma}^{-1}$ can be obtained by inverting $m \times m$ positive definite matrices, where m is fixed and independent of n. The result is a spatial BLUP (kriging) procedure which they called Fixed Rank Kriging (FRK); see also [41].

For model (3.12), the kriging predictor of $Y(\mathbf{s}_0)$ in terms of the covariance function is

$$\widehat{Y}(\mathbf{s}_0) = \mathbf{x}^{\mathrm{T}}(\mathbf{s}_0)\hat{\boldsymbol{\beta}} + \mathbf{g}^{\mathrm{T}}(\mathbf{s}_0)(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}), \tag{3.15}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}$, $\mathbf{g}^{\mathrm{T}}(\mathbf{s}_0) = \mathbf{c}^{\mathrm{T}}(\mathbf{s}_0)\boldsymbol{\Sigma}^{-1}$ and $\mathbf{c}(\mathbf{s}_0) = [C(\mathbf{s}_0, \mathbf{s}_j)]_{j=1}^{n}$.

The FRK method captures the scales of spatial dependence through a set of m (not necessarily orthogonal) basis functions, $\mathbf{B}(\mathbf{s}) = (B_1(\mathbf{s}), \ldots, B_m(\mathbf{s}))^{\mathrm{T}}$, for $\mathbf{s} \in \mathbb{R}^d$, where m is fixed. For any $m \times m$ positive definite matrix $\mathbf{G}$, we model $\mathrm{cov}\{Y(\mathbf{s}), Y(\mathbf{s}')\}$ according to

$$C(\mathbf{s}, \mathbf{s}') = \mathbf{B}^{\mathrm{T}}(\mathbf{s})\mathbf{G}\mathbf{B}(\mathbf{s}'), \quad \mathbf{s}, \mathbf{s}' \in \mathbb{R}^d, \tag{3.16}$$

which can be shown to be a valid covariance function. It is easy to see that expression (3.16) is a consequence of writing $\omega(\mathbf{s}) = \mathbf{B}^{\mathrm{T}}(\mathbf{s})\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is an m-dimensional vector with $\mathrm{var}(\boldsymbol{\eta}) = \mathbf{G}$. [7] called the model for ω a spatial random effects model. Hence, $Y(\mathbf{s}) = \mathbf{x}^{\mathrm{T}}(\mathbf{s})\boldsymbol{\beta} + \mathbf{B}^{\mathrm{T}}(\mathbf{s})\boldsymbol{\eta}$ is a spatial mixed effects model.

From expression (3.16), we can write the $n \times n$ covariance matrix as

$$\boldsymbol{\Sigma} = \mathbf{B}\mathbf{G}\mathbf{B}^{\mathrm{T}} + \tau^2\mathbf{V}. \tag{3.17}$$

Both $\mathbf{B}$, the $n \times m$ matrix whose (i, l)th element is $\mathbf{B}_l(\mathbf{s}_i)$, and $\mathbf{V}$, a diagonal matrix with entries given by the measurement error variances, are assumed known. Further,

$$\mathrm{cov}\{Y(\mathbf{s}_0), \mathbf{Z})\} = \mathbf{c}^{\mathrm{T}}(\mathbf{s}_0) = \mathbf{B}^{\mathrm{T}}(\mathbf{s}_0)\mathbf{G}\mathbf{B}^{\mathrm{T}}.$$

[7] showed that the choice of the covariance function (3.16) allows alternative ways of computing the kriging equations involving inversion of only $m \times m$ matrices. From equation (3.17),

$$\boldsymbol{\Sigma}^{-1} = \tau^{-1}\mathbf{V}^{-1/2}\left\{\mathbf{I} + (\tau^{-1}\mathbf{V}^{-1/2}\mathbf{B})\mathbf{G}(\tau^{-1}\mathbf{V}^{-1/2}\mathbf{B})^{\mathrm{T}}\right\}^{-1}\tau^{-1}\mathbf{V}^{-1/2}. \tag{3.18}$$

Then we have that, for any $n \times m$ matrix $\mathbf{P}$,

$$\mathbf{I} + \mathbf{P}\mathbf{G}\mathbf{P}^{\mathrm{T}} = \mathbf{I} + (\mathbf{I} + \mathbf{P}\mathbf{G}\mathbf{P}^{\mathrm{T}})\mathbf{P}\mathbf{G}(\mathbf{I} + \mathbf{P}^{\mathrm{T}}\mathbf{P}\mathbf{G})^{-1}\mathbf{P}^{\mathrm{T}}.$$

Premultiplying by $(\mathbf{I} + \mathbf{P}\mathbf{G}\mathbf{P}^{\mathrm{T}})^{-1}$ yields

$$(\mathbf{I} + \mathbf{P}\mathbf{G}\mathbf{P}^{\mathrm{T}})^{-1} = \mathbf{I} - \mathbf{P}(\mathbf{G}^{-1} + \mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1}\mathbf{P}^{\mathrm{T}},$$

which is a particular case of the well-known Sherman-Morrison-Woodbury formulae. Using this in equation (3.18), it yields the computational simplification

$$\boldsymbol{\Sigma}^{-1} = (\tau^2\mathbf{V})^{-1} - (\tau^2\mathbf{V})^{-1}\mathbf{B}\{\mathbf{G}^{-1} + \mathbf{B}^{\mathrm{T}}(\tau^2\mathbf{V})^{-1}\mathbf{B}\}^{-1}\mathbf{B}^{\mathrm{T}}(\tau^2\mathbf{V})^{-1}. \tag{3.19}$$

The formula (3.19) for $\boldsymbol{\Sigma}^{-1}$ involves inverting only fixed rank $m \times m$ positive definite matrices and the $n \times n$ diagonal matrix $\mathbf{V}$. Finally, the kriging predictor (3.15) is

$$\widehat{Y}(\mathbf{s}_0) = \mathbf{x}^{\mathrm{T}}(\mathbf{s}_0)\hat{\boldsymbol{\beta}} + \mathbf{B}^{\mathrm{T}}(\mathbf{s}_0)\mathbf{G}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}$ and $\boldsymbol{\Sigma}^{-1}$ is given by equation (3.19).

For a fixed number of regressors q and a fixed rank m of $\mathbf{G}$ in the covariance model that is defined by (3.16), [7] showed that the computational burden of FRK is only linear in n. The results rely on using a rich class of nonstationary covariance

functions (3.16) that arise from a spatial random effects model. Further, microscale variation in the hidden process Y could be modeled by including another diagonal matrix in equation (3.17),

$$\boldsymbol{\Sigma} = \mathbf{BGB}^{\mathrm{T}} + \xi^2\mathbf{I} + \tau^2\mathbf{V}.$$

When both diagonal matrices are proportional to each other, the measurement error parameter τ^2 and the microscale parameter ξ^2 are not individually identifiable, although their sum $\xi^2 + \tau^2$ is. The presence of the microscale variance ξ^2 was discussed by [7] but they have assumed that the process Y is smooth (i.e. $\xi^2 = 0$). The FRK formulae including ξ^2 were given by [8].

Statistics for spatio-temporal data inherits a similar need for data dimension reduction as what we saw for spatial data, possibly more so since the data size quickly becomes massive as time progresses. [9] built a spatio-temporal random effects model that allows both dimension reduction (spatially) and rapid smoothing, filtering, or forecasting (temporally). They focused on filtering and developed a methodology called Fixed Rank Filtering (FRF); see also [29]. With a similar idea as FRK, the fast statistical prediction of a hidden spatio-temporal process is achieved through spatio-temporal models defined on a space of fixed dimension; the space is defined by the random coefficients of prespecified spatio-temporal basis functions, and the coefficients are assumed to evolve dynamically. By reducing the dimensionality, FRF was proposed as a spatio-temporal Kalman filter, which is able to use past data as well as current data to great effect when estimating a process from a noisy, incomplete, and very large spatio-temporal dataset. [9] showed that the gains can be substantial when the temporal dependence is strong and there are past data at or near locations where the current data have gaps.

3.6 Gaussian Markov Random Field Approximations

Gaussian Markov random fields (GMRFs) possess appealing computational properties due to the sparse pattern of their precision matrices. The numerical factorization of the precision matrix using sparse matrix algorithms can be done at a typical cost of $O(n^{3/2})$ for two-dimensional GMRFs. The computational gains from GMRFs have been exploited to provide fast and accurate Bayesian inference for latent Gaussian field models through integrated nested Laplace approximation (INLA) by [35]. [12] made a further step of applying INLA on top of predictive process models [3] to dramatically reduce the computation in making inference for large spatial datasets. More generally, [36] demonstrated empirically that GMRFs could closely approximate some commonly used covariance functions in geostatistics, and proposed to use GMRFs as computational replacements for Gaussian random fields for example when making kriging predictions [24]. However, their approximation was restricted to Gaussian random fields that are observed over a regular lattice (or torus) and the fit itself had to be precomputed for a discrete set of parameter values. Other literature following this idea includes [1, 10, 25, 42].

[33] recently proposed an approach to find GMRFs with local neighborhood and precision matrix to represent certain Gaussian random fields with Matérn covariance structure. This is accomplished through the following two facts:

(a) The solution of a particular type of stochastic partial differential equation (SPDE) driven by Gaussian white noise is a Gaussian random field with Matérn covariance function. Specifically, let $X(\mathbf{s})$ be the solution of the following linear fractional SPDE:

$$(\kappa^2 - \Delta)^{\alpha/2} X(\mathbf{s}) = W(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \ \kappa > 0, \ \nu > 0,$$

where $W(\mathbf{s})$ is a spatial Gaussian white noise with unit variance. Then $X(\mathbf{s})$ is a Gaussian random field with the Matérn covariance

$$C(\mathbf{k}) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} (\kappa \|\mathbf{k}\|)^{\nu} K_{\nu}(\kappa \|\mathbf{k}\|),$$

where

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}}.$$

Here the Laplacian $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$, K_{ν} is the modified Bessel function of second kind with order $\nu > 0, \kappa > 0$ is a scaling parameter and σ^2 is the marginal variance. The integer value of ν determines the mean square differentiability of the underlying random field.

(b) Let X be a GMRF on a regular two-dimensional lattice indexed by (i, j), where the Gaussian full conditionals are

$$\mathrm{E}(X_{i,j} | \mathbf{X}_{-(i,j)}) = \frac{1}{a}(X_{i-1,j} + X_{i+1,j} + X_{i,j-1} + X_{i,j+1}),$$

and $\mathrm{var}(X_{i,j} | \mathbf{X}_{-(i,j)}) = 1/a$ for $|a| > 4$, where $\mathbf{X}_{-(i,j)}$ denotes the vector of X's on the lattice except at location (i, j). The coefficients in the GMRF representation of the SPDE in (a) over a regular unit-distance two-dimensional infinite lattice for $\nu = 1, 2, \ldots$, can be found by convolving the elements of the precision matrix corresponding to X by itself ν times.

The authors then generalized the above results to enable the construction of the corresponding GMRFs representation of the Matérn field on a triangulated lattice, hence the Gaussian random fields in [33] are no longer restricted to lattice data. This avoids the interpolation of irregularly spaced observations to grid points and allows for finer resolution where details are required. The drawback of this approach is that we can only find the explicit form of GMRFs for those Gaussian random fields that have a Matérn covariance structure at certain integer smoothnesses. Nevertheless, the main results in [33] cover the most important and used covariance models in

spatial statistics, and they can be extended to model Matérn covariances on the sphere, nonstationary locally isotropic Gaussian random fields, Gaussian random fields with oscillating correlation functions, and non-isotropic fields.

3.7 Multivariate Geostatistics

Multivariate spatial data have been increasingly employed in various scientific areas. We have introduced the intrinsic model, a separable covariance structure for multivariate data, to reduce computational cost in Sect. 3.2. In addition, in many statistical applications, predicting a geophysical quantity based on observations at nearby locations of the same quantity and on other related variables, so-called covariables, is of prime interest. Obviously, the analysis should take advantage of covariances between locations as well as covariances among variables. For a single variable of interest, we have discussed the tapering technique for kriging purpose in Sect. 3.3.2 and the fixed rank kriging method in Sect. 3.5.2. When information from several different variables is also available, it should be used for prediction as well. The problems implied by large amounts of data are then further amplified since many observations occur at each location. Therefore, we need methodologies to keep the analysis computationally feasible.

For spatially correlated multivariate random fields, the best linear unbiased prediction (BLUP) is often called cokriging in geostatistics. Assume that we have a primary variable and two or more secondary variables and aim at predicting the primary variable at some location. It is well-known that in a mean squared prediction error (MSPE) sense, the best predictor is the conditional expectation given variables at the other locations, where the set of conditioning variables can be either just the primary variable (i.e., kriging) or some or all of the secondary variables (i.e., cokriging).

Thus, the cokriging technique requires the solution of a large linear system based on the covariance and cross-covariance matrix of all involved variables. For large amounts of data, it is impossible to solve the linear system with direct methods. [17] proposed aggregation-cokriging for highly-multivariate spatial datasets to reduce the computational burden. This method is based on a linear aggregation of the covariables with carefully chosen weights, so that the resulting linear combination of secondary variables contributes as much as possible to the prediction of the primary variable in the MSPE sense. In other words, the secondary variables should be weighted by the strength of their correlation with the location of interest. The prediction is then performed using a simple cokriging approach with the primary variable and the aggregated secondary variables. This reduces the computational burden of the prediction from solving a $(n + \ell m) \times (n + \ell m)$ to solving a $(n+m) \times (n+m)$ linear system, where n and m are the numbers of observations of the primary and secondary variables, respectively, and ℓ is the number of secondary variables. The computational complexity is now comparable with simple bikriging, i.e., simple cokriging with only one of the secondary variables, and its optimality was discussed by [17] under different covariance structures.

Besides cokriging, Gaussian predictive process models can also be generalized to multivariate settings. In Sect. 3.5.1, we have discussed a class of models based on the idea of a univariate spatial predictive process which is motivated by kriging ideas. The predictive process projects the original process onto a subspace generated by realizations of the original process at a specified set of locations (or knots). Similarly, for multivariate spatial processes, the multivariate predictive process extends the preceding concepts to multivariate Gaussian processes. Now $\omega(\mathbf{s})$ from the model (3.12) is assumed to be a p-dimensional zero-centered multivariate Gaussian process $\boldsymbol{\omega}(\mathbf{s})$, where p is the number of variables at each location. For locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$, we write the multivariate realizations as a vector $\boldsymbol{\omega} = [\boldsymbol{\omega}(\mathbf{s}_i)]_{i=1}^n \in \mathbb{R}^{np}$. Analogous to the univariate setting, [3] again considered a set of knots $\mathscr{S}^*$ and denoted by $\boldsymbol{\omega}^*$ the realization of $\boldsymbol{\omega}(\mathbf{s})$ over $\mathscr{S}^*$. Then similar to (3.13), the multivariate predictive process is defined as

$$\tilde{\boldsymbol{\omega}}(\mathbf{s}) = \text{cov}\{\boldsymbol{\omega}(\mathbf{s}), \boldsymbol{\omega}^*\}\text{var}^{-1}(\boldsymbol{\omega}^*)\boldsymbol{\omega}^*,$$

and $\tilde{\boldsymbol{\omega}}(\mathbf{s})$ has properties that are analogous to its univariate counterpart. The projection onto a lower dimensional subspace allows the flexibility to accommodate multivariate processes in the context of large datasets.

3.8 Discussion

Whenever we deal with large spatial datasets, we face problems with storage and computation. When covariance functions produce covariance matrices that are neither sparse nor low rank, it is sometimes possible to compute the exact likelihood even for quite large datasets if they are spatially gridded data. However, exact likelihood calculations are not possible with large numbers of irregularly sited observations.

One solution that has been discussed is to truncate the covariance function to zero and use well-established algorithms to handle sparse systems. Such libraries or toolboxes are available in widely used software packages such as R or Matlab. The tapering for kriging purpose presented in Sect. 3.3.2 is based on the assumption of Matérn covariances which could be weakened, but not entirely eliminated. However, the Matérn family is already sufficiently flexible to model a broad class of processes. The tapering techniques also work for nonstationarity or anisotropic processes at least with conservative taper ranges, but the accuracy of the tapering approximation for nonstationary problems remains an open question [18].

The approximation of the likelihood in either the spatial or spectral domain is another solution to overcome computational obstacles. In the spatial domain, the composite likelihood method in Sect. 3.4.1 is based on conditional densities, which points out the difficulty in choosing conditioning sets and the need for less haphazard rules. Furthermore, the use of this approximation in Bayesian analysis poses considerable challenges, since the approximation accuracy needs to be evaluated especially if the likelihood calculation is just one part of a single step

in a MCMC algorithm [44]. The spectral methods in Sect. 3.4.2 are computationally efficient by avoiding the calculation of determinants and can be easily adapted to model nonstationary processes as a mixture of independent stationary processes [15]. However, they do not overcome the difficulty in prediction with massive data. The Gaussian predictive process model in Sect. 3.5.1 projects the parent process onto a lower dimensional subspace. Since every spatial or spatio-temporal process induces a predictive process model, it is flexible to accommodate non-stationary, non-Gaussian, possibly multivariate, possibly spatio-temporal processes in the context of large datasets. In the same spirit, the fixed rank kriging in Sect. 3.5.2 relies on using a class of nonstationary covariances where kriging can be done exactly. Those techniques can be implemented on very large datasets as the computations are linear in the size of the dataset, and they are highly flexible since they allow the underlying spatial covariances to be nonstationary. Section 3.6 described another approach based on Gaussian Markov random field approximations.

The covariance tapering and the reduced rank based methods have shown great computational gains, but they also have their own drawbacks. The covariance tapering may not be effective in accounting for spatial dependence with long range while the reduced rank based methods usually fail to accurately capture the local, small scale dependence structure. To capture both the large and small scale spatial dependence, [37] proposed to combine the ideas of the reduced rank process approximation and the sparse covariance approximation. They decomposed the spatial Gaussian process into two parts: a reduced rank process to characterize the large scale dependence and a residual process to capture the small scale spatial dependence that is unexplained by the reduced rank process. This idea was then extended to the multivariate setting by [38]. However, the application of tapering techniques to multivariate random fields remains to be explored due to the lack of flexible compactly supported cross-covariance functions.

[49] discussed some strategies to deal with computations for large datasets within the context of their convolution-based spatial nonstationary models. Basically, for parameter estimation they proposed to smooth raw estimates obtained over apriori determined grids, and for predictions they discussed the idea of local window as in [23] and tapering as in [18].

For spatio-temporal processes, we have reviewed the methods to compute separable approximations of space-time covariance matrices. A well-known shortcoming of separable covariance functions is that they do not allow for space-time interactions in the covariance. Nevertheless, the separable space-time structure allows for a simple construction of valid space-time parametric models. By assuming separability, one can further combine separable approximations with the tapering approach. In this case, it is expected that a combination of computational gains can be achieved [20]. When dealing with several variables evolving in space and time, that is, multivariate space-time random fields, the cokriging approaches are even more computationally costly. In this context, the separable approximation techniques can also be combined with other approaches for multivariate spatial processes to further facilitate computational procedures.

In summary, we have compared various methods for handling large spatial datasets from the literature reviewed in this chapter. However, it would be interesting to further compare all of them under different situations with Monte Carlo simulation studies and real data examples. We look forward to the emergence of such work.

Acknowledgements The authors thank Reinhard Furrer for valuable comments on the manuscript. Li's research was partially supported by NSF grant DMS-1007686. Genton's research was partially supported by NSF grants DMS-1007504 and DMS-1100492, and by Award No. KUSC1-016-04, made by King Abdullah University of Science and Technology (KAUST).

References

1. Allcroft, D.J., Glasbey, C.A.: A latent Gaussian Markov random field model for spatio-temporal rainfall disaggregation. J. R. Stat. Soc. Ser. C **52**, 487–498 (2003)
2. Banerjee, S., Finley, A.O., Waldmann, P., Ericsson, T.: Hierarchical spatial process models for multiple traits in large genetic trials. J. Amer. Statist. Assoc. **105**, 506–521 (2010)
3. Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H.: Gaussian predictive process models for large spatial datasets. J. R. Stat. Soc. Ser. B **70**, 825–848 (2008)
4. Caragea, P.C., Smith, R.L.: Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. J. Multivariate Anal. **98**, 1417–1440 (2007)
5. Caragea, P.C., Smith, R.L.: Approximate likelihoods for spatial processes. Technometrics. In press (2011)
6. Cressie, N., Huang, H.C.: Classes of nonseparable, spatio-temporal stationary covariance functions. J. Amer. Statist. Assoc. **94**, 1330–1340 (1999)
7. Cressie, N., Johannesson, G.: Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. Ser. B **70**, 209–226 (2008)
8. Cressie, N., Kang, E.L.: High-resolution digital soil mapping: Kriging for very large datasets. In Proximal Soil Sensing, eds R. Viscarra-Rossel, A. B. McBratney, and B. Minasny. Elsevier, Amsterdam, 49–66 (2010)
9. Cressie, N., Shi, T., Kang, E.L.: Fixed rank filtering for spatio-temporal data. J. Comput. Graph. Statist. **19**, 724–745 (2010)
10. Cressie, N., Verzelen, N.: Conditional-mean least-squares fitting of Gaussian Markov random fields to Gaussian fields. Comput. Statist. Data Anal. **52**, 2794–2807 (2008)
11. Du, J., Zhang, H., Mandrekar, V.S.: Fixed-domain asymptotic properties of tapered maximum likelihood estimators. Ann. Statist. **37**, 3330–3361 (2009)
12. Eidsvik, J., Finley, A., Banerjee, S., Rue, H.: Approximate Bayesian inference for large spatial datasets using predictive process models. Manuscript (2011)
13. Finley, A.O., Banerjee, S., MacFarlane, D.W.: A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas. J. Amer. Statist. Assoc. **106**, 31–48 (2011)
14. Finley, A.O., Sang, H., Banerjee, S., Gelfand, A.E.: Improving the performance of predictive process modeling for large datasets. Comput. Statist. Data Anal. **53**, 2873–2884 (2009)
15. Fuentes, M.: Approximate likelihood for large irregularly spaced spatial data. J. Amer. Statist. Assoc. **102**, 321–331 (2007)
16. Furrer, R., Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivariate Anal. **98**, 227–255 (2007)
17. Furrer, R., Genton, M.G.: Aggregation-cokriging for highly-multivariate spatial data. Biometrika. **98**, 615–631 (2011)

18. Furrer, R., Genton, M.G., Nychka, D.: Covariance tapering for interpolation of large spatial datasets. J. Comput. Graph. Statist. **15**, 502–523 (2006)
19. Furrer, R., Sain, S.: Spatial model fitting for large datasets with applications to climate and microarray problems. Stat. Comput. **19**, 113–128 (2009)
20. Genton, M.G.: Separable approximations of space-time covariance matrices. Environmetrics. Special Issue for METMA3 **18**, 681–695 (2007)
21. Gneiting, T., Genton, M.G., Guttorp, P.: Geostatistical space-time models, stationarity, separability and full symmetry. In Finkenstaedt, B., Held, L. and Isham, V. (eds.), Statistics of Spatio-Temporal Systems, Chapman & Hall/CRC Press, 151–175 (2007)
22. Guan, Y., Sherman, M., Calvin, J.A.: A nonparametric test for spatial isotropy using subsampling. J. Amer. Statist. Assoc. **99**, 810–821 (2004)
23. Haas, T.C.: Lognormal and moving window methods of estimating acid deposition. J. Amer. Statist. Assoc. **85**, 950–963 (1990)
24. Hartman, L., Hössjer, O.: Fast kriging of large data sets with Gaussian Markov random fields. Comput. Statist. Data Anal. **52**, 2331–2349 (2008)
25. Hrafnkelsson, B., Cressie, N.: Hierarchical modeling of count data with application to nuclear fall-out. Environ. Ecol. Stat. **10**, 179–200 (2003)
26. Jun, M., Genton, M.G.: A test for stationarity of spatio-temporal processes on planar and spherical domains. Statist. Sinica. In press (2012)
27. Jun, M., Stein, M.L.: An approach to producing space-time covariance functions on spheres. Technometrics. **49**, 468–479 (2007)
28. Jun, M., Stein, M.L.: Nonstationary covariance models for global data, Ann. Appl. Stat. **2**, 1271–1289 (2008)
29. Kang, E.L., Cressie, N., Shi, T.: Using temporal variability to improve spatial mapping with application to satellite data. Canad. J. Statist. **38**, 271–289 (2010)
30. Kaufman, C., Schervish, M., Nychka, D.: Covariance tapering for likelihood-based estimation in large spatial datasets. J. Amer. Statist. Assoc. **103**, 1556–1569 (2008)
31. Li, B., Genton, M.G., Sherman, M.: A nonparametric assessment of properties of space-time covariance functions. J. Amer. Statist. Assoc. **102**, 736–744 (2007)
32. Li, B., Genton, M.G., Sherman, M.: Testing the covariance structure of multivariate random fields. Biometrika. **95**, 813–829 (2008)
33. Lindgren, F., Rue, H., Lindström, J.: An explict link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). J. R. Stat. Soc. Ser. B, **73**, 423–498 (2011)
34. Matsuda, Y., Yajima, Y.: Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. J. R. Stat. Soc. Ser. B **71**, 191–217 (2009)
35. Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J. R. Stat. Soc. Ser. B **71**, 319–392 (2009)
36. Rue, H., Tjelmeland, H.: Fitting Gaussian Markov random fields to Gaussian fields. Scand. J. Statist. **29**, 31–50 (2002)
37. Sang, H., Huang, J.Z.: A full-scale approximation of covariance functions for large spatial data sets. J. R. Stat. Soc. Ser. B, in press (2011)
38. Sang, H., Jun, M., Huang, J.Z.: Covariance approximation for large multivariate spatial datasets with an application to multiple climate model errors. Ann. Appl. Stat., in press (2011)
39. Shaby, B., Ruppert, D.: Tapered covariance: Bayesian estimation and asymptotics. J. Comput. Graph. Statist., in press (2011)
40. Shao, X., Li, B.: A tuning parameter free test for properties of space-time covariance functions. J. Statist. Plann. Inference. **139**, 4031–4038 (2009)
41. Shi, T., Cressie, N.: Global statistical analysis of MISR aerosol data: a massive data product from NASA's Terra satellite. Environmetrics. **18**, 665–680 (2007)
42. Song, H., Fuentes, M., Gosh, S.: A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. J. Multivariate Anal. **99**, 1681–1697 (2008)

43. Stein, M.L.: A modeling approach for large spatial datasets. J. Korean Statist. Soc. **37**, 3–10 (2008)
44. Stein, M.L., Chi, Z., Welty, L.J.: Approximating likelihoods for large spatial datasets. J. R. Stat. Soc. Ser. B **66**, 275–296 (2004)
45. Varin, C., Reid, N., Firth, D.: An overview on composite likelihood methods. Statist. Sinica. **21**, 5–42 (2011)
46. Vecchia, A.V.: Estimation and model identification for continuous spatial process. J. R. Stat. Soc. Ser. B **50**, 297–312 (1998)
47. Whittle, P.: On stationary processes in the plane. Biometrika. **41**, 434–449 (1954)
48. Zhang, H., Wang, Y.: Kriging and cross-validation for massive spatial data. Environmetrics. **21**, 290–304 (2010)
49. Zhu, Z., Wu, Y.: Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data. J. Comput. Graph. Statist. **19**, 74–95 (2010)
50. Zimmerman, D.: Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models. J. Statist. Comput. Simul. **32**, 1–15 (1989)

Chapter 4
Bayesian Inference for Non-Markovian Point Processes

Peter Guttorp and Thordis L. Thorarinsdottir

Abstract The Bayesian approach to statistical inference has in recent years become very popular, especially in the analysis of complex data sets. This is largely due to the development of Markov chain Monte Carlo methods, which expand the scope of application of Bayesian methods considerably. In this paper, we review the Bayesian contributions to inference for point processes. We focus on non-Markovian processes, specifically Poisson and related models, doubly stochastic models, and cluster models. We also discuss Bayesian model selection for these models and give examples in which Bayes factors are applied both directly and indirectly through a reversible jump algorithm.

4.1 Introduction

Statistical inference for point processes originates, as pointed out by Daley and Vere-Jones, in two sources: life tables, and counting phenomena [12]. Among early sources of inferential work are Graunt, Halley and Newton in the eighteenth century on the life table side, and Newcomb, Abbe and Seidel in the second half of the nineteenth century on the counting side (for Newcomb's contributions, see [22]; the others are all described by Daley and Vere-Jones). The modern approach originated mainly in England in the 1950s and 60s, with Bartlett and Cox as the main names.

T.L. Thorarinsdottir (✉)
Department of Applied Mathematics, Heidelberg University, Germany
e-mail: thordis@uni-heidelberg.de

P. Guttorp
University of Washington, Seattle, WA
Norwegian Computing Centre, Oslo, Norway
e-mail: peter@stat.washington.edu

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_4,

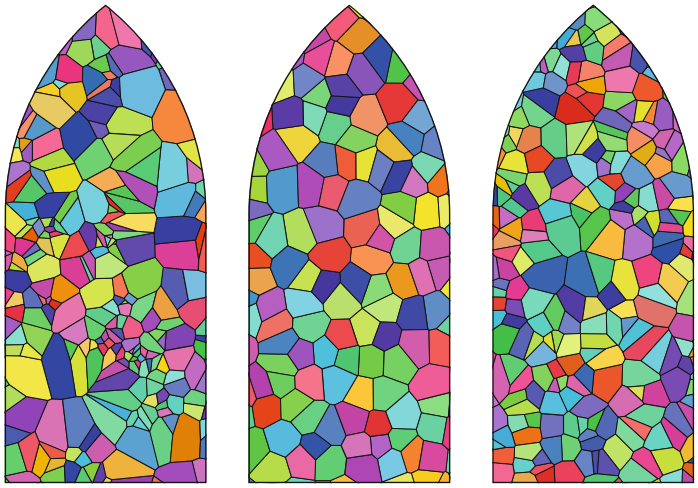

Fig. 4.1 Examples of Voroni tessellations that arise from point process patterns. *Left*: the tessellation of a process which is both clustered and regular (a Matérn type I process, Sect. 4.2.4, yields the cluster centers and a Neyman-Scott process, Sect. 4.4, the points). *Middle*: the tessellation of a regular process (a Matérn type I process). *Right*: the tessellation of a completely random process (Poisson process, Sect. 4.2.1). Generated by Ute Hahn and Dietrich Stoyan

A few examples of Voroni tessellations that arise from point process patterns are shown in Fig. 4.1.

This paper will review the Bayesian contributions to inference for point processes. We will only discuss non-Markovian processes, as lately much of the emphasis has been on Markovian models, and we consider it important not to lose sight of the non-Markovian ones. We make no pretense of a complete literature review; rather, we have chosen papers that we think are interesting or important or that we can use to make a point. A more comprehensive review paper is [47]. Chapter 4 of the recent Handbook of Spatial Statistics is devoted to spatial point processes [16].

We start in Sect. 4.2 by reviewing work on non-parametric estimation (Bayesian is always assumed unless otherwise specified) of the rate of a non-homogeneous Poisson process. Immediately we will see that many processes, and many inference problems, can be viewed from more than one point of view. We then proceed to models derived from a Poisson process using thinning, and show how one can use Bayes factors to distinguish between models of late fall precipitation in upstate New York, USA.

The next Sect. 4.3 deals with doubly stochastic models, and again we encounter the problem of how one views the inference. Section 4.4 deals with cluster processes, where we show an application to brain imaging, and Sect. 4.5 is about model selection. Here we compare the Akaike criterion and the Bayes factor for selecting between types of cluster models. Finally, a short summary is given in Sect. 4.6.

We are grateful to the organizers of the Toledo conference for the opportunity to participate and to write this paper. We also need to thank the Norwegian Computing Center in Oslo, Norway, and the University of Heidelberg, Germany, for accommodating visits by one or the other of us. Alex Lenkoski provided many helpful comments, and we thank Ute Hahn and Dietrich Stoyan for generating the patterns in Fig. 4.1.

4.2 Poisson and Related Processes

The simplest point process model, and the one for which most inference tools have been developed, is the Poisson process. The stationary case does not contain much of interest, as it is a single parameter situation, but the non-stationary case resembles non-parametric density estimation. The thinning approach to simulation will have several applications in this paper, and we also look at some models with sub-Poisson variability (often called regularity in the literature) that are constructed from an underlying Poisson process.

4.2.1 Non-parametric Estimation for Non-stationary Poisson Rate Functions

In 1978 Aalen revolutionized point process analysis by introducing a general non-parametric statistical theory for the class of multiplicative counting processes [1]. It was a frequentist theory, but received a Bayesian adaptation in the work of Lo for Poisson processes [39], and Lo and Weng for the general multiplicative processes [40]. The general multiplicative process was also dealt with in [32], but a Lévy process prior was used. Here we will focus on Lo's treatment of the Poisson process case. Consider a Poisson process with intensity measure ν. Lo showed that a gamma process prior is conjugate. To define the gamma process prior, consider a σ-finite measure α, and say the measure μ is selected by a gamma process prior if for disjoint sets $A_1, \ldots, A_k$ we have that the collection of random variables $\{\mu(A_1), \ldots, \mu(A_k)\}$ are independent gamma random variables of scale 1 and means $\alpha(A_i)$. The measure μ is then said to have shape measure α and scale parameter 1. We denote the corresponding probability measure having these finite-dimensional distributions by $P_{\alpha,1}$. We can re-scale the measure by an α-integrable positive random function β by defining $\beta\mu(A) = \int_A \beta(x)\mu(dx)$ and the corresponding probability measure is denoted $P_{\alpha,\beta}$. Lo showed that if we observe independent realizations $N_1, \ldots, N_n$ of N, and assign a prior measure $P_{\alpha,\beta}$ to the intensity measure ν, then the posterior measure is $P_{\alpha+\sum_1^n N_i, \beta/(1+n\beta)}$.

Consider now the special case where $\beta(x) = 1/\theta$, and suppose we are interested in estimating the intensity $\nu_t = \nu(0,t]$ under integrated squared error loss. It is not

difficult to verify that the Bayes estimator is

$$\frac{\theta}{\theta+n}\frac{\alpha(0,t]}{\theta}+\frac{n}{n+\theta}\frac{1}{n}\sum_{i=1}^{n}N_i(0,t],$$

i.e., a weighted average of the prior guess and the sample empirical estimate.

Generally the tools needed to estimate non-parametrically a non-homogeneous Poisson process with time dependent rate $\lambda(t)$, assumed integrable over the period of observation A, are the same as those for density estimation. Conditionally upon the total number of points $N = N(A)$ the points are distributed as the order statistics from a distribution with density

$$f(s) = \lambda(s)/\int_A \lambda(u)du \tag{4.1}$$

[10]. This fact was used to develop a kernel estimator for the intensity [13], and for a histogram type estimator in setting up a Bayesian analysis of an interesting problem in musicology [49]. In order to create a Bayesian structure, it has been popular to assign a prior related to a Gaussian process, typically of the form $\exp(X(t))$ where $X(t)$ is a Gaussian process. By the same misnomer as for the log-normal distribution, this tends to be called a log Gaussian Cox process, although it is the log intensity which is Gaussian, and in our context serves as a prior for a non-homogeneous Poisson process intensity, while the setup mathematically (albeit not conceptually) corresponds to a doubly stochastic Poisson process [9]. The doubly stochastic Poisson process is of course of interest in its own right (see Sect. 4.3). The conditional likelihood for this model, given the realization of $\lambda(s), s \in A$, is simply the usual Poisson likelihood

$$L(\lambda(s)) = \exp\left(\int_A (\log\lambda(s)dN_s - \lambda(s)ds)\right). \tag{4.2}$$

For random infinite dimensional $\lambda(s)$ the integral in the exponential of (4.2) cannot be evaluated explicitly, which makes Bayesian inference with a prior $Y(t)$ based on a Gaussian process intractable. A discretization approach to obtain a tractable expression for the likelihood has been used [11,45] and applied this to the Bayesian problem we are considering in this Sect. [8]. The idea is to approximate the continuous process $Y(t)$ by a sequence of step functions in the linear case, and values on a grid in the spatial case. Waagepetersen showed that the resulting posterior density converges to the true posterior as the discretization interval shrinks to zero [59]. Both he and others have pointed out the sensitivity of the resulting inference to the discretization scheme [8,59].

In [24] the authors took a similar route, using piecewise constant functions with random number of jumps of random size as prior on the intensity function, but not thinking of this as an approximation to a smooth prior process. It does not follow,

for example, that the posterior mean is piecewise constant. In fact, it typically comes out smooth.

Kottas used the representation in (4.1) to develop a different estimation method [33]. Treating $\gamma = \int_A \lambda(u)du$ as a separate parameter, he used explicit density estimation tools to estimate $f(s)$. We let $A = (0, T]$. Then f is estimated as a Dirichlet mixture of scaled beta densities (supported on $(0, T]$). The Dirichlet process is determined by a precision parameter α, which is given a gamma prior, and a base distribution, which is a function of the location and dispersion of the beta distributions. These are taken to be independent uniform and inverse gamma, respectively. Finally, γ is given a Jeffreys prior of the form $1/x$.

4.2.2 The Thinning Approach to Simulation

Lewis and Shedler introduced the standard approach to generating non-homogeneous Poisson processes on a set A [37]. If the rate is $\lambda(s)$ and we write $\lambda^* = \sup_{t \in A} \lambda(t)$, their *thinning* approach is to generate a homogeneous Poisson process of rate λ^*, and then keep a point at location τ with probability $\lambda(\tau)/\lambda^*$. This is, of course, a form of rejection sampling.

The Lewis-Shedler method has been extended to enable exact computation of the posterior distribution of a non-homogeneous Poisson process with a Gaussian process prior of the form $\lambda^* \sigma(X(s))$, where $\sigma(x) = (1 + \exp(-x))^{-1}$, by keeping track of the deleted locations as well as the values of the Gaussian process at both the deleted and the kept locations, which are treated as a latent variable [2]. The approach is to use a Markov chain Monte Carlo approach containing three types of steps: changing the number of deleted points, the locations of the deleted points, and the values of the Gaussian process. The joint distribution over the fixed data, the number of thinned events, the location of the thinned events, and the function values at observed and thinned events can then be written down explicitly, and used to develop an MCMC procedure for the augmented posterior distribution, without the need to evaluate integrals of Gaussian processes. This approach appears to outperform the discretization approach of the previous subsection on smooth intensity functions.

4.2.3 Extensions to Higher Dimensions

Many of the methods for point processes on the line generalize to spatial processes. In some cases these extensions are non-straightforward, mainly concerning the lack of well-ordering of $\mathbb{R}^2$. A fairly recent review is Sect. 2.4 in [34]. Interesting applications include [55] who modeled a spatial pattern of badger territories and the distribution of pores in 3D translucent alumina using an inhomogeneous Poisson process with high intensity near the edges of an unobserved Voroni tessellation.

We have chosen not to focus on parametric rate models (which abound e.g. in the software reliability literature [26, 35]), since most of these are very similar to Bayesian models for iid data.

4.2.4 Matérn Thinning

Matérn introduced three different thinnings of Poisson processes in order to produce point processes that were more regular than the Poisson [42]. Type I simply deletes all pairs of points that are within a radius R of another point. This is perhaps the simplest hard core rejection model in the literature. Type II introduces independent uniform marks t_i, called times, to each of the original points. The point with the smallest mark among all neighbors within distance R is retained. Clearly this model would have a higher rate of points than type I. Matérn also considered a third, dynamic variant, which Huber and Wolpert call Type III [27], and which Matérn thought intractable. The retained points are called "seen", while the removed points are "hidden". In the type III process the seen points are those for which no seen point with lower time mark lies within distance R. So, for example, if we have three points with increasing times, such that the first is within R of the second, and the second is within R of the third, we have no points left in a type I process, only the first point left in the type II process, but potentially two points left in the type III process (see Fig. 1 in [27] for a graphical illustration).

In order to calculate the likelihood for a type III process, Huber and Wolpert used a technique akin to that used by [2] in the previous subsection. Specifically, they suggested starting with n seen points and parameters λ and R, and then draw hidden points from a Poisson process of rate λ, and draw time marks uniformly for both seen and hidden points, until for all hidden points there is a seen point within distance R and with smaller time mark. This has the drawback that it can take quite a long time if there is a large number of seen points. Define the *shadow* of a seen point configuration as the union of balls of radius R centered at each seen point cross the interval $(t_i, 1]$ containing the possible hidden points. Let $d\Lambda(x,t)$ be the joint intensity of a Poisson point at x with mark t. Then the density (with respect to a Poisson process with uniform marks) of a seen point pattern $\mathbf{x}$ with marks $\mathbf{t}$ becomes

$$\mathbf{1}\big(\rho(\mathbf{x}) > R\big)\lambda^n \exp(|S|(1-\lambda)) \exp\big(\Lambda(D(\mathbf{x},\mathbf{t}))\big) \tag{4.3}$$

where $\rho(\mathbf{x}) = \min_{i\neq j}(x_i, x_j)$ is the smallest inter-point distance, $D(\mathbf{x},\mathbf{t})$ is the shadow of $(\mathbf{x},\mathbf{t})$, and S is the observation window. It is straightforward to verify that the acceptance-rejection approach outlined above samples directly from the likelihood. A faster perfect simulation approach was also outlined, and has been expanded upon in [44].

The likelihood calculation can now form the basis for a Bayesian approach to estimating parameter of a Matérn type III process. To our knowledge this has not

yet been implemented elsewhere in the literature. If the point process includes a time component as is the case in the example below, the density in (4.3) can be calculated directly. Otherwise, a simulation method is needed for the latent time marks.

Example 4.1. (Comparison of cluster processes for precipitation models)
Hobbs and Locatelli described mesoscale rainfall activity in cyclonic storms roughly as follows [25]. Synoptic scale weather fronts contain large mesoscale regions, rainfall bands, where precipitation activity is possible. In turn, these bands contain moving rain cells, which are the points of higher rainfall rates. Observing this from a fixed point in space (e.g., a rain gauge), we see varying amounts of rainfall over time, with precipitation tending to come in clusters. Mathematically, Le Cam was first to suggest modeling rainfall at a location by a cluster point process [36], while Kavyas and Delleur proposed a Neyman-Scott Poisson cluster process, in which the primary process is a non-homogeneous Poisson process, and were the first to fit it to observed data [31]. In a sequence of papers in the 1980s, a variety of cluster process approaches were developed (a review is provided in [21], [54] discusses more recent work), usually made stationary by considering only a short time period each year, such as a month.

In most versions of cluster point process analysis of precipitation, the primary process is assumed unobserved. This may be reasonable if only rain gauge data is used. However, one would often be able to assess the arrival of weather fronts using different types of data. Guttorp used so-called event-based data from the MAP3S acid rain monitoring network to assess features of the secondary process [20]. This is the same data set that we will be using for our analysis, see Fig. 4.2. The MAP3S/PCN (Multistate Atmospheric Power Product Pollution Study / Precipitation Chemistry Network) network of nine monitoring stations in the northeastern United States was initiated in 1976. We will focus on station 1, located on Whiteface Mountain in New York, at an altitude of 610 meters. The data were obtained from the Battelle-Pacific Northwest Laboratories ADS (Acid Deposition System) data base. They are described in [17], and in [41]. The data were collected on an event basis, using samplers that open during precipitation, and close during dry periods. The definition of an event in the MAP3S network was left to the operator of the station; the Whiteface operator made a meteorologically based decision on what constitutes a new event.

For either station, each event may contain several precipitation incidents, indicated by separate lid openings. Since storm fronts do not arrive according to a Poisson process (the fronts are physically separate), we do not expect a Poisson cluster process to be an adequate description of precipitation. We thus perform a comparison of a homogeneous Poisson cluster model and the type III Matérn model described in the previous subsection. Here, we view the fall data from 1976 to 1982 as seven independent realizations of fall precipitation events at Whiteface Mountain.

In a Bayesian framework, Bayes factors offer a natural way of scoring models based on the evidence provided by the data [29]. Specifically, suppose that $p(\mathbf{x} \mid \theta, M)$ is the density function of the observed point pattern $\mathbf{x}$ under model M given the model-specific parameter vector θ. Let the prior density of θ (assumed

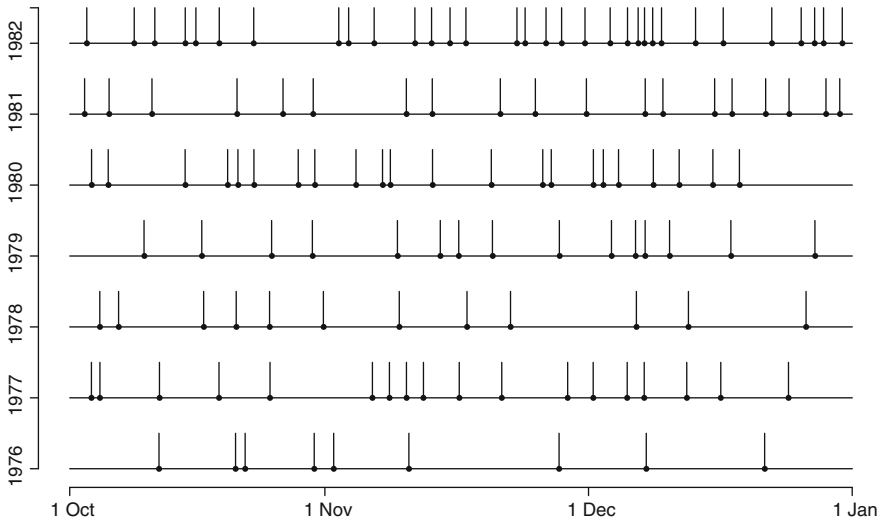

Fig. 4.2 Fall (October through December) precipitation events observed at Whiteface Mountain, New York, 1976–1982

to be proper) be given by $\pi(\theta)$. The marginal likelihood of $\mathbf{x}$ under model M is given by

$$m(\mathbf{x} \mid M) = \int p(\mathbf{x} \mid \theta, M)\pi(\theta \mid M)d\theta. \tag{4.4}$$

Two models, M_1 and M_2 may then be compared by calculating the Bayes factor

$$\mathrm{B}_{12}(\mathbf{x}) = \frac{m(\mathbf{x} \mid M_1)}{m(\mathbf{x} \mid M_2)}. \tag{4.5}$$

For our data set, the Matérn Type III density in (4.3) becomes

$$p(\mathbf{x} \mid \lambda, R, M_{\mathrm{Ma}}) = \mathbf{1}\big(\rho(\mathbf{x}) > R\big)\lambda^n \exp(7T + \lambda(nR - 7T)),$$

where $n = 127$ is the total number of observed points, $T = 92$ is the number of days in the observation period, and $\rho(\mathbf{x}) = 0.75$ is the minimum inter-point distance over all the seven realizations. Similarly, the density for the homogeneous Poisson process is given by,

$$p(\mathbf{x} \mid \lambda, M_{\mathrm{Po}}) = \lambda^n \exp(7T(1 - \lambda)).$$

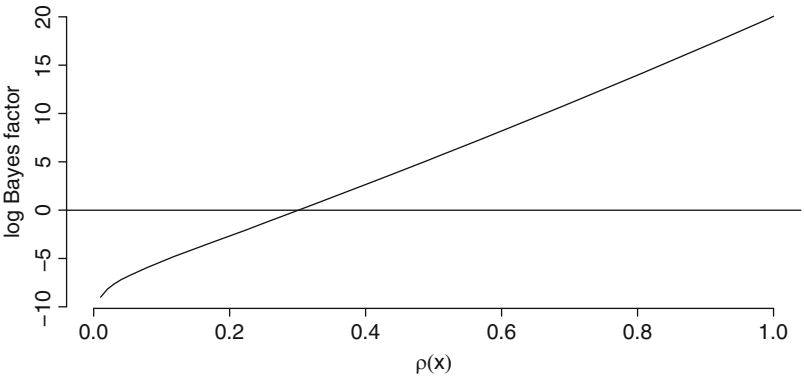

Fig. 4.3 The Bayes factor for comparing the Matérn Type III model and the homogeneous Poisson model for the Whiteface Mountain precipitation data, as a function of the minimum inter-point distance, $\rho(\mathbf{x})$. The Bayes factors are plotted on a log-scale; values greater than zero favor the Matérn Type III model

Here, we assume that $\lambda, R > 0$. We assign the parameter λ a conjugate prior density and set it to be exponential with rate parameter $\nu = 2$ in both models, while the parameter R in the Matérn Type III density is assigned a uniform prior on $(0, T)$, see Fig. 4.4. This choice of prior distributions allows for explicit calculation of the marginal likelihoods and hence, the Bayes factor. The Bayes factor for equiprobable models becomes

$$\mathrm{B}_{\mathrm{Ma,Po}}(\mathbf{x}) = \frac{(7T+\nu)}{Tn^2}\left(\left(\frac{7T+\nu}{7T+\nu-n\rho(\mathbf{x})}\right)^n - 1\right) = 273618, \tag{4.6}$$

which strongly favors the Matérn Type III model the latter being consistent with our hypothesis. As shown in Fig. 4.3, the value of the Bayes factor is highly dependent on the value of the minimum inter-point distance $\rho(\mathbf{x})$.

Based on the results above, we continue with the analysis of the Matérn Type III model only. The full conditional posterior distribution for λ is given by a $\Gamma(n+1, 7T+\nu-nR)$ distribution and

$$p(R \mid \mathbf{x}, \lambda, M_{\mathrm{Ma}}) = \mathbf{1}\big(0 < R < \rho(\mathbf{x})\big)\frac{\lambda n}{\exp(\lambda n \rho(\mathbf{x})) - 1}\exp(\lambda n R).$$

Figure 4.4 shows the posterior distributions for R and λ which are obtained from 50,000 simulations using a Gibbs sampler and the inverse transform. The posterior distributions are much sharper than the prior distributions and the posterior means are very close to the maximum likelihood estimates. The maximum likelihood estimates are given by $\hat{R} = 0.75$ and $\hat{\lambda} = 0.23$, while we obtain the posterior means $\tilde{R} = 0.74$ and $\tilde{\lambda} = 0.23$.

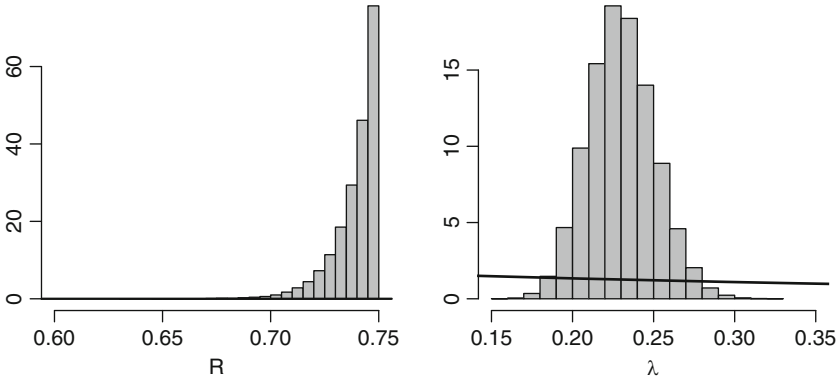

Fig. 4.4 Posterior distributions for the parameters R (*left*) and λ (*right*) in the Matérn Type III model for precipitation events at Whiteface Mountain. The respective prior distributions are denoted by solid black lines

4.3 Doubly Stochastic Processes

The doubly stochastic Poisson process, introduced by Cox in [9] and so named by Bartlett in [6] is obtained by letting the rate $\lambda(t)$ of the Poisson process vary according to a positive stochastic process, say $\Lambda(t)$. There are instances of doubly stochastic Poisson processes that are identical to cluster processes (for example, the shot noise process driven by a stationary Poisson process is identical to a Neyman-Scott Poisson cluster process, see p. 171-172 in [12]). It is worth noting here that, except when the rate process is determined by the scientific situation, it is difficult to analyze a doubly stochastic process without having repeated observations, since the model is indistinguishable from a non-homogeneous Poisson process based on a single path [46]. Thus, how you view your analysis can be seen as a matter of preference or convenience. We have not been able to find any Bayesian analyses of data where repeated observations are available, so that one can tell apart the doubly stochastic mechanism from the non-homogeneous Poisson process model.

Wolpert and Ickstadt modeled a spatial Poisson process with random intensity, where the intensity measure is a kernel mixture with a gamma measure [63]. As an example, they analyze the density and spatial correlation of hickory trees. The same data were also analyzed in Chap. 10.4 of [46] in a Bayesian setting using a non-homogeneous Poisson process with a log-Gaussian prior process, where the Gaussian process has constant mean β, variance σ^2, and an exponential correlation function with range parameter α. The hyper-parameters $\beta, \sigma^2, \kappa = \log(\alpha)$ need prior distributions as well. The authors used Jeffreys priors for the mean and variance, and a uniform prior between -2 and 4 for κ. As discussed in Sect. 4.2, a discrete approximation to the prior process was used. The analysis was very sensitive to the prior on κ, and compared to a frequentist method of moment analysis using the g-function, the Bayesian method indicates a substantially larger

correlation range. As pointed out above, this can also be viewed as a parametric Bayesian analysis of the doubly stochastic Poisson process obtained using a log Gaussian rate function.

In [19] a doubly stochastic Poisson process was considered with a gamma process (as in [39]) being its rate function $\Lambda(t)$. This process has parameter functions α (the rate function measure) and β (the scale process). In the case of constant scale $\beta = b$, the resulting process is what the authors call a negative binomial process of type 2. To perform a Bayesian analysis, they assigned a gamma process prior to the rate function measure α, and computed a closed form expression for the posterior distribution of α given the data. The authors did not view the distribution of the rate function $\Lambda(t)$ as a prior distribution.

In the highly influential paper on integrated nested Laplace approximations (INLA), the authors illustrated how their numerical alternative to Markov chain Monte Carlo methods can be applied to a doubly stochastic Poisson process where the intensity process is log Gaussian, although the method would work for any intensity process that is a positive function of a Gaussian process such that the resulting doubly stochastic Poisson process is valid. The calculation of over 20,000 marginal distributions, applied to the rain-forest data set also analyzed in [61], took four hours of computing time. Again, the Gaussian process was discretized to a fine grid. To get similar precision with MCMC methods would be prohibitive computationally. It is possible within INLA to calculate Bayes factors, as noted in Sect. 6.2 of the paper. However, the prior distributions used for the underlying random field are usually improper. The Bayes factor is thus only determined up to an unknown ratio of constants unless the parameters associated with the improper prior appear in all the models among which selection is performed.

4.4 Cluster Processes

The general cluster process consists of a primary process X of parent points τ_i, to each of which is associated a secondary point process $Z_i + \tau_i$ [12]. The structure of the primary and secondary processes is a suitable source of classification. Thus we can for example separate Poisson cluster processes (in which the primary process is Poisson) from general cluster processes (with a general primary process). On the line the most common secondary processes are of the Bartlett-Lewis type in which a random number of secondary points are laid out according to a renewal process, and the Neyman-Scott type where the secondary points are iid around the parent point (or cluster center). The named processes that abound in the literature (Cox cluster process, Matérn cluster process, Thomas process etc.) are simply special cases, and it does not seem useful to us to have a nomenclature which separates the particular distributional assumption. For example, we would call the (generalized) Thomas process a Poisson cluster process of Neyman-Scott type using a Poisson cluster size distribution and normally distributed dispersion. Most Poisson cluster

processes are non-Markovian; the exception being those with uniformly bounded cluster diameters [5].

Van Lieshout and Baddeley developed likelihood expressions for cluster processes with Poisson distributed offspring sizes, and developed Bayesian inference for processes where the prior distribution of the parent process is a Markov inhibition process [38]. Of course, one could assign a Poisson process or a Matérn type III prior, and the results would be very similar. The main tool is a Markov chain Monte Carlo algorithm that uses a birth and death process (or, in a special case, coupling from the past), and the techniques are applied to a classical data set.

McKeague and Loizeaux considered Neyman-Scott processes in the plane, and also used an inhibition process as prior on the parent process [43]. They used perfect sampling, and applied their tools to an example involving leukemia cases, where unobserved cluster centers are estimated to lie close to some hazardous waste sites.

The idea of self-exciting processes is to have the rate depend on the development of the model in the past. If this dependence can be written as a linear functional, there is an alternative representation of this process as a cluster process (see pp. 183–185 in [12]). Gamerman used a variant where the intensity is piecewise constant, but dependent on the events in the previous piece [15]. One could of course also think of this as a doubly stochastic model. Gamerman writes down equations for filtering and prediction as well as for Bayesian estimation of the rates in each interval.

A Neyman-Scott process with negative binomial cluster size distribution and truncated bivariate normal dispersion has been used to model minke whale populations [60]. The data are obtained from line transect samples, and are modeled as a random thinning of the cluster process. The parameter of interest is the product of the rate of the cluster centers and the mean cluster size, called the whale intensity. It is estimated using Markov chain Monte Carlo, even though the exact likelihoods are computable.

Example 4.2. (Modeling activation in the human brain)
Functional magnetic resonance imaging (fMRI) is a technique for non-invasive in vivo recording of brain activation. It is based on the different magnetic properties of oxygenated and deoxygenated haemoglobin; images obtained with the method show changing blood flow in the brain associated with neural activation. Figure 4.5 shows such data set, where the subject was not exposed to stimuli during the recording of the data. Despite the lack of specific stimuli, changes in the signal appear over time, some of which show covariation in different regions of the brain.

In [58] and [30], a Bayesian spatio-temporal point process model for such data was proposed. Purely spatial processes for this type of data have also been proposed [23, 57]. The activation is described by a marked point process Φ, where the point process is latent and corresponds to the unobserved neural activation while the marks are observed and describe the associated observed MR signal changes due to changes in the blood oxygenation level. It is thus the latent point process, Ψ, and the associated intensity function that are of main interest for the statistical analysis.

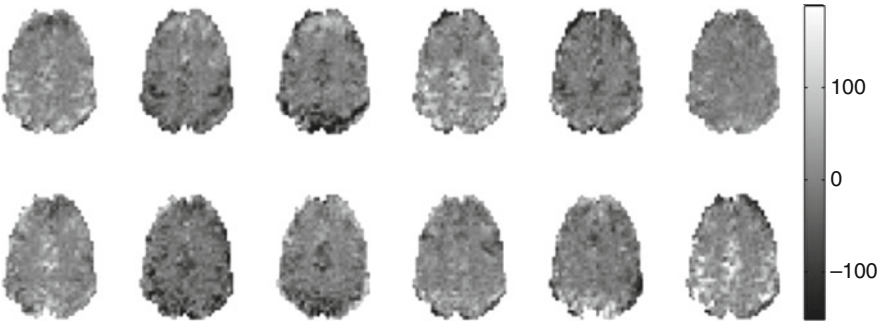

Fig. 4.5 Development of the magnetic resonance (MR) signal activity over time in a single slice through the human brain. From left to right and top to *bottom*: the activity at time $t = 12, 30, 48, \ldots, 210$ s. Note that the images shown here have been pre-processed to correct for movement related artifacts and the signal changes have been enhanced so that they can be observed with the naked eye. From [58]

Assume that we have observed data $\{Z_{tx}\}$, where $t \in [0, T]$ denotes time and $x \in \mathscr{X}$ denotes a spatial location, or a voxel, in the brain region $\mathscr{X}$ which is a bounded subset of $\mathbb{R}^2$ or $\mathbb{R}^3$. Here, $\mathscr{X}$ is a single slice through the brain, $\mathscr{X} \subset \mathbb{R}^2$. To account for edge effects in the time domain, we assume that Ψ is a process on $[T_-, T] \times \mathscr{X}$, where $T_- < 0$ is chosen such that it is very unlikely that a neural activation starting before time T_- will affect an observed MR signal on $[0, T]$. The marked process is denoted by $\Phi = \{[t_i, x_i; m_i]\}$ with $(t_i, x_i) \in \Psi$ and marks $m_i \in \mathbb{R}^d$.

The resulting model for the observed MR signal intensity at time t and voxel x becomes

$$Z_{tx} = \mu_x + \sum_i f_{tx}(t_i, x_i; m_i) + \sigma\varepsilon_{tx}, \tag{4.7}$$

where μ_x is the baseline signal at voxel x and ε_{tx} is an error term with mean 0 and variance 1. The function f_{tx} describes the contribution to the observed signal intensity at voxel x and time t caused by a neuronal activation at $(t_i, x_i) \in \Psi$. This function is assumed to be separable in space and time with $f_{tx}(t_i, x_i; m_i) = g(t - t_i; m_i)h(x - x_i; m_i)$ and $m_i = (\theta_{1i}, \theta_{2i}, \theta_{3i}) \in \mathbb{R}^3_+$, where

$$h(y; m) = \theta_1 \exp\left(-\frac{\|y\|^2}{2\theta_2}\right)$$

and

$$g(u; m) = \int_0^{\theta_3} \frac{1}{\sqrt{2\pi}3} \exp\left(-\frac{(u - v - 6)^2}{18}\right) dv.$$

Here, $\|\cdot\|$ denotes the Euclidean norm. The mark parameters thus have the following interpretation: θ_{1i} describes the magnitude of the signal change due to neural activation i, θ_{2i} describes the spatial extend of this change, and θ_{3i} its temporal duration.

For simplicity, assume that the marks m_i and the variance σ^2 are fixed. The aim of the statistical analysis is then to recover the latent point process Ψ and its intensity function based on the observations $\{Z_{tx}\}$ under the model described by (4.7). Further, we may replace Z_{tx} in (4.7) by $Z_{tx} - \overline{Z}_{\cdot x}$ and f_{tx} by $f_{tx} - \bar{f}_{\cdot x}$. The new data have $\mu_x = 0$ and the same correlation structure as the original data if the number of observed images is sufficiently large.

The prior distribution of Ψ is chosen as Poisson with intensity λ. There is thus no interaction between points in the prior distribution and any interactions found in the posterior distribution derive from interactions observed in the likelihood. The intensity function λ is assumed to be of the form

$$\lambda(t,x) = \sum_{k=1}^{K} \lambda_k \mathbf{1}\big(x \in \mathscr{X}_k\big), \tag{4.8}$$

where the sets $\mathscr{X}_k \subseteq \mathscr{X}$ are disjoint. Their union may be the whole observed brain region $\mathscr{X}$ but need not be. The sets $\mathscr{X}_k$ should be specified by the experimenter while the parameters λ_k are unknown. The intensity function can be written as $\lambda(t,x) = c\lambda_2(x)$, where $c > 0$ and $\int_{\mathscr{X}} \lambda_2(x)dx = 1$. It follows from (4.8), that λ_2 can be written as

$$\lambda_2(x) = \sum_{k=1}^{K} \pi_k \frac{\mathbf{1}\big(x \in \mathscr{X}_k\big)}{|\mathscr{X}_k|},$$

where $|\cdot|$ denotes area and $\pi_k > 0$ with $\sum_{k=1}^{K} \pi_k = 1$. The parameters $c, \pi = (\pi_1, \ldots, \pi_K)$ are assigned non-informative prior distributions.

The posterior density is of the form

$$p(c, \pi, \psi \mid z) \propto p(z \mid \psi)\, p(\psi \mid c, \pi)\, p(c)\, p(\pi),$$

where the likelihood is given by

$$p(z \mid \psi) = [2\pi\sigma^2]^{-n(z)/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{t,x} \left[z_{tx} - \sum_{(t_i,x_i)\in\psi} f_{tx}(t_i, x_i; m) \right]^2 \right).$$

A fixed scan Metropolis within Gibbs algorithm is used to simulate from the posterior density where in each scan c, π, and ψ are updated in turn. The full conditional distributions for c and π are given by a Gamma and a Dirichlet distribution, respectively, while the full conditional distribution for ψ is

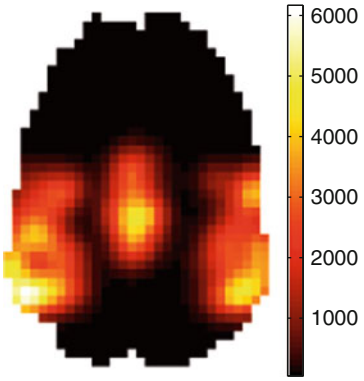

Fig. 4.6 The posterior spatial activation pattern in the three regions of interest cumulated over time. The three regions are the left and right motor cortex and a middle region. From [30]

$$p(\psi \mid c, \pi, z) \propto c^{n(\psi)} \prod_{k=1}^{K} \pi_k^{n_k(\psi)} \exp\left(-\frac{1}{2\sigma^2} \sum_{t,x} \left[z_{tx} - \sum_{(t_i, x_i) \in \psi} f_{tx}(t_i, x_i; m) \right]^2 \right), \tag{4.9}$$

where $n_k(\psi)$ denotes the number of points in ψ that fall within $\mathscr{X}_k$. Note that the full conditional distribution for ψ is in fact a pairwise interaction density. The point process ψ is simulated from the density in (4.9) using a Birth-Death-Move algorithm as described in [46].

Based on an earlier analysis of the same data set in [7], the prior intensity in (4.8) was set to be positive in three sub-regions of interest, the left and right motor cortex and a middle region. The resulting posterior spatial intensity pattern for ψ when cumulated over time is shown in Fig. 4.6. The posterior spatial intensity is clearly inhomogeneous in contrast to the homogeneous prior intensity with strong indications for clustering in the spatial domain.

4.5 Model Selection

Model selection for point process models is commonly carried out by investigating the summary statistics of the point pattern prior to the model fitting. Formal Monte Carlo tests of goodness-of-fit to the homogeneous Poisson process or comparison of the nearest-neighbor distance distribution function and the spherical contact distribution function can provide the modeler with evidence for regularity or clustering in the point pattern as compared to complete randomness [4, 28]. Such comparisons can produce important guidance for choosing the correct class of models, yet these model classes are very broad, rendering the information less valuable.

Statistical inference for point process models is usually very computationally intensive, and it is often not feasible to perform inference for a single data set under many different models. For this reason, scientific understanding of the data, combined with expert knowledge of the model class, is often combined to select a priori a single model for a given data set, once the appropriate class of models has been established. However, if the scientific question of interest relates to specific details in the modeling, such as particularities in the clustering mechanism of the point pattern, a more formal procedure for model comparison is called for.

The Akaike information criterion (AIC), which is given by

$$\mathrm{AIC} = -2 \log L + 2k, \tag{4.10}$$

where L is the maximum likelihood value and k is the number of parameters in the model, is by far the most popular model comparison criterion used in the point process literature. The AIC has the advantage that it can be applied to any likelihood based inference method. However, it has been noted that it tends to favor more complicated models for larger data sets [48]. This is a clear disadvantage in a setting where the modeling easily becomes computationally intractable. We discuss this issue further in Example 4.3.

Bayes factors (see Example 4.1) were first used in a point process context by Akman and Raftery, who compared parametric intensity models for non-homogeneous Poisson processes on the line [3]. The focus of their work was to develop conditions for which the Bayes factor could be determined under vague prior information. In this context, they call the Bayes factor $\mathrm{B}_{12}^{(n)}(\mathbf{x}, T)$ *operational* if for $U_n(T) = \{u = (u_1, \ldots, u_n) : 0 \le u_1 \le \cdots \le u_n \le T\}$, there exists a positive integer n such that

$$\sup_{T>0} \sup_{u \in U_n(T)} \mathrm{B}_{12}^{(n)}(u, T) < \infty.$$

Then, m, the smallest such integer, is the smallest number of observed events needed for a comparison of M_1 and M_2. Furthermore, if $\mathrm{B}_{12}^{(n)}(u, T)$ is a bounded function of u for each fixed n and T, and invariant to scale changes in the time variable,

$$\mathrm{B}_{12}^{(n)}(u, T) = \mathrm{B}_{12}^{(n)}(au, aT) \quad \forall a > 0,$$

for all n, u, T, then the Bayes factor is operational. It is thus, under fairly general conditions, sufficient to define the prior distributions such that the Bayes factor becomes time-invariant for it to be well defined. Akman and Raftery demonstrated this explicitly for log-polynomial intensity models.

Partial Bayes factors have been used for hypothesis testing to classify a point pattern as either a homogeneous Poisson pattern or a mixture of a homogeneous Poisson pattern and a hard-core Strauss process [62]. Here, the term *partial Bayes factor* refers to calculating the Bayes factor in (4.5) using a summary statistic $\mathbf{y}$ rather than the full data $\mathbf{x}$, as the marginal likelihood is intractable for the mixture

model considered in the study. The partial Bayes factor is equivalent to (4.5) if and only if $\mathbf{y}$ is a sufficient statistic for $\mathbf{x}$ under both M_1 and M_2.

To our knowledge, [3] and [62] are the only applications of Bayesian model selection criteria reported in the literature in the context of point process models such as those discussed in this paper. In Example 4.1, we showed how Bayes factors may be calculated directly for simple models. In the following example, we consider using a reversible jump algorithm for Bayesian model selection when direct calculation of the Bayes factor is not feasible.

Example 4.3. (Model selection for point processes of Neyman-Scott type)
Here, we compare using AIC and Bayes factors for model selection within the class of Neyman-Scott cluster processes. More precisely, we compare two different models of Neyman-Scott type which differ in the dispersion process for the secondary points. Model M_1 has a homogeneous Poisson cluster process, a Poisson cluster size distribution, and the dispersion distribution is given by a normal distribution. This model has also been called the modified Thomas process [14]. Model M_2, on the other hand, can be seen as a mixture of two such processes, where the dispersion variance differs for the two components of the mixture. Palm likelihood inference for M_1 and M_2 was considered in [56] and Prokešová and Jensen showed that the Palm likelihood estimator for these models is consistent and asymptotically normally distributed [50].

Model M_1 has a latent cluster center process and three unknown parameters: the intensity of the cluster process, κ, the mean cluster size, α, and the dispersion variance, ω^2. We generate ten samples from this model on $B = [0,1] \times [0,1]$ for $(\kappa, \alpha, \omega) = (50, 30, 0.03)$ and perform Palm likelihood inference and Bayesian inference for each sample under both model M_1 and M_2. The procedure is then repeated with data samples generated from model M_2. Model M_2 has two latent cluster center processes and five unknown parameters. We set the true parameters as $(\kappa_1, \kappa_2, \alpha, \omega_1, \omega_2) = (25, 25, 30, 0.02, 0.04)$, where κ_i is the intensity and ω_i^2 is the dispersion variance for cluster process $i = 1, 2$. Examples of such point patterns are shown in Fig. 4.7.

Bayesian inference for processes of similar type is discussed in e.g. [46], [47], and [60]. Contrary to the models considered in Example 4.1, we cannot calculate the marginal likelihood (4.4) of a dataset $\mathbf{x}$ under the models M_1 and M_2 directly. Instead, we define a reversible jump algorithm where we jump between the models M_1 and M_2 [18]. The Bayes factor can then be obtained directly from the MCMC sample by comparing the time spent in M_1 and the time spent M_2.

The random intensity function of M_2 is given by

$$\alpha Z(\xi \mid \Psi, \omega) = \alpha \left[\frac{1}{2\pi\omega_1^2} \sum_{c \in \Psi_1} \exp\left(-\frac{\|c - \xi\|^2}{2\omega_1^2} \right) + \frac{1}{2\pi\omega_2^2} \sum_{c \in \Psi_2} \exp\left(-\frac{\|c - \xi\|^2}{2\omega_2^2} \right) \right],$$

where $\omega = (\omega_1, \omega_2)$, $\Psi = (\Psi_1, \Psi_2)$ denotes the cluster center processes. To account for edge effects, we define the center processes on the extended window

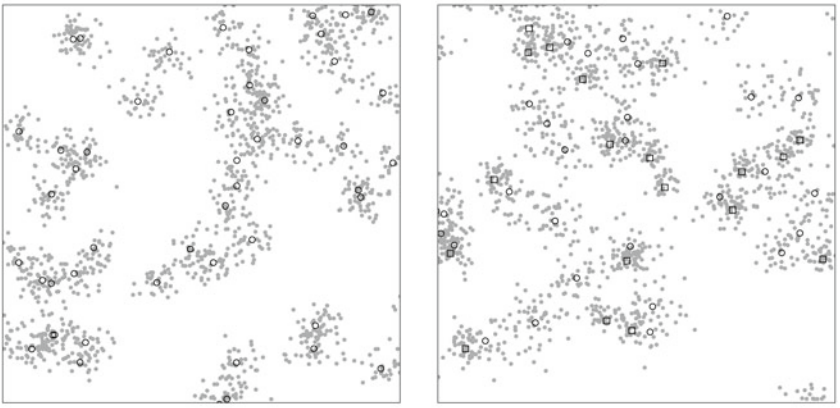

Fig. 4.7 Examples of simulated point patterns of Neyman-Scott type in the plane. *Left*: Poisson cluster center process with Poisson number of offsprings and normal dispersion process. *Right*: mixture of two such processes which differ in the dispersion variance. The observed point patterns are indicated with gray dots, while the black squares and circles indicate the latent cluster center processes

$B_{\text{ext}} = [-0.1, 1.1] \times [-0.1, 1.1]$. The density of a Poisson process on B with intensity function κ with respect to a homogeneous Poisson process X_1 with intensity λ is given by

$$p(\mathbf{x} \mid \kappa) = \exp\left(\lambda|B| - \int_B \kappa(\xi) d\xi\right) \prod_{\xi \in \mathbf{x}} \kappa(\xi).$$

As noted by [46] (p. 151), the choice of X_1 is not important for maximum likelihood inference and for MCMC simulations from a single model. However, when performing a reversible jump step between models with different number of latent processes, we need to choose λ with care in order to obtain balanced proposals, see below.

The joint posterior distribution of the latent processes and the parameters in M_2 is thus given by

$$p(\psi, \kappa, \alpha, \omega \mid \mathbf{x}) \propto p(\mathbf{x} \mid \alpha Z(\cdot \mid \psi, \omega)) p(\psi_1 \mid \kappa_1) p(\psi_2 \mid \kappa_2) p(\kappa) p(\alpha) p(\omega),$$

and the joint posterior distribution under M_1 is an obvious simplification. Our MCMC simulation algorithm consists of the following steps:

(a) Updating the latent process ψ;
(b) Updating the parameter κ;
(c) Updating the parameter α;
(d) Updating the parameter ω;
(e) Proposing to jump between M_1 and M_2.

Steps (a–d) are repeated 25 times under the same model between proposals to jump between models. For step (a), we use the Birth-Death-Move algorithm described in [46]. If we are currently in model M_2, we propose one change for each of the latent processes ψ_1 and ψ_2. We assign conjugate priors to the parameters κ and α which result in closed form full conditional distributions for these parameters. More precisely, we set $\kappa \sim \Gamma(50, 1)$, $\kappa_1, \kappa_2 \sim \Gamma(12.5, 0.5)$ and $\alpha \sim \Gamma(30, 1)$, where the gamma distributions are parameterized in terms of shape and rate. The full conditional distributions are then

$$\begin{aligned}
\kappa \mid \psi &\sim \Gamma(50 + n(\psi), 1 + |B_{\text{ext}}|) \\
\kappa_i \mid \psi_i &\sim \Gamma(12.5 + n(\psi_i), 0.5 + |B_{\text{ext}}|), \quad \text{for } i = 1, 2 \\
\alpha \mid \mathbf{x}, Z(\cdot \mid \psi, \omega) &\sim \Gamma\left(30 + n(\mathbf{x}), 1 + \int_B Z(\xi \mid \psi, \omega) d\xi\right).
\end{aligned}$$

A Metropolis-Hastings step is needed to update the dispersion parameter ω. We define the prior distribution for ω in terms of the precision and set $1/\omega^2 \sim \Gamma(1, 0.001)$. Under model M_2, we simulate initial values for ω_1 and ω_2 from the prior distribution until $\omega_1 < \omega_2$. This is needed for identifiability, as M_2 is otherwise invariant to permutations of the labels $i = 1, 2$. The joint prior distribution of (ω_1, ω_2) is thus 2 times the product of the individual prior components; this plays a role in the reversible jump step (e). To update the dispersion parameter ω under M_1, we generate a proposal $1/\omega^{2^*} \sim \Gamma(1/\omega^2, 1)$ and accept it with probability

$$\min\left\{\frac{p(\mathbf{x} \mid \alpha Z(\cdot \mid \psi, \omega^*) q(\omega \mid \omega^*)}{p(\mathbf{x} \mid \alpha Z(\cdot \mid \psi, \omega) q(\omega^* \mid \omega)}, 1\right\},$$

where $q(\omega^* \mid \omega)$ is the proposal density for ω^* given the current state of the chain. Under M_2, the parameters ω_1 and ω_2 are updated in a similar way. However, a proposal is rejected immediately if the condition $\omega_1 < \omega_2$ is violated by the proposal.

The reversible jump step (e) is similar to the reversible jump step for normal mixtures described in [52]. To move from M_2 to M_1 we need to merge the two cluster processes into one process. This is proposed by setting

$$\begin{aligned}
\psi^* &= \psi_1 \cup \psi_2 \\
\kappa^* &= \kappa_1 + \kappa_2 \\
\omega^* &= \sqrt{\frac{\kappa_1 \omega_1^2 + \kappa_2 \omega_2^2}{\kappa_1 + \kappa_2}}.
\end{aligned}$$

The reversible split move from M_1 to M_2 is now largely determined. There are two degrees of freedom involved in the split which we determine with a two-dimensional random vector u given by

$$u_1 \sim \text{Beta}(2, 2), \quad u_2 \sim \text{Beta}(2, 2).$$

Here, we set

$$\kappa^* = (\kappa_1^*, \kappa_2^*) = (u_1\kappa, (1-u_1)\kappa), \tag{4.11}$$

$$\omega^* = (\omega_1^*, \omega_2^*) = \left(\sqrt{\frac{u_2}{u_1}}\omega, \sqrt{\frac{1-u_2}{1-u_1}}\omega \right),$$

and reject the proposal immediately if $\omega_1^* < \omega_2^*$ does not hold. It still remains to allocate the points in ψ to either ψ_1^* or ψ_2^*. This is performed by allocating each point in ψ at random to either ψ_1^* with probability κ_1^*/κ or to ψ_2^* with probability κ_2^*/κ.

The acceptance probability for a split move is

$$\min\left\{ \frac{p(\psi^*, \kappa^*, \alpha, \omega^*|\mathbf{x})}{p(\psi, \kappa, \alpha, \omega|\mathbf{x})q(u)}|J|, 1 \right\},$$

where $q(u)$ is the density function of u and J is the Jacobian of the transformation described in (4.11),

$$|J| = \frac{\omega\kappa}{2\sqrt{u_1(1-u_1)}\sqrt{u_2(1-u_2)}}.$$

As mentioned above, we need to choose the densities of the latent cluster processes carefully in order to obtain balanced proposals. Let X, X_1, and X_2 be homogeneous Poisson processes on B_{ext} with intensities λ, λ_1, and λ_2, respectively, such that

$$\lambda = \frac{n(\psi)}{|B_{\text{ext}}|}, \quad \lambda_1 = \frac{n(\psi_1)}{|B_{\text{ext}}|}, \quad \lambda_2 = \frac{n(\psi_2)}{|B_{\text{ext}}|}.$$

The log-ratio of the density of (Ψ_1^*, Ψ_2^*) with respect to (X_1, X_2) and the density of Ψ with respect to X is then given by

$$\log\left(\frac{p(\psi_1^*|\kappa_1^*)p(\psi_2^*|\kappa_2^*)}{p(\psi|\kappa)} \right)$$
$$= n(\psi_1^*)\left[\log\frac{\kappa_1^*}{\kappa} - \log\frac{n(\psi_1^*)}{n(\psi)} \right] + n(\psi_2^*)\left[\log\frac{\kappa_2^*}{\kappa} - \log\frac{n(\psi_2^*)}{n(\psi)} \right],$$

which penalizes for a lack of balance between the proposed intensities and the corresponding point patterns. The acceptance probability for a merge move is calculated in a similar fashion. The algorithm was implemented in R [51].

The Palm likelihood inference is performed as described in [56], where the maximization is repeated five times for each sample using different starting values each time. We found that this was necessary, as different starting values would often give different results. The AIC in (4.10) is then calculated for each sample based on

Table 4.1 Model selection results for comparing M_1 and M_2 based on Akaike information criterion (AIC) and Bayes factors (BF) for simulated data. The table reports the classification results for each of the model selection criteria based on ten simulated data sets from each model.

Correct model	AIC		BF	
	M_1	M_2	M_1	M_2
M_1	8	2	10	0
M_2	3	7	0	10

the optimal result obtained over the five runs. The MCMC chain is run for 300,000 iterations over the steps (a–d). We assessed the convergence by running several such chains for each data set which give nearly identical results. The starting values for both inferenceo methods are set as

$$\kappa \sim \text{Po}(50), \quad \alpha \sim \text{Po}(30), \quad \frac{1}{\omega^2} \sim \Gamma(1, 0.001),$$

under M_1 and similar under M_2. For the Bayesian inference, the initial latent center processes are simulated from a Poisson model and the chain is started randomly in either M_1 or M_2. The Palm likelihood inference takes about 30–40 minutes on a standard desktop computer for a single data set. Running one MCMC chain takes about 1.5–2 hours on the same computer.

The results of the simulation study are reported in Table 4.1. In the Bayesian framework, all the MCMC chains would initially jump back and forth between the two models and then settle in the correct model. Under M_1, this initial burn-in period was very short, or only about 5,000 iterations. However, the mixing was slower under M_2, and about 100.000 iterations were needed before all the chains would settle in M_2. In the frequentist framework, a data set would be classified as belonging to either M_1 or M_2 based on the minimum AIC obtained for that data set. As Table 4.1 shows, 25% of the data sets were wrongly classified by this method. We did, however, not find any indications of the AIC preferring either the simpler or the more complicated model. Generally, though, we would obtain a much greater difference between the two AIC scores when M_2 was chosen as the correct model.

4.6 Summary

Statistical inference for point process models was initially performed in a frequentist manner, with the earliest work on Bayesian inference being published about three decades ago. In this paper, we have reviewed some Bayesian contributions for non-Markovian processes. Our aim was not to provide a complete literature review; rather, we have chosen to focus on those papers that we find especially important or interesting. In particular, we have tried to emphasize the variety of areas to which non-Markovian point process models have been applied.

We have further emphasized the use of Bayesian methodology for model selection. We show how Bayes factors can be used to determine model probabilities for simple models without performing a full inference under each model. For more complicated models, this is usually no longer the case. In an example, we show how a reversible jump algorithm can be used to determine model probabilities when the marginal likelihoods for the competing models cannot be computed directly. Traditionally, model selection methods for point processes mainly aim at detecting repulsion or clustering in the point pattern and there seems to be a lack of methods that apply beyond this initial distinction. The results presented here suggest that Bayesian methodology might be applied to fill this gap, although further research is needed.

References

1. Aalen, O.: Nonparametric inference for a family of counting processes. Ann. Stat. **6**(4), 701–726 (1978)
2. Adams, R.P., Murray, I., MacKay, D.J.C.: Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In Proceedings of the 26^{th} International Conference on Machine Learning, Montreal, Canada (2009)
3. Akman, V.E., Raftery, A.E.: Bayes factor for non-homogeneous Poisson processes with vague prior information. J. R. Stat. Soc. Ser. B **48**(3), 322–329 (1986)
4. Baddeley, A.: Modelling strategies. In Handbook of Spatial Statistics. Diggle, P. Guttorp, P., Fuentes, M., Guttorp, P. (Eds.) Chapmann & Hall/CRC, Boca Raton (2010)
5. Baddeley, A., Van Lieshout, M., Møller, J.: Markov properties of cluster processes. Adv. Appl. Prob. **28**(2), 346–355 (1996)
6. Bartlett, M.: The spectral analysis of point processes. J. R. Stat. Soc. Ser. B **25**(2), 264–296 (1963)
7. Beckmann, C.F., Deluca,M., Devlin, J.T., Smith, S.M.: Investigations into resting-state connectivity using independent component analysis. Philos. T. Roy. Soc. B **360**, 1001–1013 (2005)
8. Beneš, V., Bodlák, K., Møller, J., Waagepetersen, R. A case study on point process modelling in disease mapping. Image Anal. Sterol. **24**, 159–168 (2005)
9. Cox, D.R.: Some statistical methods connected with series of events (with discussion). J. R. Stat. Soc. Ser. B **17**, 129–164 (1955)
10. Cox, D. R., Lewis, P.A.: Statistical Analysis of Series of Events. Methuen, London (1966)
11. Cressie, N., Rathbun, S.L.: Asymptotic properties of estimators for the parameters of spatial inhomogeneous poisson point processes. Adv. Appl. Prob. **26**, 124–154 (1994)
12. Daley, D., Vere-Jones. D.: An Introduction to the Theory of Point Processes (2nd ed.). Volume I: Elementary Theory and Methods. Springer, Berlin (2005)
13. Diggle, P.: A kernel method for smoothing point process data. Appl. Stat. **34**, 138–147 (1985)
14. Diggle, P., Besag, J., Gleaves, J.T.: Statistical analysis of point patterns by means of distance methods. Biometrics **32**(3), 659–667 (1976)
15. Gamerman, D.: A dynamic approach to the statistical analysis of point processes. Biometrika **79**(1), 39–50 (1992)
16. Gelfand, A.E., Diggle, P., Fuentes, M., Guttorp, P.: Handbook of spatial statistics. Chapman & Hall/CRC, Boca Raton (2010)
17. Gentleman, R., Zidek, J.V., Olsen, A.R.: acid rain precipitation monitoring data base. Technical Report 19, Department of Statistics, University of British Columbia, Vancouver (1985)

18. Green, P.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82**(4), 711–732 (1995)
19. Gutiérrez-Peña, E., Nieto-Barajas, L.E.: Bayesian nonparametric inference for mixed Poisson processes. In J. M. Bernado, M. J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West (Eds.), Bayesian Statistics 7, 163–179. Oxford University Press, Oxford (2003)
20. Guttorp, P. Analysis of event-based precipitation data with a view toward modeling. Water Resour. Res. **24**(1), 35–43 (1988)
21. Guttorp, P.: Stochastic modeling of rainfall. In M. F. Wheeler (Ed.), Environmental Studies: Mathematical, Computational, and Statistical Analysis, 171–187. Springer, New York (1996)
22. Guttorp, P.: Simon Newcomb–astronomer and statistician. In C. C. Heyde and E. Seneta (Eds.), Statisticians of the Centuries, 197–199. Springer, New York (2001)
23. Hartvig, N.V.: A stochastic geometry model for functional magnetic resonance images. Scand. J. Stat. **29**, 333–353 (2002)
24. Heikkinen, J., Arjas, E.: Non-parametric Bayesian estimation of a spatial Poisson intensity. Scand. J. Stat. **25**(3), 435–450 (1998)
25. Hobbs, P.V., Locatelli, J.D.: Rainbands, precipitation cores and generating cells in a cyclonic storm. J. Atmos. Sci. **35**, 230–241 (1978)
26. Huang, Y.S., Bier, V.M,: A natural conjugate prior for the nonhomogeneous Poisson process with an exponential intensity function. Commun. Stat.–Theor. M. **28**(6), 1479–1509 (1999)
27. Huber, M.L., Wolpert, R.L. Likelihood-based inference for Matérn Type-III repulsive point processes. Adv. Appl. Prob. **41**, 958–977 (2009)
28. Illian, J., Penttinen, A., Stoyan, H., Stoyan, D.: Statistical analysis and modelling of spatial point patterns. Wiley, New York (2008)
29. Jeffreys, H.: Some tests of significance, treated by the theory of probability. P. Camb. Philis. Soc. **31**, 203–222 (1935)
30. Jensen, E.B.V., Thorarinsdottir, T.L.: A spatio-temporal model for functional magnetic resonance imaging data - with a view to resting state networks. Scand. J. Stat. **34**, 587–614 (2007)
31. Kavvas, M.L., Delleur, J.W.: A stochastic cluster model for daily rainfall sequences. Water Resour. Res. **17**, 1151–1160 (1981)
32. Kim, Y.: Nonparametric Bayesian estimators for counting processes. Ann. Stat. **27**(2), 562–588 (1999)
33. Kottas, A.: Dirichlet process mixtures of beta distributions, with applications to density and intensity estimation. In Proceedings of the Workshop on Learning with Nonparametric Bayesian Methods. 23rd ICML, Pittsburgh, PA (2006)
34. Kottas, A., Sanso, B.: Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. J. Stat. Plan. Inf. **37**(10, Sp. Iss. SI), 3151–3163 (2007)
35. Kuo, L., Yang, T.: Bayesian computation for nonhomogeneous Poisson processes in software reliability. J. Am. Stat. Assoc. **91**(434), 763–773 (1996)
36. Le Cam, L.M.: A stochastic description of precipitation. In J. Neyman (Ed.), Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability **3**, California, Berkeley, 165–186 (1960)
37. Lewis, P.A.W., Shedler, G.S.: Simulation of a nonhomogeneous Poisson process by thinning. Naval Logistics Quarterly **26**, 403–413 (1979)
38. Van Lieshout, M.N.M., Baddeley, A.J.: Extrapolating and interpolating spatial patterns. In A. B. Lawson and D. Denison (Eds.), Spatial Cluster Modelling, 61–86. Chapman & Hall/CRC, Boca Raton (2002)
39. Lo, A.Y.: Bayesian nonparametric statistical inference for Poisson point process. Z. Wahrscheinlichkeit **59**, 55–66 (1982)
40. Lo, A.Y., Weng, C.S.: On a class of Bayesian nonparametric estimates: II. hazard rate estimates. *Ann. I. Stat. Math. 41*(2), 227–245 (1989)
41. MAP3S/RAINE Research Community: The MAP3S/RAINE precipitation chemistry network: statistical overview for the period 1976-1979. Atmos. Environ. **16**, 1603–1631 (1982)

42. Matérn, B.: Spatial variation, **49** of Meddelanden frn Statens Skogsforskningsinstitut. Statens Skogsforskningsinstitut, Stockholm (1960)
43. McKeague, I.W., Loizeaux, M.: Perfect sampling for point process cluster modelling. In A. B. Lawson and D. Denison (Eds.), Spatial Cluster Modelling 87–107. Chapman & Hall/CRC, Boca Raton (2002)
44. Møller, J., Huber, M.L., Wolpert, R.L.: Perfect simulation and moment properties for the Matérn type III process. Stoch. Proc. Appl. **120**, 2142 – 2158 (2010)
45. Møller, J., Syversveen, A.R., Waagepetersen, R. Log Gaussian Cox processes. Scand. J. Stat. **25**, 451–482 (1998)
46. Møller, J., Waagepetersen, R.: Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall/CRC, Boca Raton (2004)
47. Møller, J., Waagepetersen, R.: Modern statistics for spatial point processes. Scand. J. Stat. **34**(4), 643–684 (2007)
48. Ogata, Y.: Seismicity analysis through point-process modeling: a review. Pure Appl. Geophys. **155**, 471–507 (1999)
49. Peeling, P., Li, C.F., Godsill, S.: Poisson point process modeling for polyphonic music transcription. J. Acoust. Soc. Am. **121**(4), EL168–EL175 (2007)
50. Prokešová, M., Jensen, E.B.V.: Asymptotic Palm likelihood theory for stationary point processes. Thiele Research Report nr. 2/2010, Department of Mathematical Sciences, University of Aarhus, Aarhus (2010)
51. R Development Core Team R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0 (2009)
52. Richardson, S., Green, P.J.: On Bayesian analysis of mixtures with an unknown number of components. J. R. Stat. Soc. Ser. B **59**(4), 731–792 (1997)
53. Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B **71**, 319–392 (2009)
54. Salim, A., Pawitan, Y.: Extensions of the Bartlett-Lewis model for rainfall processes. Stat. Model. **3**, 79–98 (2003)
55. Skare, Ø., Møller, J., Jensen, E.B.V.: Bayesian analysis of spatial point processes in the neighbourhood of Voronoi networks. Stat. Comp. **17**, 369–379 (2007)
56. Tanaka, U., Ogata, Y., Stoyan, D.: Parameter estimation and model selection for Neyman-Scott point processes. Biometrical J. **49**, 1–15 (2007)
57. Taskinen, I.: Cluster priors in the Bayesian modelling of fMRI data. Ph. D. thesis, Department of Statistics, University of Jyväskylä, Jyväskylä. (2001)
58. Thorarinsdottir, T.L., Jensen, E.B.V.: Modelling resting state networks in the human brain. In R. Lechnerová, I. Saxl, and V. Beneš (Eds.), *Proceedings S^4G: International Conference on Stereology, Spatial Statistics and Stochastic Geometry*, Prague (2006)
59. Waagepetersen, R.: Convergence of posteriors for discretized log Gaussian Cox processes. Stat. Probabil. Lett. **66**(3), 229–235 (2004)
60. Waagepetersen, R., Schweder, T.: Likelihood-based inference for clustered line transect data. J. Agric. Biol. Envir. S. **11**(3), 264–279 (2006)
61. Waagepetersen, R.P.: An estimating function approach to inference for inhomogeneous Neyman-Scott processes. Biometrics **63**, 252–258 (2007)
62. Walsh, D.C., Raftery, A.E.: Classification of mixtures of spatial point processes via partial Bayes factors. J Comput. Graph. Stat. **14**, 139–154 (2005)
63. Wolpert, R.L. Ickstadt, K.: Poisson/Gamma random field models for spatial statistics. Biometrika **85**(2), 251–267 (1998)

Chapter 5
A Review on Spatial Extreme Modelling

Jean-Noël Bacro and Carlo Gaetan

Abstract In this chapter we review recent advances in modelling spatial extremes. After a brief illustration of the extreme value theory for univariate and multivariate values, we concentrate on spatial max-stable processes. Statistical inference and simulation for this processes are subject of a close examination. Max-stable processes are also contrasted with spatial hierarchical models. The review ends with summarizing some open problems.

5.1 Introduction

Applications of extreme value methods mainly concern environmental and climate processes. During this last thirty years, the extreme value theory has developed probabilistic models and inferential procedures to characterize and work with extreme events. Such events are, almost by definition, rare and unexpected. Univariate methods have been first used to describe the extremal behaviour at a particular location in space and extreme modelling for single stationary time series is well-established (see, for example, [3, 23] and the references therein). The multivariate extreme value theory offers various notions to capture the main characteristics of the underlying dependence structure [27] but has to be extended to characterize extremal behaviour of spatial processes. Recently, major efforts are concentrated on developing models and methods for extreme values taking into account the more complex structure of spatial phenomena. If standard spatial statistics methods

J.-N. Bacro (✉)
I3M, Université Montpellier II, 34095 Montpellier, Cedex 5, France
e-mail: bacro@math.univ-montp2.fr

C. Gaetan
Dipartimento di Scienze Ambientali, Informatica e Statistica, Università Ca' Foscari-Venezia, I-30121 Venezia, San Giobbe, Cannaregio 873
e-mail: gaetan@unive.it

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_5,

perform well for the inference on the mean behaviour of spatial processes, they are clearly inappropriate when dealing with extreme realizations. Owing to the extreme nature of the considered events, the class of max-stable processes plays a central role and deserves special attention. Introduced by [20] as an extreme analogy of Gaussian processes, max-stable processes have been first considered by [10, 13, 63] for spatial applications. More recently a hierarchical approach has been proposed [9, 16, 29] in order to model spatial dependence. Both approaches have relative merits that we will discuss in some extent later. A comparison of both can be found in [19].

The paper is organized as follows. First in Sect. 5.2 we recall some results on univariate and multivariate extreme value theory. Sect. 3 deals with max-stable processes. Classic spectral representations from [57] as well as Brown-Resnick ones from [39] are detailed. Statistical inference for spatial max-stable processes is presented in Sect. 5.4. Estimation methods for the extremal coefficient function and for the model parameters are reviewed. The simulation of max-stable processes is considered in Sect. 5.5. Well established unconditional simulation techniques are described there. Instead for the conditional case we outline a recent method that copes with the prediction for max-stable processes. The hierarchical approach offers an alternative approach for modelling spatial dependence and its merits are discussed in Sect. 5.6. We end the chapter discussing some open problems in modelling spatial and spatio-temporal extreme data.

5.2 Extreme Values Theory for Univariate and Multivariate Data

5.2.1 Univariate Extremes

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables having a distribution function $F(\cdot)$. Consider $M_n = \max(X_1, \ldots, X_n)$ and assume that there exist sequences of positive constants $(a_n)_{n\geq 0}$ and real constants $(b_n)_{n\geq 0}$ such that $(M_n - b_n)/a_n$ converges as $n \to \infty$ to a non-degenerate distribution function $G(\cdot)$. Then $G(\cdot)$ is the generalized extreme value (GEV) distribution [25, 33, 36]

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \tag{5.1}$$

where a_+ denotes $\max(a, 0)$, $-\infty < \xi < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$. Here μ is a location parameter, σ is a scale parameter and ξ is a shape parameter. The case $\xi = 0$ is obtained by taking the limit as $\xi \to 0$. If $(M_n - b_n)/a_n$ converges to a

GEV distribution $G(\cdot)$, $F(\cdot)$ is said to be in the domain of (max) attraction of $G(\cdot)$ (usual notation: $F \in \mathscr{D}(G)$).

An alternative characterization of extreme value behaviour comes from modelling exceedances of high threshold and uses the point process theory [18, 51, 52].

Let $G(\cdot)$ be a GEV distribution and assume $F \in \mathscr{D}(G)$. Consider the sequence of point processes $(P_n)_{n\geq 1}$ on $\mathbb{R}^2$ defined as

$$P_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right), i = 1, \ldots, n \right\}.$$

Due to the scaling, all but the most extreme values become scaled to the lower endpoint of the rescaled distribution. On region $A_x = \{(0,1) \times (x, \infty)\}$, $(P_n)_{n\geq 1}$ can be approximated by a Poisson process with intensity measure

$$\Lambda\{(0,1) \times (x, \infty)\} = \left[1 + \xi \frac{x-\mu}{\sigma}\right]_+^{-1/\xi}.$$

The resulting distribution of the maximum for large n is a GEV distribution

$$\mathrm{P}\left((M_n - b_n)/a_n \leq x\right) = \mathrm{P}(\text{no points in } A_x) \approx \exp\left\{ -\left[1 + \xi \frac{x-\mu}{\sigma}\right]_+^{-1/\xi} \right\}.$$

In the applications, the point process model is often parametrized in order to work on maxima over blocks such as months, years .... If n observations are available corresponding to m blocks of data, the intensity measure is written as

$$\Lambda\{(a,b) \times (x, \infty)\} = (b-a)m \left[1 + \xi \frac{x-\mu}{\sigma}\right]_+^{-1/\xi}$$

to ensure that the parameters μ, σ, ξ correspond to the GEV distribution of the block maxima [62]. If n_b is the number of data in one block, i.e. $m = n/n_b$, then the distribution of the block maximum $M_{n_b} = \max(X_1, \ldots, X_{n_b})$ can be computed as

$$\mathrm{P}(M_{n_b} \leq x) \approx \exp\left\{-\Lambda\left((0, n_b/n) \times (x, \infty)\right)\right\}.$$

Threshold exceedances are modelled by the generalized Pareto distribution (GPD)

$$H(x) = 1 - \left(1 + \xi \frac{x}{\tilde{\sigma}}\right)_+^{-1/\xi},$$

where ξ is a shape parameter and $\tilde{\sigma} > 0$ is a scale parameter. In this approach, all observations over a high threshold, say u, are considered. From Pickands's result [51], it is well known that the conditional distribution $\mathrm{P}(X - u \leq y | X > u)$ can be approximated by a GPD having the same shape parameter ξ as the GEV distribution related to the maximum of X. Such results can be deduced following the point

process approach. If (T_i, X_i) designates a point of the process such that $X_i > u$, then, for large u,

$$\mathrm{P}(X_i > x | X_i > u) \approx \left\{ \frac{1 + \xi(x-\mu)/\sigma}{1 + \xi(u-\mu)/\sigma} \right\}_+^{-1/\xi} = \left\{ 1 + \frac{\xi(x-u)}{\tilde{\sigma}} \right\}_+^{-1/\xi}$$

with $\tilde{\sigma} = \sigma + \xi(u - \mu)$. This corresponds to the usual representation using the GPD, specifying the scale parameter $\tilde{\sigma}$ in terms of the GEV parameters and the threshold u.

5.2.2 Multivariate Extremes

The multivariate extreme value (MEV) theory focuses on component-wise maxima. Unlike the univariate case, the class of MEV distributions does not have a parametric representation but there are considerable restrictions on the dependence structure between marginal distributions. From a probability point of view, MEV distributions are well characterized [30, 53]. More precisely, let $(X_{i,1}, \ldots, X_{i,p})'$, $i = 1, 2, \ldots$ be a sequence of p-dimensional i.i.d. random vectors and $(M_{n,1}, \ldots, M_{n,p})'$ the vector of component-wise maxima, $M_{n,j} = \max\{X_{1,j}, \ldots, X_{n,j}\}$. If we assume that for suitable normalizing sequences, $(a_{n,j})_{n \geq 1} > 0$, $(b_{n,j})_{n \geq 1}$, the marginal limit distribution for each univariate component exists, then the limit distribution of the normalized vector

$$\left(\frac{M_{n,1} - b_{n,1}}{a_{n,1}}, \ldots, \frac{M_{n,p} - b_{n,p}}{a_{n,p}} \right) \tag{5.2}$$

can be characterized. By construction, each component of (5.2) must have a limit distribution of GEV type. Therefore, there is no loss of generality in keeping a specific marginal distribution since the other ones can be obtained through marginal transformations. Usually the results are stated assuming that marginal distribution $F_{X_j} \in \mathscr{D}(G_{X_j})$, where G_{X_j} is the unit Fréchet marginal distribution, i.e. $G_{X_j}(x) = \exp(-1/x)$ for $x > 0$.

There exist two fundamental representations for a MEV distribution [52, 53].

- A multivariate distribution G is a limit distribution of the random vector (5.2) with unit Fréchet margins if and only if

$$G(x_1, \ldots, x_p) = \exp(-V(x_1, \ldots, x_p)) \tag{5.3}$$

for $(x_1, \ldots, x_p)' \in \mathbb{R}_+^p$, where

$$V(x_1, \ldots, x_p) = \int_{S_p} \max_{j=1,\ldots,p} \left(\frac{w_j}{x_j} \right) dH(w_1, \ldots, w_p)$$

for a positive measure H defined on the $(p-1)$-dimensional unit simplex, S_p, such that $\int_{S_p} w_j dH(w_1, \ldots, w_p) = 1$, $j = 1, \ldots, p$.

- Any MEV distribution G with unit Fréchet margins can be written as

$$G(x_1, \ldots, x_p) = \exp\left(-A(w_1, \ldots, w_p) \sum_{j=1}^{p} x_j^{-1}\right) \tag{5.4}$$

where $w_j = x_j^{-1} / \sum_{k=1}^{p} x_k^{-1}$, $j = 1, \ldots, p$, and $A(\cdot)$ is a convex function on S_p satisfying $\max(w_1, \ldots, w_p) \leq A(w_1, \ldots, w_p) \leq 1$, for all $(w_1, \ldots, w_p)' \in S_p$.

The functions $V(\cdot)$ and $A(\cdot)$ are known as the *exponent measure* [53] and *dependence function* [52], respectively. Note that $A(\cdot)$ is often written as a function of $p-1$ arguments which sum to unity. For given GEV marginal distributions, the class of distributions given by (5.3) is referred to as the family of MEV distributions. In the sequel, for simplicity, we will focus on bivariate extreme value distributions for (X, Y). In that case, representation (5.4) is equivalent to MEV distribution but this is not generally the case [69]. For $p = 2$, we can rewrite (5.3) as

$$G(x, y) = \exp\left\{-\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w)\right\}, \tag{5.5}$$

where $H(\cdot)$ is such that $\int_0^1 dH(w) = 2$, $\int_0^1 w dH(w) = 1$. Instead formula (5.4) becomes

$$G(x, y) = \exp\left\{-\left(\frac{1}{x} + \frac{1}{y}\right) A\left(\frac{x}{x+y}\right)\right\} \tag{5.6}$$

where $A(0) = A(1) = 1$, $-1 \leq A'(0) \leq 0$, $0 \leq A'(1) \leq 1$, $A''(w) \geq 0$ and $\max(w, 1-w) \leq A(w) \leq 1$, $0 \leq w \leq 1$. The dependence function $A(\cdot)$ is related to the measure $H(\cdot)$ through

$$A(w) = \int_0^1 \max(w(1-u), (1-w)u) dH(u).$$

The functions $A(\cdot)$ and $H(\cdot)$ can deal with the total dependence case, i.e. $X = Y$ with probability one,

$$G(x, y) = \exp\{-\max(x^{-1}, y^{-1})\}$$

as well as with the independence case, i.e.

$$G(x, y) = G_X(x) G_Y(y) = \exp\{-(x^{-1} + y^{-1})\}.$$

In fact, choosing $A(w) = 1$ or $H(0.5) = 2$ and $A(w) = \max(w, (1-w))$ or $H(0) = H(1) = 1$ leads, respectively, to total dependence and independence.

Several parametric models for multivariate extremes have been introduced [11, 12, 69] especially for the bivariate case [11, 12, 14, 34, 35, 37, 68]. For example, the so-called logistic-Gumbel model

$$G(x, y) = \exp\left\{-(x^{-1/\alpha} + y^{-1/\alpha})\right\}^{\alpha},$$

$x > 0, y > 0$, $0 < \alpha \leq 1$ has gained a lot of favour since it can cover the total dependence case ($\alpha \to 0$) and the independence ($\alpha = 1$) one only through one parameter. As in dimension one, a notion of domain of attraction has been introduced. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sequence of i.i.d. random vectors having distribution function $F(\cdot, \cdot)$. As previously, let $(M_{1,n}, M_{2,n})$ designate the corresponding vector of component-wise maxima. Assume that both X_i and Y_i are unit Fréchet distributed and let $G(\cdot, \cdot)$ be a bivariate extreme value law with unit Fréchet margins. Then, $F \in \mathscr{D}(G)$ if

$$\mathrm{P}\left(M_{1,n}/n \leq x, M_{2,n}/n \leq y\right) \overset{n\to\infty}{\longrightarrow} G(x, y)$$

When $G(x, y) = \exp\{-(x^{-1} + y^{-1})\}$, the random variables X and Y are said to be asymptotically independent. Independence implies asymptotic independence but the converse is false.

5.3 Spatial Max-Stable Processes

Max-stable processes arise from an infinite-dimensional generalisation of multivariate extreme value theory. Let $Z(\cdot)$ be a stochastic process with non-degenerate marginals over an index set S. From [20], $Z(\cdot)$ is a max-stable process if all of its N-dimensional distributions satisfy the max-stability property: for each $n \geq 1$, there exists sequences $a_n(s) > 0$, $b_n(s)$, with $s \in S$, such that for any subset $D = \{s_1, \ldots, s_N\}$ of S

$$\mathrm{P}\left(\frac{Z(s_1) - b_n(s_1)}{a_n(s_1)} \leq z_1, \ldots, \frac{Z(s_N) - b_n(s_N)}{a_n(s_N)} \leq z_N\right)^n = G(z_1, \ldots, z_N).$$

Here $G(z_1, \ldots, z_N)$ is a MEV distribution. Choosing $S \subset \mathbb{R}^d$ for some $d \geq 1$ allows us to consider general processes such as time, spatial or spatio-temporal ones. All the marginal distributions of $Z(\cdot)$ are GEV ones and univariate marginals can always be considered as Fréchet ones up to a rescaling and shift. A max-stable process with unit Fréchet marginal distributions is usually referred as a *simple* max-stable process.

Currently, we can recognize two main approaches for constructing such processes. The first approach relies on (mixed) moving maxima processes [22, 57, 63] whereas the second one involves maxima on stochastic processes with the same correlation structure [39, 57].

In an unpublished paper, [63] exploits a representation of a continuous time max-stable process given by [20] and introduces the following max-stable process

$$Z(s) = \max_{i \geq 1} X_i f(s, U_i),$$

where $\{X_i, U_i\}_{i \geq 1}$ is a realization of a Poisson process on $(0, \infty) \times S_U$ with intensity measure $x^{-2} dx \, \nu(du)$. Here $f(\cdot)$ is a non-negative function on $S \times S_U$, with finite integral with respect a positive measure ν on S_U.

Under the constraint $\int_{S_U} f(s, u) \nu(du) = 1$, we get that $Z(\cdot)$ process is simple. In that paper, Smith gives an intuitive physical interpretation in terms of rainfall-storms, i.e. X_i and U_i are, respectively, the size and the type of the 'storm' i.

In general it is difficult to derive a closed form expression of the multivariate distribution of max-stable processes, apart from the bivariate case.

If $S_U = S$ and $f(s, u) = \phi_{u, \Sigma}(s)$ is the d-variate normal density with vector mean u and covariance matrix Σ, [63] derives the expression of the bivariate distribution of $Z(s)$ and $Z(s + h)$

$$\mathrm{P}(Z(s) \leq z_1, Z(s + h) \leq z_2) = \exp\left\{ -\frac{1}{z_1} \Phi\left(\frac{a(h)}{2} + \frac{1}{a(h)} \log \frac{z_2}{z_1} \right) - \frac{1}{z_2} \Phi\left(\frac{a(h)}{2} + \frac{1}{a(h)} \log \frac{z_1}{z_2} \right) \right\}, \qquad h \in \mathbb{R}^d, \tag{5.7}$$

where $\Phi(\cdot)$ is the standard normal distribution function and $a(h) = \sqrt{h' \Sigma^{-1} h}$.

This representation is known as *Gaussian extreme value process* and is used to model environmental data such as extremes of rainfall [10, 13] or extremes of wind speed [14]. In [61] the covariance matrix Σ is allowed to vary across S.

Different choices of function $f(\cdot)$ are also possible. [22] derived the expression of the bivariate distributions when $f(\cdot)$ is a Student and a Laplace multivariate density.

The initial proposal of Smith is extended by [57], introducing a random shape $Y(\cdot)$ instead of the deterministic function $f(\cdot)$. The Schlather's formulation is the following:

$$Z(s) = \max_{i \geq 1} X_i Y_i(s - U_i), \tag{5.8}$$

where $\{X_i, U_i\}_{i \geq 1}$ is a realization of a Poisson process on $(0, \infty) \times \mathbb{R}^d$ with intensity measure $x^{-2} dx \times \mu^{-1} du$ and $\{Y_i(\cdot)\}_{i \geq 1}$ are independent copies of a non-negative random function $Y(\cdot)$, defined on $\mathbb{R}^d$, such that $\mu = \mathbb{E}[\int_S Y(s) ds] \in (0, \infty)$. The Schlather's extension keeps the previous physical interpretation as illustrated in [42]. These authors term the process (5.8) *storm process* and derive its multivariate distribution for different choices of the random functions Y, that specifies the spatial observation support.

[57] has also introduced a class of max-stable process that is based on stationary stochastic processes with finite expectation. Let $\{X_i\}_{i \geq 1}$ designate a realization of

a Poisson process on $(0, \infty)$ with intensity measure $\mu^{-1}x^{-2}dx$ and $\{W_i(\cdot)\}_{i\geq 1}$ be independent copies of $W(\cdot)$, a stationary stochastic process on $\mathbb{R}^2$, with $\mu = \mathbb{E}(\max\{0, W(0)\}) \in (0, \infty)$. The stochastic process $Z(\cdot)$,

$$Z(s) = \max_{i\geq 1} X_i W_i(s), \tag{5.9}$$

is a stationary simple max-stable process. The multivariate distribution of $Z(\cdot)$ for any subset $\{s_1, \ldots, s_m\}$ of S is

$$\mathrm{P}\left(Z(s_1) \leq z_1, \ldots, Z(s_m) \leq z_m\right) = \exp\left\{-\mathbb{E}\left(\sup_{1\leq i\leq m} \frac{W(s_i)}{z_i}\right)\right\}.$$

When the Poisson process has intensity measure $\sqrt{2\pi}x^{-2}dx$ and $W(\cdot)$ is a zero-mean stationary Gaussian process, with unit variance and correlation function $\rho(\cdot)$, the max-stable process (5.9) is known as the *extremal Gaussian process* [57]. The bivariate distribution of an extremal Gaussian process is given by:

$$\mathrm{P}(Z(s) \leq z_1, Z(s+h) \leq z_2) = \exp\left\{-\frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right) \times \left(1 + \sqrt{1 - 2(\rho(h)+1)\frac{z_1 z_2}{(z_1+z_2)^2}}\right)\right\}.$$

A distinctive feature of extremal Gaussian process is that the independence between pairs of observations does not occur at any distance (see also Sect. 5.4). For the applications, such a property could be questionable and attempts to modify this property can be found in the literature. A first model defined as a mixture of the Schlather model and an asymptotically independent stochastic process is proposed [54]. This model, sometimes called the *independent Schlather model*, is of limited interest in a spatial context and [19] propose another instance of (5.9), the *geometric Gaussian process*. For such max-stable process, $W(\cdot)$ is a log-Gaussian stochastic process

$$W(s) = \exp\left\{\sigma\epsilon(s) - \frac{\sigma^2}{2}\right\}$$

where $\epsilon(\cdot)$ is a zero-mean stationary Gaussian process, with variance σ^2 and correlation function $\rho(\cdot)$. Actually the geometric Gaussian process is an instance of a two-dimensional *Brown-Resnick process* [6, 39], see below, and its bivariate distributions are the same as (5.7) provided that $a(h) = \sqrt{2\sigma^2(1-\rho(h))}$ [54]. In this case, the dependence can range from the complete one to the independence, according to the $\rho(\cdot)$ value.

([21], Corollary 9.4.5) introduce a representation for simple max-stable processes on $S = [0, 1]$ which is closely related to (5.9). In their representation, $\{W_i(\cdot)\}_{i\geq 1}$

are independent copies of a positive stochastic process $W(\cdot)$ in $C[0,1]$, the space of continuous functions on $[0,1]$ equipped with the supremum norm. The authors assume also that $\mathbb{E}(W_i(s)) = 1$ for all $s \in [0,1]$ and $\mathbb{E}(\sup_{s\in[0,1]} W(s)) < \infty$. This representation has been used in [7] for simulating extreme rainfalls.

Historically, the first example of max-stable process in the literature has been the Brown-Resnick process on a line [6]

$$Z(s) = \max_{i\geq 1} X_i \exp\left\{\varepsilon_i(s) - \frac{|s|}{2}\right\}, \qquad s \in \mathbb{R},$$

where $\{X_i\}_{i\geq 1}$ is a realization of a Poisson process on $(0,\infty)$ with intensity $x^{-2}dx$, $\{\varepsilon_i(\cdot)\}_{i\geq 1}$ are independent copies of a Brownian motion on the real line and we assume that $\{X_i\}_{i\geq 1}$ is independent from $\{\varepsilon_i(\cdot)\}_{i\geq 1}$. Note that the process $W(s) = \exp\{\varepsilon(s) - |s|/2\}$ meets the conditions required in ([21], Corollary 9.4.5).

[39] propose a natural generalization of the Brown-Resnick process by replacing the Brownian motion on the real line by a more general Gaussian process on $\mathbb{R}^d$, $d \geq 2$, namely

$$Z(s) = \max_{i\geq 1} X_i \exp\left\{\varepsilon_i(s) - \frac{\sigma^2(s)}{2}\right\} \tag{5.10}$$

where $\varepsilon_i(\cdot)$ are independent copies of a Gaussian process $\varepsilon(\cdot)$ with stationary increments, variance $\sigma^2(\cdot)$ and variogram $2\gamma(\cdot)$. This simple max-stable process has bivariate distribution given by

$$\mathrm{P}(Z(s+h) \leq z_1, Z(s) \leq z_2) = \exp\left\{-\frac{1}{z_1}\Phi\left(\frac{\sqrt{2\gamma(h)}}{2} + \frac{\log(z_1/z_2)}{\sqrt{2\gamma(h)}}\right) - \frac{1}{z_2}\Phi\left(\frac{\sqrt{2\gamma(h)}}{2} + \frac{\log(z_2/z_1)}{\sqrt{2\gamma(h)}}\right)\right\}.$$

The Brown-Resnick process is stationary although the Gaussian process $\varepsilon(\cdot)$ may not be. Moreover, the variogram $2\gamma(\cdot)$ characterizes the Brown-Resnick process and careful choices [39] for it allow us to end up with the Smith and Schlather models. Finally, [39] show that the Brown-Resnick process has connections with mixed moving maxima processes. More precisely, a Brown-Resnick process corresponding to a Gaussian process $\varepsilon(\cdot)$ with stationary increments has a mixed moving maxima representation provided that $\varepsilon(s) - \sigma^2(s)/2$ tends to $-\infty$ almost surely as $|s| \to \infty$.

Other characterizations and representations of max-stable processes are possible. The *multivariate maxima of moving maxima* processes [65, 76, 77] are well known max-stable processes but not yet exploited in a spatial context. The ongoing research [38, 67, 73, 74] is exploring the link of max-stable processes with sum-stable processes and these recent contributions seem promising to solve the prediction task (see Sect. 5.5).

5.4 Inference for Spatial Extreme Model

5.4.1 The Extremal Coefficient Function

Measuring extremal dependence between two random variables by means of their correlation is inappropriate because the correlation measures the dependence around the mean values. Moreover some extreme value distributions do not have moments of order two.

[63] suggests to measure the dependence by means of the extremal coefficient. For a bivariate vector (X, Y) with bivariate extreme distribution and common marginal distribution $G(\cdot)$ the extremal coefficient θ is defined by

$$\mathrm{P}(max(X, Y) \leq x) = G^{\theta}(x).$$

The values of θ range from one to two where the perfect dependence corresponds to $\theta = 1$, instead for independent maxima we have $\theta = 2$. To simplify the presentation, in the sequel we consider random variables with unit Fréchet distribution. In this case we have

$$\mathrm{P}(max(X, Y) \leq x) = \exp\left(-\frac{\theta}{x}\right) = \exp\left(-\frac{V(1,1)}{x}\right) = \exp\left(-\frac{2A(1/2)}{x}\right). \tag{5.11}$$

For a max-stable process $Z(\cdot)$, the extremal coefficient function [59], $\theta(\cdot)$, is given by

$$\mathrm{P}\left(\max(Z(s), Z(s+h)) \leq z\right) = \exp\{-\theta(h)/z\}.$$

Moreover [59] prove some properties of $\theta(\cdot)$:

1. $2 - \theta(\cdot)$ is a semi-definite positive function;
2. The function $\theta(\cdot)$ is not differentiable at 0, unless $\theta(h) = 1$, for all h;
3. If $Z(\cdot)$ is an isotropic stochastic process, $\theta(\cdot)$ has at most one discontinuity at 0 and is continuous elsewhere.

The extremal coefficient function for max-stable models in Sect. 5.3 can be derived from their bivariate distribution and we get:

Gaussian extreme value process: $\theta(h) = 2\Phi\left(\frac{\sqrt{h'\Sigma^{-1}h}}{2}\right)$;

Extremal Gaussian process: $\theta(h) = 1 + \sqrt{(1-\rho(h))/2}$;

Brown-Resnick process: $\theta(h) = 2\Phi\left(\frac{\sqrt{2\gamma(h)}}{2}\right)$.

Note that for the extremal Gaussian process $\lim_{|h|\to\infty} \theta(h) = 1 + 1/\sqrt{2} < 2$, therefore the (asymptotic) independence of the maxima does not occur at any distance. Moreover, for the Brown-Resnick process we have independence if $2\gamma(\cdot)$ is unbounded.

[42] specify compatibility relationships between extremal coefficients at various spatial supports, extending results for max-stable random vectors [59] and on rectangle spatial supports [10, 13].

Estimators of a single value $\theta(h)$ are proposed by [15, 59, 63] and [4]. Assume that we have T independent copies, $Z_h^{(t)} = (Z_{h,1}^{(t)}, Z_{h,2}^{(t)})'$, $t = 1, \ldots, T$ of the bivariate vector $Z_h = (Z(s), Z(s+h))'$, with unit Fréchet margins. The random variable $1/Z(s)$ has exponential distribution and $\min(1/(Z(s), 1/Z(s+h))$ has an exponential distribution with rate $\theta(h)$. This suggests as natural estimator [63]

$$\widehat{\theta}(h) = \frac{T}{\sum_{t=1}^{n} \min([Z_{h,1}^{(t)}]^{-1}, [Z_{h,2}^{(t)})]^{-1})}.$$

[59] derive a maximum likelihood estimator maximizing the censored log-likelihood

$$l(\theta(h)) = \#\left\{t : \max_{k=1,2}(Z_{h,k}^{(t)} \bar{Z}_{h,k}) > u\right\} \log\theta(h) - \theta(h) \sum_{t=1}^{T} \left[\max\left\{u, \max_{k=1,2}(Z_{h,k}^{(t)} \bar{Z}_{h,k})\right\}\right]^{-1} \tag{5.12}$$

where u is a high threshold level and $\bar{Z}_{h,k} = T^{-1} \sum_{t=1}^{T} 1/Z_{h,k}^{(t)}$.

In Geostatistics, it is common to characterize the spatial bivariate structure of a process $Z(\cdot)$ by means of the *madogram*

$$\nu(h) = \frac{1}{2}\mathbb{E}|Z(s+h) - Z(s)|.$$

[15] suggest to consider the madogram for the transformed max-stable process $F(Z(\cdot))$, where $F(\cdot)$ is the unit Fréchet distribution. In this case they show that

$$\nu_F(h) = \frac{1}{2}\mathbb{E}|F(Z(s+h)) - F(Z(s))| = \frac{1}{2}\frac{\theta(h)-1}{\theta(h)+1}$$

which leads to the natural estimator for $\theta(h)$

$$\widehat{\theta}(h) = \frac{1 + 2\widehat{\nu}_F(h)}{1 - 2\widehat{\nu}_F(h)},$$

where $\widehat{\nu}_F(h) = \sum_{t=1}^{T} |Z^{(t)}(s) - Z^{(t)}(s+h)|/2T$.

A generalization of the [15] estimator has been proposed by [4]. In a large simulation study, [4] compared four estimators for $\theta(\cdot)$, including the non-parametric estimator $\hat{\theta}(h) = 2\hat{A}_h(1/2)$ deduced from (5.11), where $\hat{A}_h(1/2)$ is the nonparametric

estimator [8] for the Pickands function for the pair $(Z(s), Z(s+h))$. Their simulation results point out that no one of the considered estimators outperforms the other, but the two madogram based estimators lead to slightly better results. Recently, a madogram based estimator for the dependence function has been proposed by [47].

The estimates $\widehat{\theta}(h)$ can be used to fit the parameters of a max-stable model. Assuming that the max-stable model depends on an unknown parameter ψ, characterizing the spatial dependence, [63] suggests to minimize the weighted sum of squares criterion

$$S(\psi) = \sum_{h_i \in \mathscr{H}} \left(\frac{\widehat{\theta}(h_i) - \theta(h_i;\psi)}{se(\widehat{\theta}(h_i))} \right)^2$$

which contrasts the theoretical values $\theta(h_i;\psi)$ with the estimates $\hat{\theta}(h_i)$. Here $\mathscr{H}$ is a set of spatial lags h_i chosen for evaluating the estimate $\widehat{\theta}(h_i)$. Here $\theta(\cdot;\psi)$ is the theoretical extremal coefficient function and $se(\widehat{\theta}(h_i))$ is the standard error related to the estimates $\widehat{\theta}(h_i)$, that can be evaluated using, for instance, a jackknife or bootstrap estimator.

A plug-in estimate for the parameter ψ based on the Pickands dependence function is suggested by [22] and applied in [7].

5.4.2 Composite Likelihood Methods

The heart of difficulties in doing likelihood inference for max-stable processes is that the full likelihood is difficult to evaluate, because we do not have a closed form expression for the distribution of the whole observation vector. The composite likelihood (CL) is an inference function derived by multiplying likelihoods of marginal or conditional events. The terminology has been introduced by [44], but his precedent in spatial statistics was the Besag's pseudo-likelihood [5]. Today, around this approximation, there is a renewed interest testified in a thoughtful overview of its applications [71].

Let $\{f(\cdot,\psi), \psi \in \Psi\}$ be a parametric family of joint densities for the observations $X_1, \ldots, X_m \in \mathbb{R}^m$ and consider a set of events $\{A_i : A_i \subseteq \mathfrak{F}, i \in I \subseteq \mathbb{N}\}$, where $\mathfrak{F}$ is a σ-algebra on $\mathbb{R}^m$. The logarithm of CL is defined as

$$cl(\psi) = \sum_{i \in I} \log f((X_1, \ldots, X_m)' \in A_i; \psi).$$

Now suppose that we have observed a maximum $Z^{(t)}(s_i)$ at site s_i and over a temporal block t, with $i = 1, \ldots, n$ and $t = 1, \ldots, T$ and the observations are temporally independent. If we are interested only in the parameters of the marginal

distributions, we can consider the estimates coming from the marginal composite likelihood constructed pretending that the observations are spatially independent,

$$cl_{\text{ind}}(\psi) = \sum_{t=1}^{T} \sum_{i=1}^{n} \log f(Z^{(t)}(s_i); \psi)$$

The standard errors have to be adjusted (see below) otherwise they will be underestimated and this procedure has been firstly suggested by [64].

If the parameters of interest are also related to dependence we can consider the pairwise log-likelihood [50, 61]:

$$cl_{\text{pair}}(\psi) = \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j>i}^{n} \log f(Z^{(t)}(s_i), Z^{(t)}(s_j); \psi)$$

Under suitable conditions [50] the maximum composite likelihood estimator for ψ can be proved consistent and asymptotically Gaussian with asymptotic variance the inverse of the Godambe information matrix

$$I_a(\psi) = H_a(\psi)[J_a(\psi)]^{-1} H_a(\psi),$$

where $H_a(\psi) = \mathbb{E}(-\nabla^2 cl_a(\psi))$ and $J_a(\psi) = \mathbb{V}(\nabla cl_a(\psi))$, and $a = \text{ind}, \text{pair}$.

Pairwise likelihood estimation has been implemented in the R packages `SpatialExtremes` [55] and `ExtremalProc` [49] available on the CRAN repositories.

5.5 Simulation of Spatial Max-Stable Processes

Spatial simulations allow the recovery of information at any site. When the simulations fit the observed data, they are said to be conditional, otherwise they are termed unconditional. In particular, the conditional simulation appears as a remarkable tool since different conditional simulations can be combined to derive a predictor of any quantity of interest (quantiles, probabilities of exceedance, return levels, ...).

Simulation methods are well established for Gaussian spatial models [41]. Instead there exist few methods for unconditional simulation of max-stable processes [42, 54, 57] and, to the best of our knowledge, one attempt connected to the conditional case [75].

5.5.1 Unconditional Simulations

According to formula (5.8), (5.9) and (5.10), a max-stable process $Z(\cdot)$ is defined through a maximum over an infinite number of copies of a stochastic process.

In practice only a finite number of realizations of the random process can be generated. Nevertheless, there are theoretical results and *ad-hoc* procedures that allow us to obtain exact or approximate simulations of some max-stable models. [57] gives conditions which ensure exact simulations for the max-stable processes (5.8) and (5.9). We will illustrate the simulation algorithm in this last case.

Suppose that the stochastic process $W(\cdot)$ is uniformly bounded by a finite and positive constant C. Setting $T_0 = 0$, for $k = 1, 2, \ldots$, the following procedure simulates a realization from a max-stable process (5.9):

1. generate $E_k \sim \mathscr{E}$ and put $T_k = T_{k-1} + E_k$, $X_k = T_k^{-1}$;
2. generate $W_k(\cdot) \sim W(\cdot)$;
3. if $C\,X_k > \max\limits_{1 \le i \le k} X_i W_i(s)$, go to (1); else return $Z(s) = \max\limits_{1 \le i \le k} X_i W_i(s)$.

Here $\mathscr{E}$ stands for the exponential distribution with unit mean. By construction $(T_i)_{i \ge 1}$ is a Poisson process on $(0, \infty)$ with intensity measure dt and $(X_i)_{i \ge 1}$ is a Poisson process on $(0, \infty)$ with the required intensity measure $x^{-2}dx$. The algorithm ends in a finite number of iterations because X_k decreases to 0, as $k \to \infty$. When the condition about uniformly boundness is not fulfilled, an approximate algorithm can be introduced. For instance, good results are obtained when we choose C such that $\mathrm{P}(W(s) > C)$ is small enough.

An analogous algorithm can be devised for the max-stable process (5.8) provided that the random shape $Y(\cdot)$ is uniformly bounded and has finite support ([57], Theorem 4). Both of these algorithms are implemented in the R packages *Random Fields* [58] and *SpatialExtremes* [55].

More recently [42] introduced the exact simulation of the Poisson storm process, exploiting specific properties of the random storms.

Simulations of Brown-Resnick models are more difficult to obtain. Ideas used for the storm processes seem to lead to simulation results with a poor quality. In their experiments, [48] and [40] report simulations of the Brown-Resnick process using the process definition and a fixed number k of iterations that appear non stationary, if the semi-variogram $\gamma(\cdot)$ increases fast.

5.5.2 Conditional Simulations

We have already remarked that it is difficult to obtain in closed form the multivariate distribution of the max-stable process making the exact conditional simulation infeasible.

[75] provide an approximate answer to this problem. Suppose we have observed a max-stable process $Z(\cdot)$ over n sites $s_1, \ldots s_n$. Starting from the extremal integral representation [67] of a max-stable process $Z(\cdot)$, the distribution of the multivariate max-stable vector $(Z(s_1), \ldots, Z(s_n))'$ can be approximated arbitrarily well (for large p) by the distribution of the multivariate vector $(\tilde{Z}(s_1), \ldots, \tilde{Z}(s_n))'$, with $\tilde{Z}(s_i) = \max_{k=1,\ldots,p} \phi_k(s_i) Y_k$, for $i = 1, \ldots, n$. Here the $\phi_k(\cdot)$'s are suitable

non-negative deterministic functions and Y_k's are independent standard α-Fréchet random variables, with distribution $\mathrm{P}(Y_k \leq y) = \exp(-\sigma_k^{\alpha} y^{-\alpha})$, for $\alpha, \sigma_k, y > 0$.

The Wang and Stoev's algorithm generate samples from the regular conditional probability of $(Y_1, \ldots, Y_p)'$ given $(\tilde{Z}(s_1), \ldots, \tilde{Z}(s_n))' = (Z(s_1), \ldots, Z(s_n))'$ and predict $Z(s^*)$ at any arbitrary location s^* from the samples, $(Y_1^{(sim)}, \ldots, Y_p^{(sim)})'$, using the discrete approximation of max-stable process:

$$\tilde{Z}^{(sim)}(s^*) = \max_{k=1,\ldots,p} \phi_k(s^*) Y_k^{(sim)}.$$

The simulation algorithm is implemented in the R package `maxLinear` [72].

5.6 Spatial Hierarchical Models

Spatial hierarchical models [2] stem from the remark that given three random variables Z, Y and X, we can always decompose the joint distribution of the triplet (Z, Y, X) by successive conditioning

$$[Z, Y, X] = [Z|Y, X][Y|X][X]$$

where $[A]$ denotes the distribution of random variable A. For spatial data, model building is broken into three stages:

1. data model: $[data|process, parameters]$
2. process model: $[process|parameters]$
3. parameter model: $[parameters]$

At stage 1 we specify how the *data* are generated by a distribution driven by a latent spatial *process*, i.e. the likelihood, at stage 2 we give the distribution of the *process* which depends on some *parameters*. The final stage specifies the distribution of the *parameters*, casting the model in a Bayesian framework.

In the literature [9, 16, 29] univariate extreme value models are used as building blocks in the first stage.

Suppose that we want to model the maximum $Z(s)$ at site s and we assume that conditional on three latent spatial processes $\mu(\cdot)$,$\sigma(\cdot)$ and $\xi(\cdot)$, $Z(s)$ is a random variable with $GEV(\mu(s), \sigma(s), \xi(s))$ density

$$f(Z(s); \mu(s), \sigma(s), \xi(s)) = \frac{1}{\sigma(s)} \left[1 + \frac{\xi(s)}{\sigma(s)} (Z(s) - \mu(s))\right]_+^{-1/\xi(s)-1} \exp\left\{-\left[1 + \frac{\xi(s)}{\sigma(s)} (Z(s) - \mu(s))\right]_+^{-1/\xi(s)}\right\} \quad (5.13)$$

Moreover we assume that at observed sites $(s_1, \ldots s_n)$ the random variables $Z(s_i)$, $i = 1, \ldots, n$, are independent on $(\mu(s_i), \sigma(s_i), \xi(s_i))$.

[9] suggest to model threshold exceedances at each location using the point process characterization. More precisely, let $\mathscr{E}(s) = \{Z_i(s) : Z_i(s) - u > 0\}$ be the set of exceedances of a threshold u at location s and $N(s) = |\mathscr{E}(s)|$ the number of these exceedances, the likelihood at each site s is given by

$$f(\mathscr{E}(s); \mu(s), \sigma(s), \xi(s)) = \exp\left\{-m\left[1 + \frac{\xi(s)}{\sigma(s)}(u - \mu(s))\right]_+^{-1/\xi(s)}\right\} \prod_{Z_k(s)\in\mathscr{E}(s)} \left\{\frac{1}{\sigma(s)}\left[1 + \frac{\xi(s)}{\sigma(s)}(Z_k(s) - u)\right]_+^{-1/\xi(s)-1}\right\} \tag{5.14}$$

where m is the number of blocks.

[16] models the exceedances using the GPD distribution and specifying a Binomial distribution for the number of exceedances

$$f(\mathscr{E}(s); \tilde{\sigma}(s), \xi(s), \pi(s)) = \binom{N_b}{N(s)} \pi(s)^{N(s)}(1 - \pi(s))^{N_b - N(s)} \prod_{Z_k(s)\in\mathscr{E}(s)} \left\{\frac{1}{\tilde{\sigma}(s)}\left[1 + \frac{\xi(s)}{\tilde{\sigma}(s)}(Z_k(s) - u)\right]_+^{-1-1/\xi(s)}\right\} \tag{5.15}$$

where N_b is the number of observations in a block and $\pi(s)$ is the probability that an observation exceeds u.

In modelling exceedances we assume that at observed sites $(s_1, \ldots s_n)$ the random vectors, $\mathscr{E}(s_i)$, $i = 1, \ldots, n$, are independent on $(\mu(s_i), \sigma(s_i), \xi(s_i))$ or on $(\tilde{\sigma}(s_i), \xi(s_i), \pi(s_i))$ and that the exceedances observed at each site are temporal independent. Actually exceedances often occur in clusters so we need to decluster the data. The first step in a declustering procedure [62] is to identify the clusters. Two clusters of observations over time t at each site s are separated when, after observing $Z_t(s) > u$, we have m consecutive observations over time below u. Thus, we decluster the data by keeping only the highest measurement of that cluster. Selecting the cluster maximum for each cluster makes the data approximately independent. Obviously the critical issue for declustering is the choice of m. Tools for selecting m have been proposed by [43] and [24].

In the stage two, the specification of the latent spatial processes $a(\cdot)$, $a = \mu, \sigma, \tilde{\sigma}, \xi, \pi$, follows a geostatistical approach namely

$$h_a(a(s)) = g_a(s; \beta_a) + X_a(s; \psi_a),$$

where $h_a(\cdot)$ is a known link function, $g_a(\cdot, \beta_a)$ is a regression function with unknown parameter β_a representing the large scale component and $X_a(\cdot, \psi_a)$ is a zero-mean Gaussian stochastic process with unknown parameter ψ_a representing the small scale component.

Moreover, for simplicity, spatial processes $X_a(\cdot;\psi_a)$ are supposed mutually independent, even if [66] suggested that scale and shape parameters should have negative dependence.

Zero-mean spatial Gaussian processes are specified through the covariance function $c_a(s,s';\psi_a)$. Its explicit formulation requires the inversion of a potentially large matrix for doing Bayesian hierarchical inference. A possible solution is approximating the spatial process $X_a(\cdot;\psi)$ by a discrete convolution [29].

Prior distributions for the parameters β_a are usually chosen uninformative, instead for the parameters ψ_a it is difficult to elicit prior information (see [16] for an example). For variance parameters in $c_a(s,s';\psi_a)$ Inverse Gamma priors are a possible choice.

Adopting a hierarchical formulation has limitations and advantages. Firstly models are max-stable only conditionally and the resulting predictive distributions will be a mixture of max-stable distributions. Moreover conditional independence entails that the spatial variation is taken into account only through the parameters of the marginal max-stable distributions. Thus when we do simulations from these models, the resulting patterns lack of local spatial structure, with an effect similar to a nugget effect. For weather extremes this fact could be an issue.

However hierarchical modelling shows its potential benefit when we want to improve the knowledge about extreme events over a long period of time. Examples are climate parameters like return levels [16] or trends in extreme rainfalls [29].

A way of introducing extra spatial dependence in the likelihood is through copulas. A copula C is a multivariate distribution on the $n-$dimensional hypercube $[0,1]^n$. Suppose that $G_i(\cdot)$ is the marginal distribution of a random variable Z_i, $i = 1,\dots,n$, then there exists a function $C(\cdot)$ such that the distribution $G(z_1,\dots,z_n) = C(G_1(z_1),\dots,G_n(z_n))$ is a multivariate distribution with marginal distributions $G_i(\cdot)$. Note that when we deal with continuous random variables, as in the present case, the copula is unique [60]. Advantages and disadvantages about copula modeling of extreme values are discussed in [46].

[56], assuming the conditional model (5.13) for the maxima, use a Gaussian copula. Firstly, they transform the data into

$$Y(s) = \exp\left\{-\left[1+\frac{\xi(s)}{\sigma(s)}\left(Z(s)-\mu(s)\right)\right]_+^{-1/\xi(s)}\right\}$$

then they build the likelihood as

$$f(Z(s_1),\dots,Z(s_n);\Theta) = \phi_{0,\Sigma_G}(\Phi^{-1}(Y(s_1)),\dots,\Phi^{-1}(Y(s_n)))$$

where $\Theta = \{\psi_G, (\mu(s_i),\sigma(s_i),\xi(s_i), i=1,\dots,n\}$, $\phi_{m,\Sigma}$ is the n-variate Gaussian density, with mean vector m and covariance matrix Σ, and Φ is the distribution of a standard Gaussian random variable. The covariance matrix $\Sigma_G = [c_G(s_i,s_j;\psi_G)]_{i,j=1}^n$ comes from the covariance function $c_G(s,s';\psi_G)$ of a fixed spatial Gaussian process. Using a Gaussian copula has computational

advantages [56] and allows us a great flexibility in choosing covariance models, however Gaussian copula entails asymptotic independence for a pair of maxima $(Z(s), Z(s'))$ observed at two sites $s \neq s'$. A solution to overcome this problem is using a multivariate t-Student copula with covariance matrix Σ_T and d_T degrees of freedom [32].

A different spatial hierarchical model has been introduced by [7] and its distinctive feature is that the marginal distributions are consistent with the max-stable theory. Let Y be a random variable with GPD distribution $1 - (1 + y)^{-1}$, $y \geq 0$, and $X(\cdot)$ a positive stochastic process, independent from Y such that: $\mathbb{E}[X(s)] = 1$ and $\sup_{s \in S} X(s) = u$. The GPD process $Z(\cdot) = (Y + 1)X(\cdot)$ has marginal distribution with GPD tail, namely

$$\mathrm{P}\{Z(s) > z + 1\} = (1 + z)^{-1}$$

for $1 + z > u$.

5.7 Discussion and Perspectives

Recent results in the literature offer tools for modelling spatial extreme data and, as it has been illustrated in the previous sections, both spatial max-stable processes and hierarchical models appear as interesting tracks to work with spatial extremes. Nevertheless, there are some topics for discussion that we propose to the reader.

Max-stable spatial models for block maxima are usually defined in a stationary context [61] and nonstationarity is worked out by means of the marginal distributions of the data [50].

Their parametric inference relies on the composite likelihood approach, but nowadays we have considered only the bivariate densities. For spatial Gaussian processes these choices can be supported because empirical covariances are sufficient statistics for the dependence parameter, but for max-stable processes densities with higher dimensionality could be considered [31]. Moreover spatial models for threshold exceedances and related statistical inference techniques appear welcome. Research on a composite likelihood inference for spatial threshold exceedances is currently under investigation [1].

Hierarchical modelling is a natural way for incorporating known physical constraints, through informative priors. Recently [26] have proposed a hierarchical model that is consistent with the max-stable theory. Apparently their model lacks a spatial application.

As underlined in Sect. 5.5, there is a demand for valuable methods for conditional simulations of spatial extreme processes. Moreover, when available information is provided by two datasets, global climatic models outputs and rain-gauge data for example, sometimes we need to take into account the multi-support spatial structure. Recent results on downscaling extreme values are available [28, 45] and change of

scales for extreme values appear as a fundamental problem, particularly in a climate change framework.

For many environmental processes, extreme events have a spatial and temporal component. There are some attempts to model both components in a Bayesian hierarchical framework [32, 70] but, to our best knowledge, no real application of recent advances [17] for space-time processes with regularly varying innovations has been found.

Acknowledgments Gaetan's research was partially supported by MIUR grant 2008MRFM2H. Bacro's research was partially supported by the ANR-McSim and GICC-Miraccle projects.

References

1. Bacro, J.N., Gaetan, C.: Composite likelihood estimation for spatial max-stable processes using threshold exceedances. Submitted for publication (2011)
2. Banerjee, S., Carlin, B.P., Gelfand, A.E.: Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC Press, Boca Raton (2004)
3. Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of Extremes: Theory and Applications. John Wiley & Sons, New York (2004)
4. Bel, L., Bacro, J.N., Lantuéjoul, C.: Assessing extremal dependence of environmental spatial fields. Environmetrics. **19**, 163–182 (2008)
5. Besag, J.: Spatial interaction and the statistical analysis of lattice systems (with discussion). J. Roy. Stat. Soc. B. **36**, 192–236 (1974)
6. Brown, B., Resnick, S.: Extremes values of independent stochastic processes. J. Appl. Probab. **14**, 732–739 (1977)
7. Buishand, T.A., de Haan, L., Zhou, C.: On spatial extremes: with application to a rainfall problem. Ann. Appl. Stat. **2**, 624–642 (2008)
8. Capéraà, P., Fougères, A.L., Genest, C.: A non parametric estimation procedure for bivariate extreme value copulas. Biometrika. **84**, 567–577 (1997)
9. Casson, E., Coles, S.G.: Spatial regression models for extremes. Extremes. **1**, 449–468 (1999)
10. Coles, S.G.: Regional modelling of extreme storms via max-stable processes. J. Roy. Stat. Soc. B. **55**, 797–816 (1993)
11. Coles, S.G., Tawn, J.A.: Modelling extreme multivariate events. J. Roy. Stat. Soc. B. **53**, 377–392 (1991)
12. Coles, S.G., Tawn, J.A.: Statistical methods models for multivariate extremes: an application to structural design. Appl. Stat. **43**, 1–48 (1994)
13. Coles, S.G., Tawn, J.A.: Modelling extremes of the areal rainfall process. J. Roy. Stat. Soc. B. **58**, 329–347 (1996)
14. Coles, S.G., Walshaw, D.: Directional modelling of extreme wind speed. J. Appl. Stat. **43**, 139–157 (1994)
15. Cooley, D, Naveau P., Poncet, P.: Variograms for spatial max-stable random fields. In: Bickel P, Diggle P, Fienberg S, Gather U, Olkin I, Zeger S, Bertail P, Soulier P, Doukhan P (eds) Dependence in Probability and Statistics. Lect. Notes Stat. **187**, 373–390, Springer, New York (2006)
16. Cooley, D., Nychka, D., Naveau, P.: Bayesian spatial modeling of extreme precipitation return levels. J. Am. Stat. Assoc. **102**, 824–840 (2007)
17. Davis, R.A., Mikosch, T.: Extreme value theory for space-time processes with heavy-tailed distributions. Stoch. Proc. Appl. **118**, 560–584 (2008)

18. Davison, A.C., Smith, R.L.: Models for exceedances over high thresholds. J. Roy. Stat. Soc. B. **3**, 393–442 (1990)
19. Davison, A., Padoan, S., Ribatet, M.: Statistical modelling of spatial extremes. Accepted for publication, Statistical Science (2011)
20. de Haan. L.: A spectral representation for max-stable processes. Ann. Probab. **12**, 1194–1204 (1984)
21. de Haan, L., Ferreira, A.: Extreme Value Theory: an Introduction. Springer, New-York (2006)
22. de Haan, L., Pereira, T.T.: Spatial extremes: models for the stationary case. Ann. Stat. **34** 146–168 (2006)
23. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events: for Insurance and Finance, corrected edn. Springer, New York (1997)
24. Ferro, C.A.T., Segers, J.: Inference for clusters of extreme values. J. Roy. Stat. Soc. B. **65**, 545–556 (2003)
25. Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample. P. Camb. Philos. Soc. **24**, 180–190 (1928)
26. Fougères, A., Nolan, J.P., Rootzén, H.: Models for dependent extremes using stable mixtures. Scand. J. Stat.: **36**, 42–59 (2009)
27. Fougères, A.L.: Multivariate extremes. In: Finkenstädt B, Rootzén H (eds) Extreme Values in Finance, Telecommunications, and the Environment. Chapman & Hall/CRC, Boca Raton, 373–388 (2004)
28. Friederichs, P.: Statistical downscaling of extreme precipitation events using extreme value theory. Extremes. **13**, 109–132 (2010)
29. Gaetan, C., Grigoletto, M.: A hierarchical model for the analysis of spatial rainfall extremes. J. Agr. Biol. Envir. St. **12** 434–449 (2007)
30. Galambos, J.: The Asymptotic Theory of Extreme Order Statistics, 2nd edn. Wiley, New York (1987)
31. Genton, M.G., Ma, Y., Sang, H.: On the likelihood function of Gaussian max-stable processes. Biometrika **98**, 481–488 (2011)
32. Ghosh, S., Mallick, B.K.: A hierarchical Bayesian spatio-temporal model for extreme precipitation events. Environmetrics **22**, 192–204 (2011)
33. Gnedenko, B.V.: Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. **44**, 423–453 (1943)
34. Gumbel, E.: Distribution de valeurs extremes en plusieurs dimensions. Publications de l'Institut de Statistique de l'Université de Paris. **9**, 171–173 (1960)
35. Hüsler, J., Reiss, R.: Maxima of normal random vectors: between independence to complete dependence. Stat. Probabil. Lett. **7**, 283–286 (1989)
36. Jenkinson, A.F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Q. J. Roy. Meteor. Soc. **81**, 158–171 (1955)
37. Joe, H.: Families of min-stable multivariate exponential and multivariate extreme value distributions. Stat. Probabil. Lett. **9**, 75–81 (1990)
38. Kabluchko, Z.: Spectral representations of sum- and max-stable processes. Extremes. **12**, 401–424 (2009)
39. Kabluchko, Z., Schlather, M., de Haan, L.: Stationary max-stable fields associated to negative definite functions. Ann. Probab. **37**, 2042–2065 (2009)
40. Oesting, M., Kabluchko, Z., Schlather, M. : Simulation of Brown-Resnick processes. Extremes. To appear, doi: 10.1007/s10687-011-0128-8 (2011).
41. Lantuéjoul, C.: Geostatistical Simulation. Springer, Berlin, Heidelberg (2002)
42. Lantuéjoul, C., Bacro, J.N., Bel, L. : Storm processes and stochastic geometry. Extremes. To appear, doi: 10.1007/s10687-010-0121-7 (2011)
43. Ledford, A.W., Tawn, J.A.: Diagnostics for dependence within time series extremes. J. Roy. Stat. Soc. B.: **65**, 521–543 (2003)
44. Lindsay, B.: Composite likelihood methods. Contemp. Math. **80**, 221–239 (1988)

45. Mannshardt-Shamseldin, E., Smith, R., SR Sain, L.M., Cooley, D.: Downscaling extremes: a comparison of extreme value distributions in point-source and gridded presipitation data. Ann. Appl. Stat. **4**, 484–502 (2010)
46. Mikosch, T.: Copulas: tales and facts (with discussion). Extremes 9:3–62 (2006)
47. Naveau, P., Guillou, A., Cooley, D., Diebolt, J.: Modelling pairwise dependence of maxima in space. Biometrika, **96**, 1–17 (2009)
48. Oesting, M., Kabluchko, Z., Schlather, M.: Simulation of Brown-Resnick processes. Extremes. To appear, doi: 10.1007/s10687-011-0128-8 (2011)
49. Padoan, S.: ExtremalProc: modelling and analysis of extremal processes. R package version 0.1-2 (2010)
50. Padoan, S.A., Ribatet, M., Sisson, S.: Likelihood-based inference for max-stable processes. J. Am. Stat. Assoc. **105**, 263–277 (2010)
51. Pickands, J.: Statistical inference using extreme order statistics. Ann. Stat. **3**, 119–131 (1975)
52. Pickands, J.: Multivariate extreme value distributions. In: Proceedings of the 43rd Session ISI, Buenos Aires, 859–878 (1981)
53. Resnick, S.: Extreme values, Regular variation and Point Processes. Springer Verlag, New York (1987)
54. Ribatet, M.: Geostatistics for extremes. Manuscript, Université Montpellier II, Montpellier (2010)
55. Ribatet, M.: SpatialExtremes: Modelling Spatial Extremes. R package version 1.6-0 (2010)
56. Sang, H., Gelfand, A.: Continuous spatial process models for spatial extreme values. J. Agr. Biol. Envir. St. **15**, 49–65 (2010)
57. Schlather, M.: Models for stationary max-stable random fields. Extremes. **5**, 33–44 (2002)
58. Schlather, M.: RandomFields: Simulation and Analysis of Random Fields. R package version 1.3.41 (2009)
59. Schlather, M., Tawn, J.A.: A dependence measure for multivariate and spatial extreme values: properties and inference. Biometrika. **90**, 139–156 (2003)
60. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris. **8**, 229–231 (1959)
61. Smith, E.L., Stephenson, A.G.: An extended Gaussian max-stable process model for spatial extremes. J. Stat. Plan. Infer. **139**, 1266–1275 (2009)
62. Smith, R.L.: Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. Stat. Sci. **4**, 367–393 (1989)
63. Smith, R.L.: Max-stable processes and spatial extremes. Preprint, University of Surrey, Surrey (1990)
64. Smith, R.L.: Regional estimation from spatially dependent data. Preprint University of North Carolina, Chapel Hill (1990)
65. Smith, R.L., Weissman, L.: Characterization and estimation of the multivariate extremal index. Preprint University of North Carolina, Chapel Hill (1996)
66. Stephenson, A.G., Tawn, J.: Bayesian inference for extremes: accounting for the three extremal types. Extremes. **7**, 291–307 (2004)
67. Stoev, S.A., Taqqu, M.S.: Extremal stochastic integrals: a parrallel between max-stable and alpha-stable processes. Extremes **8**, 237–266 (2006)
68. Tawn, J.: Bivariate extreme value theory: models and estimation. Biometrika. **75**, 397–415 (1988)
69. Tawn, J.: Modelling multivariate extreme value distributions. Biometrika, **77**, 245–253 (1990)
70. Turkman, K., Amaral Turkman, M., Pereira, J.: Asymptotic models and inference for extremes of spatio-temporal data. Extremes. **13**, 375–397 (2010)
71. Varin, C., Reid, N., Firth, D.: An overview of composite likelihood methods. Stat. Sinica. **21**, 5–42 (2011)
72. Wang, Y., Stoev, S.A.: Conditional sampling for spectrally discrete max-stable random fields. Advances in Applied Probability **43**, 461–483 (2011)
73. Wang, Y., Stoev, S.A.: On the association of sum- and max-stable processes. Stat. Probabil. Lett. **80**, 480–488 (2010)

74. Wang, Y., Stoev, S.A.: On the structure and representations of max-stable processes. Adv. Appl. Probab. **42**, 855–877 (2010)
75. Wang, Y., Stoev, S. A. : Conditional sampling for spectrally discrete max-stable random fields. Advances in Applied Probability **43**, 461–483 (2011).
76. Zhang, Z., Smith., R.L.: The behavior of multivariate maxima of moving maxima processes. J. Appl. Probab. **41**, 1113–1123 (2004)
77. Zhang, Z., Smith, R.L.: On the estimation and application of max-stable processes. J. Stat. Plan. Infer. **12**, 1135–1153 (2010)

Chapter 6
Relations Between Designs for Prediction and Estimation in Random Fields: An Illustrative Case

Werner G. Müller, Luc Pronzato, and Helmut Waldl

Abstract Two approaches are considered to design experiments for a correlated random field when the objective is to obtain precise predictions over the whole experimental domain. Both take the uncertainty of the estimated parameters of the correlation structure of the random field into account. The first one corresponds to a compound D-optimality criterion for both the trend and covariance parameters. The second one relies on an approximation of the mean squared prediction error already proposed in the literature. It is conjectured, and shown on a paradigmatic example, that for some particular settings both approaches yield similar optimal designs, thereby revealing a sort of accordance between the two criteria for random fields. However, our example also shows that a strict equivalence theorem as in the uncorrelated case is not achievable. As a side issue we cast doubts on the ubiquity of equidistant space-filling designs.

6.1 Introduction

For the development of design theory for experiments with independent observations, the so called equivalence theorem [10], henceforth KWET, and its extensions, have played a major role. It allows to quickly check whether given designs are optimal, and led to the development of efficient algorithms for constructing good designs. One of the key aspects of the KWET is the establishment of the

W.G. Müller (✉) · H. Waldl
Department of Applied Statistics, Johannes-Kepler-University Linz, Altenberger Strase 69, A-4040 Linz, Austria
e-mail: werner.mueller@jku.at; helmut.waldl@jku.at

L. Pronzato
Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis, Les Algorithmes, 2000 route des lucioles, BP 121, 06903 Sophia Antipolis cedex, France
e-mail: pronzato@i3s.unice.fr

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_6,

equivalence of optimal designs between two criteria of optimality, one related to parameter estimation, the other related to prediction (classically between D- and G-optimality).

Unfortunately, in the design and analysis for correlated random fields, given by

$$Y_t(x) = \eta(x, \beta) + \varepsilon_t(x), \quad t \in \mathscr{T} \subset \mathbb{R}, \tag{6.1}$$

most of the conditions of the KWET are not met. Here, $t \in \mathscr{T}$ indexes the different realizations of the field, β is an unknown vector of parameters in $\mathbb{R}^p$, x a known vector of regressors belonging to some set $\mathscr{X}$, and the random term $\varepsilon(x)$ has zero mean, (unknown) variance σ^2 and a parameterized spatial error correlation structure such that $\mathbb{E}[\varepsilon_t(x)\,\varepsilon_t(x')] = \sigma^2 c(x, x'; \gamma)$. It is often assumed that the deterministic term has a linear structure, i.e., $\eta(x, \beta) = f^\top(x)\beta$, and that the random field $\varepsilon_t(x)$ is Gaussian, allowing estimation of β, σ and γ by Maximum Likelihood.

Note that setup (6.1) is used in such diverse areas of spatial data analysis (cf. [4]) as mining, hydrogeology, natural resource monitoring and environmental science, and has become the standard modelling paradigm in computer simulation experiments, following the seminal paper [18]. In the analysis of deterministic computer models all realizations of the field are identical, $Y_t(x) \equiv Y_{t'}(x)$ for all t, t', and $\lim_{x' \to x} c(x', x; \gamma) = c(x, x; \gamma) = 1$ for all x.

In this note we will argue why we believe that an accordance of prediction and estimation based criteria can also be suspected in setup (6.1). At the same time, however, our example will serve to demonstrate that a strict KWET can not be achieved in this setting.

6.2 Motivation

We must carefully distinguish between two fundamentally distinct problems that are both usually designated as "prediction" problems. One may refer to [16] for a detailed overview of the differences between the two situations. In the following we will refer to ξ as the design, the collection of inputs x, and to $\mathscr{X}$ as the design space, the set of potential inputs. $\mathscr{S}$ stands for the set of unsampled inputs. Note that, due to correlations, we can only consider exact designs and thus n needs to be fixed beforehand.

6.2.1 Prediction of a Distinct Realization (Parameter Estimation): $\{Y_t(x), x \in \xi\} \to \{E\{Y_{t'}(x) \mid Y_t(z), z \in \xi\}, x \in \mathscr{X}\}$

Here, observations in the model (6.1) at a given $t \in \mathscr{T}$ and sites ξ are used to predict a future realization of the random field at a different $t' \neq t$. This problem has usually been assessed under the assumption that realizations at different times

are independent and identically distributed (see eg. [13]). Under this assumption, prediction amounts to estimation of the deterministic term $\eta(x, \beta)$ and requires estimation of the parameters β in (6.1), say by $\hat{\beta}^n$. The prediction of a future realization of the field $Y_{t'}(x)$ is then simply $\hat{\eta}^n(x) = \eta(x, \hat{\beta}^n)$. Note that here the influence of γ is somewhat hidden, but estimators of $\hat{\beta}^n$ will generally depend upon it, so that precise estimation of $\hat{\gamma}$ is required as well.

For this case, it has been suggested in [15] to maximize a compound criterion with weighting factor α,

$$\Phi[\xi \,|\, \alpha] = |M_\beta(\xi, \gamma)|^\alpha \cdot |M_\gamma(\xi, \gamma)|^{(1-\alpha)}, \tag{6.2}$$

which consists of determinants of information matrices corresponding to trend and covariance parameters, stemming from

$$\mathbb{E}\left\{ \begin{matrix} -\frac{\partial^2 lnL(\beta,\gamma)}{\partial\beta\partial\beta^\top} & -\frac{\partial^2 lnL(\beta,\gamma)}{\partial\beta\partial\gamma^\top} \\ -\frac{\partial^2 lnL(\beta,\gamma)}{\partial\gamma\partial\beta^\top} & -\frac{\partial^2 lnL(\beta,\gamma)}{\partial\gamma\partial\gamma^\top} \end{matrix} \right\} = \begin{pmatrix} M_\beta(\xi,\gamma) & 0 \\ 0 & M_\gamma(\xi,\gamma) \end{pmatrix}, \tag{6.3}$$

where, for the linear model

$$M_\beta(\xi, \gamma) = \frac{1}{n\sigma^2} \sum_{x_i \in \xi} \sum_{x_{i'} \in \xi} f(x)[C_n^{-1}(\gamma)]_{i,i'} f^\top(x'),$$

and

$$\{M_\gamma(\xi, \gamma)\}_{jj'} = \frac{1}{2} \operatorname{tr}\left\{ C_n^{-1}(\gamma) \frac{\partial C_n(\gamma)}{\partial \gamma_j} C_n^{-1}(\gamma) \frac{\partial C_n(\gamma)}{\partial \gamma_{j'}} \right\},$$

using notation $\{C_n(\gamma)\}_{ii'} = c(x_i, x_{i'}; \gamma)$, $i, i' = 1, \ldots, n$. In terms of experimental design, the framework is not much different from the standard one (for which the KWET holds), the difference being that for fixed t the errors $\varepsilon_t(x)$ in (6.1) are correlated.

6.2.2 *Prediction of the Same Realization (Inter-/extrapolation): $\{Y_t(x), x \in \xi\} \to \{E\{Y_t(x)|Y_t(z), z \in \xi\}, x \in \mathscr{S}\}$*

The situation here is very different from that encountered in the setting in Sect. 6.2.1. Notice that we require $\mathscr{S} \cap \xi = \emptyset$. Also note that, even in the idealized framework where β and the parameters σ^2 and γ of the covariance function of $\varepsilon_t(x)$ in (6.1) are known, predicting the value of $Y_t(x)$ at an unsampled site x requires the collection of (neighboring) observations on the particular realization of the field, whereas in a setting where the $\varepsilon_t(x)$ are (spatially) uncorrelated, prediction at $x \notin \xi$ is simply given by $\eta(x, \beta)$ when β is known and observations are useless.

Contrasting parameter estimation and interpolation/extrapolation, the second task is usually treated by minimizing a functional of the so-called kriging variance $\mathrm{Var}[\hat{Y}_t(x \mid \xi)] = \mathbb{E}[(\hat{Y}_t(x \mid \xi) - Y_t(x))^2]$ at site x, interpreted as the unconditional Mean-Squared Prediction Error (MSPE) for the best linear unbiased predictor at x. For instance one may minimize the maximum of it over a set $\mathscr{X}$,

$$\min_{\xi} \max_{x \in \mathscr{X}} \mathbb{E}[(\hat{Y}_t(x \mid \xi) - Y_t(x))^2]. \tag{6.4}$$

Here, $\hat{Y}_t(x \mid \xi)$ denotes the best-linear unbiased predictor of $Y_t(x)$ based on the design points in ξ and associated observations $Y_t(\xi) = [Y_t(x_1), \ldots, Y_t(x_n)]^\top$.

Assume that γ is known. In the linear setting (universal kriging, with $\eta(x, \beta) = f^\top(x)\beta$ in (6.1) a polynomial in x), it takes the form

$$\hat{Y}_t(x \mid \xi) = f^\top(x)\hat{\beta} + c_n^\top(x, \gamma) C_n^{-1}(\gamma)[Y_t(\xi) - F_n \hat{\beta}],$$

where $\{c_n(x, \gamma)\}_i = c(x, x_i; \gamma)$, $i = 1, \ldots, n$, and $\hat{\beta} = \hat{\beta}(\gamma)$ is the weighted Least-Squares estimator of β in the linear regression model, that is

$$\hat{\beta}(\gamma) = [F_n^\top C_n^{-1}(\gamma) F_n]^{-1} F_n^\top C_n^{-1}(\gamma) Y_t(\xi),$$

with $F_n = [f(x_1), \ldots, f(x_n)]^\top$. We can write

$$\hat{Y}_t(x \mid \xi) = v_n^\top(x, \gamma) Y_t(\xi)$$

with $v_n^\top(x, \gamma) \in \mathbb{R}^n$. The MSPE is given by

$$\begin{aligned} MSPE_\xi(x, \sigma^2, \gamma) = \sigma^2 \big\{ & 1 - c_n^\top(x, \gamma) C_n^{-1}(\gamma) c_n(x, \gamma) \\ & + g_n^\top(x, \gamma) [F_n^\top C_n^{-1}(\gamma) F_n]^{-1} g_n(x, \gamma) \big\} \end{aligned}$$

with $g_n(x, \gamma) = f(x) - F_n^\top C_n^{-1}(\gamma) c_n(x, \gamma)$. Note that the MSPE depends on (σ^2, γ), with σ^2 intervening only as a multiplicative factor. For a recent discussion of the related design problem in a different context see [8].

The situation gets more complicated when the covariance parameters are estimated (by Maximum Likelihood) from the same dataset. Indeed, the resulting additional uncertainty then needs to enter the design criterion. For instance, following the approach of [9] and using a first-order expansion of the MSPE for the estimated parameters $(\hat{\sigma}^{2^n}, \hat{\gamma}^n)$ around their true value, we obtain as an approximation an additional correcting term for the MSPE related to the observations collected. Assume for simplicity that σ^2 is known, we then get the corrected kriging variance by the approximation

$$\widehat{MSPE}_\xi(x,\hat{\gamma}^n) = MSPE_\xi(x,\sigma^2,\hat{\gamma}^n) + \operatorname{tr}\left\{ M_\gamma^{-1}(\xi,\hat{\gamma}^n) \frac{\partial v^\top(x,\theta)}{\partial \gamma}\bigg|_{\hat{\gamma}^n} C_n(\hat{\gamma}^n) \frac{\partial v(x,\theta)}{\partial \gamma^\top}\bigg|_{\hat{\gamma}^n} \right\}, \quad (6.5)$$

where $M_\gamma(\xi,\gamma)$ is the part of the (expected) information matrix related to the parameters γ, see (6.3). (When σ^2 is unknown and estimated on the same dataset, we need to consider the full information matrix $M_{\sigma^2,\gamma}(\xi,\sigma^2,\gamma)$ for parameters σ^2 and γ, and then replace in (6.5) $M_\gamma^{-1}(\xi,\hat{\gamma}^n)$ by the part of $M_{\sigma^2,\gamma}^{-1}(\xi,\hat{\sigma^2}^n,\hat{\gamma}^n)$ corresponding to γ; a similar modification can be used in (6.2).) Consequently, [26] (for a similar criterion see also [23]) regards

$$\min_\xi \max_{x\in\mathscr{X}} \widehat{MSPE}_\xi(x,\gamma) \quad (6.6)$$

for some nominal γ as the (local) design problem, which is termed EK-(empirical kriging-)optimality in the same paper. The objective here is to take into account the dual effect of the design (obtaining accurate predictions at unsampled sites and improving the accuracy of the estimation of the covariance parameters, those two objectives being conflicting, see [16]) through the formulation of a single criterion.

6.3 A Suspected Accordance Between Criteria

Let us briefly review one of the essential statements of the KWET [10], which was formulated for the classical linear regression setup with uncorrelated errors: it relates D-optimal designs for estimating the regression coefficients β, which maximize $|M_\beta|$, to so-called G-optimum designs, which minimize the maximum prediction variance, i.e.

$$\max_{x\in\mathscr{X}} \operatorname{Var}[\hat{Y}_t(x)].$$

Those criteria are equivalent when considering approximate designs, in the sense that the D-efficiency of a G-optimal approximate design is 100% and vice versa. Thus the respective efficiencies for exact designs can be expected to be high. It is thus natural that in [15] is formulated the conjecture that it may always be possible to find an α that allows to find designs optimizing $\Phi[\xi|\alpha]$ with high EK-efficiency (that is efficiency with respect to EK-optimality), thus establishing a relationship in the spirit of the KWET. Note that, using (6.5) and due to the unbiasedness of the kriging predictor, we can rewrite (6.6) as

$$\min_\xi \max_{x\in\mathscr{X}} \left\{ \operatorname{Var}[\hat{Y}_t(x)] + \operatorname{tr}\left\{ M_\gamma^{-1} \operatorname{Var}[\partial \hat{Y}_t(x)/\partial\gamma] \right\} \right\}, \quad (6.7)$$

which particularly highlights the affinity with G-optimality. The reasoning is much similar as for the two-step design suggested in [24], which however confines itself to the covariance parameters and seeks to find a balance not in the criterion itself, but in applying two different criteria successively. For more details and the derivation of (6.7) and variants see [1, 25, 26].

It has been observed, in particular in the references above, that both criteria are seeking to find a compromise between space-filling behavior (i.e. the trend parameter and the kriging variance component respectively) and short distances (i.e. the covariance parameter and the correcting term component respectively). We are thus led to believe that respectively good efficiencies can be produced for both setups. This would be very advantageous since EK-optimal designs are much more difficult to generate than parameter estimation designs (they require embedded optimizations over the candidate sets, see (6)). A quasi KWET-relationship would thus allow to replace the very demanding optimization (6.6) by the much less intensive (6.2) without much loss in efficiency.

In the next section we will give an extremely simple example that on the one hand supports the suspected accordance between the two criteria, but that on the other hand clearly shows that a strict equivalence between them cannot be achieved, contrary to the uncorrelated case (the KWET). This example was chosen from the wealth of our computations in various diverging setups and is intended to serve as a paradigmatic case.

6.4 Example

We will use the Ornstein-Uhlenbeck process on $\mathscr{T} = [0,1]$, which is a special case of (6.1) with $\eta(x, \beta) = \beta$, i.e. $f(x) \equiv 1$, and $c(x, x'; \gamma) = \exp|x - x'|/\gamma = \rho^{|x-x'|}$, setting $\sigma^2 \equiv 1$ to avoid identifiability problems (see, e.g. [21]). Also from now on we will be using the alternate parametrization ρ for ease of interpretation, with M_ρ replacing M_γ.

For this example, we have analytic results that correspond to the case $\alpha = 1$ in (2). The optimality of space-filling designs is proved in [11] and [22]. For the setup with an additional slope parameter it is shown in [5] that the points 0 and 1 must be included in the design, that for growing ρ the design tends to a space-filling design and that the efficiencies of space filling designs can be quite high also for small ρ (for small numbers of observations). The similar behavior for designs based on minimizing the (maximum or average of the) kriging variance is a widely acknowledged fact and has led to the predominance of equidistant designs for computer simulation experiments (cf. [3]). Prediction and estimation based criteria are recently compared in [2], who implicitely establish a KWET-type relationship in this restricted case.

However, contrasting results are known for the case $\alpha = 0$ in (2). Here it is shown [14] that the optimal designs collapse into one point; see also [22].

In the following we will only report results for $n = 3$ and $\rho = \frac{1}{100}$, although similar, albeit perhaps more trivial, results were achieved for other choices of ρ.

For growing n there is expectedly a tendency towards approaching the case $\alpha = 1$, which, however, is counterbalanced by growing ρ. This relation requires more detailed investigations in the future, but the below given necessarily limited cases seem to encapture the general behaviour well.

6.4.1 Two Design Points Fixed

We will start our investigations by fixing $x_1 = 0$ and $x_2 = 1$ and we will be looking for the optimal position x_3^* for the third design point. This is inspired by the findings of [5] in the linear case and will allow a more comprehensive exposition. The correlation matrix is now given simply by

$$C_3(\rho) = \begin{pmatrix} 1 & \rho & \rho^{|x_3|} \\ \rho & 1 & \rho^{|1-x_3|} \\ \rho^{|x_3|} & \rho^{|1-x_3|} & 1 \end{pmatrix}$$

and the resulting kriging variance as a function of the remaining point x_3 (and the points for prediction x) is displayed in the left panel of Fig. 6.1 with the axis for x_3 in front. It is evident that $x_3 = 0.5$ as expected minimizes the maximum over x of the kriging variance. That this is not the case for the corrected EK-criterion (6.6) can be easily seen from the right panel of Fig. 6.1, indicating that the minimum is reached for a point close to the endpoints of the region. In fact, the minimizing argument is $x_3^* = 0.934$ (or $x_3^* = 0.066$ respectively), which gives a lower corrected kriging variance (though also a much higher kriging variance) than the center point. This discrepancy of the two criteria is well documented in Fig. 6.2.

It now remains to be seen, whether the design $(0, 0.934, 1)$ (or at least one with an efficiency close to it) can be achieved employing criterion (6.2) with a particular choice of α. Its components are of the expected form, the $|M_\beta|$ being concave with a maximum value of $29/11$ at $x_3 = 0.5$ and the $|M_\rho|$ being convex with a limit value of 236.765 for $x_3 \to 0$ or $x_3 \to 1$ respectively. We can in fact tune the compound criterion $\Phi[x_3|\alpha]$ in such a way that it gives an optimum at $x_3 = 0.934$ (or

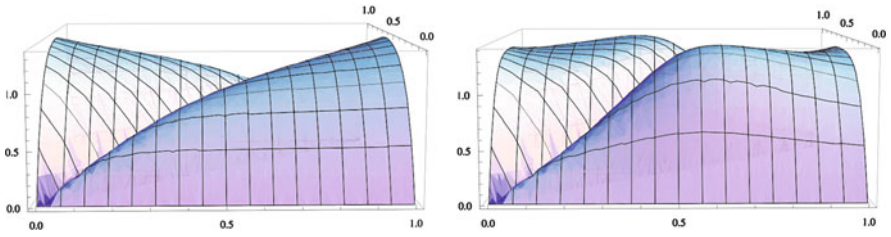

Fig. 6.1 Kriging variance (*left panel*) and corrected kriging variance (*right panel*) as a function of x_3 (*front axis*) and x

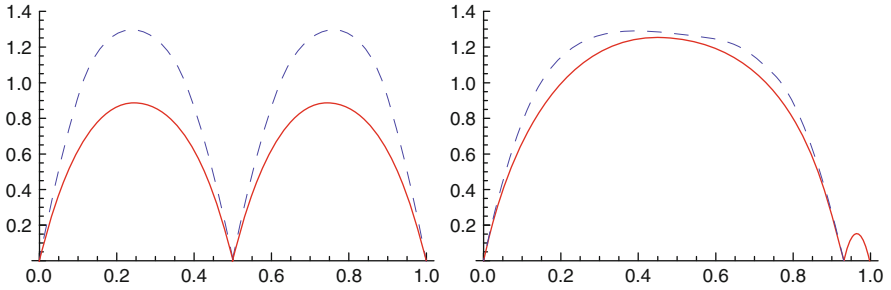

Fig. 6.2 Kriging variance (*solid line*) and corrected kriging variance (*dashed line*) as a function of x for $x_3 = 0.5$ (*left panel*) and $x_3 = 0.934$ (*right panel*)

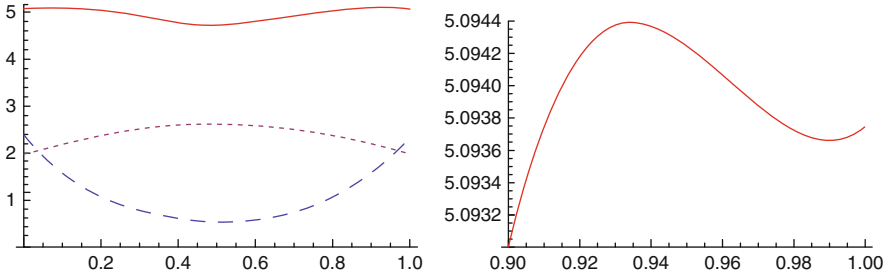

Fig. 6.3 $3|M_\beta|$ (dotted), $|M_\rho|/100$ (*dashed*) and $\Phi[x_3|\alpha = 0.8025]$ (*solid*) as functions of x_3 (*left panel*) and enlarged portion of the graph (*right panel*)

0.066 respectively) by choosing $\alpha = 0.8025$, which can be seen from the enlarged picture in the right panel of Fig. 6.3. Thus for this example we have achieved exact "equivalence" between the prediction and the estimation based criterion.

6.4.2 One Design Point Fixed

Let us continue the example by lifting the restriction to the endpoints and allowing two of the design points to vary freely. It is natural for reasons of symmetry to then fix one of the three points in the center, i.e. $x_2 = 0.5$ and the other two equally distant from the boundaries, i.e. $x_3 = 1 - x_1$. One can now plot the correcting second term in (6.7) as a function of x_3 and x and see that the situation differs much from the above, see Fig. 6.4.

It is indeed so that although again the optimal designs corresponding to the kriging variance and the EK-criterion differ (for the former we find $x_1^* = 0.120$ and $x_3^* = 0.880$, while for the latter $x_1^* = 0.099$ and $x_3^* = 0.901$), their structure and the form of the respective functions, see Fig. 6.5, are rather similar.

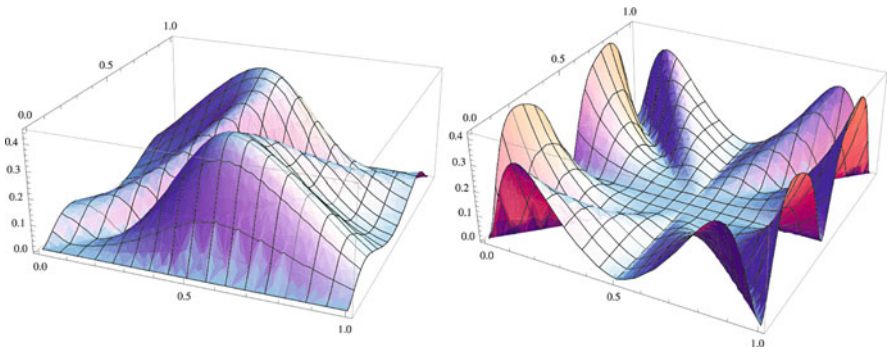

Fig. 6.4 Correcting terms in (6.7) as functions of x_3 (*front axis*) and x for $x_1 = 0$ and $x_2 = 1$ (*left panel*) and $x_2 = 0.5$ and $x_3 = 1 - x_1$ (*right panel*)

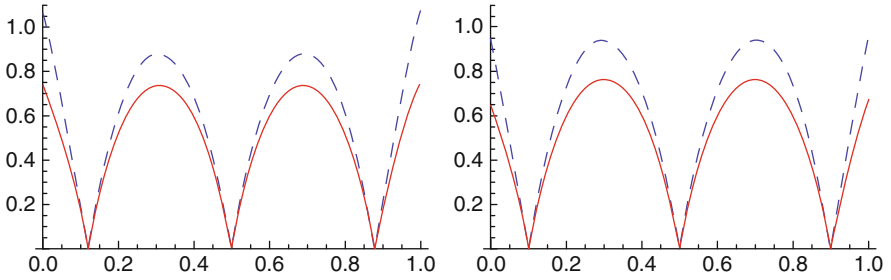

Fig. 6.5 Kriging variance (*solid line*) and "corrected" kriging variance (*dashed line*) as a function of x for $x_1^* = 0.120 = 1 - x_3^*$ (*left panel*) and $x_1^* = 0.099 = 1 - x_3^*$ (*right panel*)

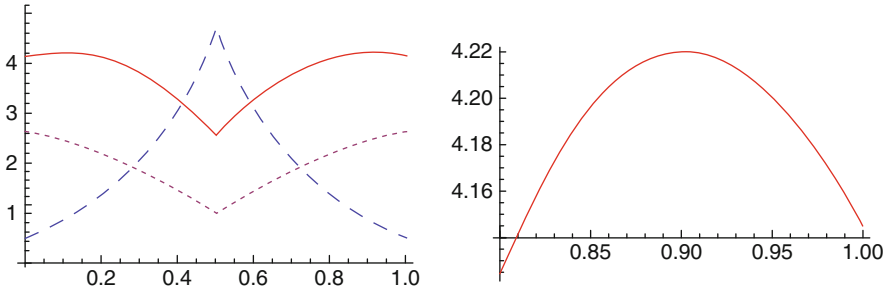

Fig. 6.6 $3|M_\beta|$ (*dotted*), $|M_\rho|/100$ (*dashed*) and $\Phi[x_3|\alpha = 0.8477]$ (*solid*) as functions of $x_3 = 1 - x_1$ (*left panel*) and enlarged portion of the graph (*right panel*)

Once more we are looking for an α, which will yield a similar design for the compound criterion as for EK-optimality. The functions $|M_\beta|$ and $|M_\rho|$ are as expected with spikes at $x_1 \to 1 - x_3 = 0.5$ and yet again we can choose $\alpha = 0.8477$ to yield exactly $x_1^* = 0.099 = 1 - x_3^*$ (see Fig. 6.6), the EK-optimal design.

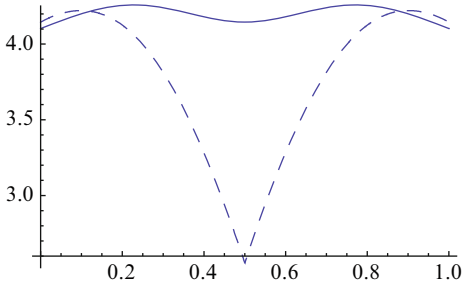

Fig. 6.7 $\Phi[x_3|\alpha = 0.8477]$ as functions of x_3 with $x_1 = 0, x_2 = 1$ (*solid line*) and $x_3 = 1 - x_1$ with $x_2 = 0.5$ (*dashed line*)

6.4.3 No Design Point Fixed (Three Point Optimal Designs)

Let us now finally lift all restrictions and allow the three points to vary freely. A grid search reveals that the EK-optimal design is indeed $\{0.0975, 0.5, 0.9025\}$. Unfortunately now, the $\Phi[\cdot\,|\,\alpha]$-optimal designs are asymmetric for all $\alpha \in [0.8021; 0.919]$, so exact correspondence of the optimal designs cannot be achieved for these α-values. For $\alpha > .919$ the $\Phi[\cdot\,|\,\alpha]$-optimal design is $\{0, 0.5, 1\}$ which is also different from the EK-optimal design. Finally, for $\alpha < 0.8021$ the $\Phi[\cdot\,|\,\alpha]$-optimal design collapses to a two-point design with one double point and for $\alpha < 0.4785$ to a one-point design with a triple point. Thus even in this simple example the conjecture in the strict sense is disproved. However, a plot of the criterion (2) with $\alpha = 0.8477$ as a function of $x_3 = 1 - x_1$ when $x_2 = 0.5$ and as a function of x_3 when $x_1 = 0, x_2 = 1$ (the optimum design being obtained in the latter case, for the value $x_3 = 0.772$ or $x_3 = 0.228$), reveals that the EK-optimal design is only marginally suboptimal for $\Phi[\cdot\,|\,0.8477]$ (see Fig. 6.7). Thus we can expect local optima (e.g. yielded from an exchange algorithm) to still perform rather well.

6.4.4 Relative Efficiencies of the $\Phi[\cdot\,|\,\alpha]$-Optimal and EK-Optimal Designs

Let us now more closely analyze whether the $\Phi[\cdot\,|\,\alpha]$-optimal design is EK-efficient and vice versa. We define the EK-efficiency of $\Phi[\cdot\,|\,\alpha]$-optimal designs $\xi_{\Phi[\cdot\,|\,\alpha]}$ as the ratio of the EK criterion function of the EK-optimal design ξ_{EK} and $\xi_{\Phi[\cdot\,|\,\alpha]}$:

$$\mathrm{Eff}_{\mathrm{EK}}^{-1}(\xi_{\Phi[\cdot\,|\,\alpha]}) = \frac{\max_{x\in\mathscr{X}} \widehat{MSPE}_{\xi_{\Phi[\cdot\,|\,\alpha]}}(x,\gamma)}{\max_{x\in\mathscr{X}} \widehat{MSPE}_{\xi_{\mathrm{EK}}}(x,\gamma)}$$

If $\xi_{\Phi[\cdot\,|\,\alpha]}$ is not unique for some α, we take the $\Phi[\cdot\,|\,\alpha]$-optimal design with the highest EK-efficiency. Analogously, the $\Phi[\cdot\,|\,\alpha]$-efficiency of EK-optimal designs

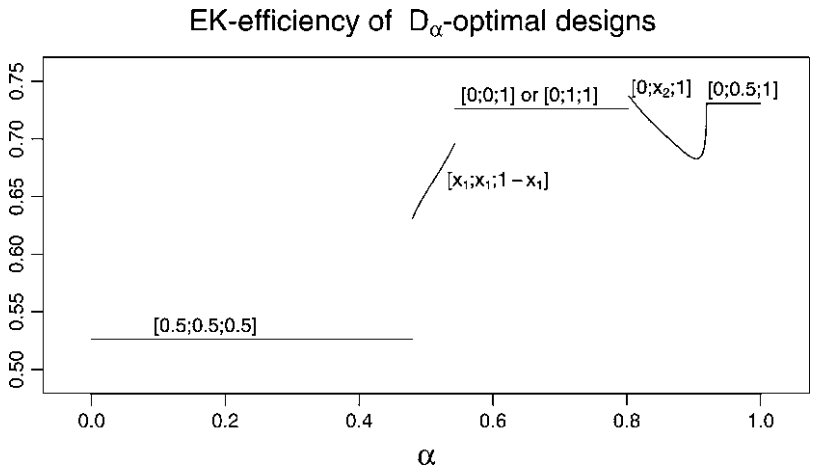

Fig. 6.8 EK-efficiencies of $\xi_{\Phi[\cdot\,|\,\alpha]}$ and different α-values

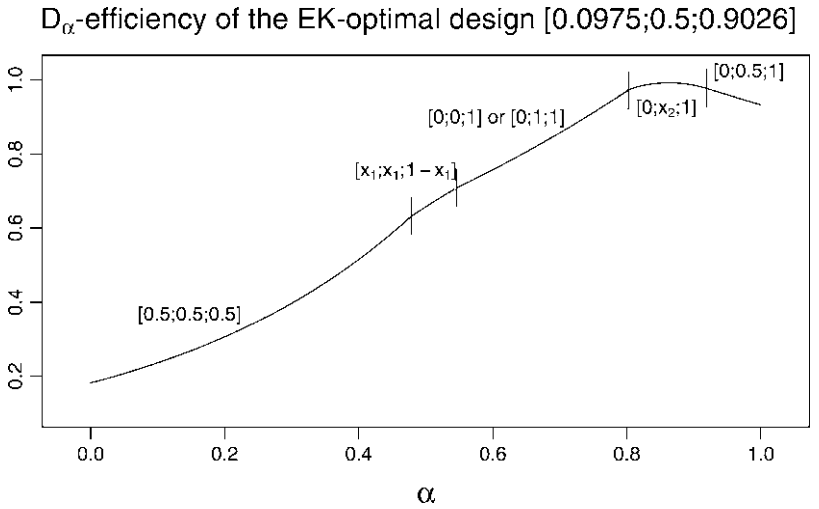

Fig. 6.9 $\Phi[\cdot\,|\,\alpha]$-efficiencies of $\xi_{\text{EK}} = \{0.0975, 0.5, 0.9026\}$ and different α-values

is defined as the ratio of the $\Phi[\cdot\,|\,\alpha]$ criterion function of EK-optimal and $\Phi[\cdot\,|\,\alpha]$-optimal designs:

$$\text{Eff}^{-1}_{\Phi[\cdot\,|\,\alpha]}(\xi_{EK}) = \frac{\Phi[\xi_{\Phi[\cdot\,|\,\alpha]}|\alpha]}{\Phi[\xi_{\text{EK}}|\alpha]} = \frac{|M_\beta(\xi_{\Phi[\cdot\,|\,\alpha]}, \gamma)|^\alpha \cdot |M_\gamma(\xi_{\Phi[\cdot\,|\,\alpha]}, \gamma)|^{(1-\alpha)}}{|M_\beta(\xi_{\text{EK}}, \gamma)|^\alpha \cdot |M_\gamma(\xi_{\text{EK}}, \gamma)|^{(1-\alpha)}}$$

Continuing our example of Sect. 6.4.3 we computed the $\Phi[\cdot\,|\,\alpha]$- and EK-efficiencies for all α-values. The result can be seen in the Figs. 6.8 and 6.9. In the graphs also the

different $\Phi[\cdot\,|\,\alpha]$-optimal designs and the corresponding α-intervals are specified. The best EK-efficiency of $\Phi[\cdot\,|\,\alpha]$-optimal designs is reached at $\alpha = 0.8021$, $\xi_{\Phi[\cdot\,|\,\alpha=0.8021]} = \{0, 0.0574, 1\}$ has an EK-efficiency of 73.7%. If we look at the $\Phi[\cdot\,|\,\alpha]$-efficiencies of the EK-optimal design $\xi_{\mathrm{EK}} = \{0.0975, 0.5, 0.9026\}$ we get the highest efficiency for $\alpha = 0.8629$: $\mathrm{Eff}_{\Phi[\cdot\,|\,\alpha=0.8629]}(\xi_{EK}) = 99.23\%$.

6.4.5 More Design Points and Different Linear Predictors

In a first generalization of our example in Sect. 6.4.3 we studied larger designs for the same model. The EK-optimal design for this model is always symmetric about the center of the design interval and the design points are spaced almost equidistant. The criterion (6.6) is minimal when $\widehat{MSPE}_{\xi}(x, \gamma)$ has its maximum at the boundaries of the design interval and the midpoints between the design points simultaneously (see Fig. 6.5 for $n = 3$).

For $n > 3$ the highest EK-efficiency for $\Phi[\cdot\,|\,\alpha]$-optimal designs is reached when $\xi_{\Phi[\cdot\,|\,\alpha]}$ is equidistant. The corresponding α-values are $\alpha \in [0.816; 1]$ for $n = 4$, $\alpha \in [0.8075; 1]$ for $n = 5$ etc. i.e. the $\Phi[\cdot\,|\,\alpha = 1]$-optimal design is always the design with the highest EK-efficiency. The maximal EK-efficiency for $\xi_{\Phi[\cdot\,|\,\alpha]}$ increases with n (82.84% for $n = 4$, 88.63% for $n = 5$ etc.).

For $n > 3$ the maximal $\Phi[\cdot\,|\,\alpha]$-efficiency for the EK-optimal designs is also reached when $\xi_{\Phi[\cdot\,|\,\alpha]}$ is equidistant. Since the maximum of the criterion function (6.2) decreases with decreasing α and the $\Phi[\cdot\,|\,\alpha]$-optimal design is equidistant for $\alpha \in [\alpha_n; 1]$ we get the maximal $\Phi[\cdot\,|\,\alpha]$-efficiency of EK-optimal designs for $\alpha = \alpha_n$. For $n = 4$ we have $\alpha_4 = 0.816$ and a maximal $\Phi[\cdot\,|\,\alpha]$-efficiency of 97.76%, for $n = 5$ we have $\alpha_5 = 0.8075$ and a maximal $\Phi[\cdot\,|\,\alpha]$-efficiency of 96.72% etc.

With increasing n the EK-optimal design converges to the equidistant design which is also $\Phi[\cdot\,|\,\alpha]$-optimal for $\alpha = 1$. Therefore both relative efficiencies converge to 1 in our example.

The reported results only change slightly if we change the linear predictor of our example to $\eta(x, \beta) = \beta_0 + \beta_1 x$: The $\Phi[\cdot\,|\,\alpha]$-optimal design with $\alpha = 1$ always has maximal EK-efficiency (81.86% for $n = 3$, 72.53% for $n = 4$, 81.39% for $n = 5$ etc.). We could also observe a maximal $\Phi[\cdot\,|\,\alpha]$-efficiency of EK-optimal designs which is similar to the example without regressor (70.39% for $n = 3$ at $\alpha = 0.8343$, 78.8% for $n = 4$ at $\alpha = 0.7182$ etc.). Also here the relative efficiencies converge to 1 as n increases.

The presented examples, albeit of only seemingly limited scope, give hope that the accordance formulated in Sect. 6.3 (a quasi KWET-type relationship) will continue to hold in more complex and realistic settings. They also show that the EK-optimal designs (and the corresponding $\Phi[\cdot\,|\,\alpha]$-optimal designs) can be quite far from the often suggested and frequently employed equidistant space-filling designs.

6.5 Outlook

Suppose therefore that a value α can always be found such that the optimal design for problem (6.2) is reasonably efficient for (6.7). It is then much advantageous to solve (6.2) rather than minimize (6.5) since the former does not require maximization over the candidate set $\mathscr{X}$. Several approaches have been suggested to solve (6.2) with $\alpha = 1$, see [6, 12]. It remains to be checked whether such approaches can be used when $\alpha < 1$. The main difficulty here is to choose a suitable α beforehand. From the discontinuities in Fig. 6.8 one may guess that this is a formidable task. Since it is reasonable in most applications to assume that $\mathscr{X}$ is finite, the evaluation of $M\widehat{MSPE}_\xi(\gamma) = \max_{x\in\mathscr{X}} \widehat{MSPE}_\xi(x,\gamma)$ has a moderate computational cost. One can then simply compute optimal designs for (6.2) for a series of values of α, and retain the best one in terms of $M\widehat{MSPE}_\xi(\gamma)$.

The corrected MSPE (6.5) can also be used for the sequential construction of designs. Let ξ^{n_0} denote some initial design of size n_0. At step k of such a sequential construction, $k \geq n_0$, choose x_{k+1} as

$$x_{k+1} = \arg\max_{x\in\mathscr{X}} \widehat{MSPE}_{\xi^k}(x,\gamma)$$

and then update ξ^k into $\xi^{k+1} = \{\xi^k, x_{k+1}\}$. Again, when $\mathscr{X}$ is finite, the sequential construction above has a moderate computational cost. If one wishes to minimize the integrated MSPE, $I\widehat{MSPE}_\xi(\gamma) = \int_{\mathscr{X}} \widehat{MSPE}_\xi(x,\gamma)\mu(dx)$, for some measure of interest μ, one can choose instead at step k

$$x_{k+1} = \arg\min_{x\in\mathscr{X}} \int_{\mathscr{X}} \widehat{MSPE}_{\{\xi^k,x\}}(x,\gamma)\mu(dx)\,.$$

The parameters γ can be estimated after each generation of a new sampling point x_{k+1}, rendering the sequential designs above adaptive. Algorithms for the construction of adaptive designs for (6.2) are also of interest. We believe that such investigations could yield the development of cheap algorithms for the sequential construction of designs that would take into account the prediction task and at the same time the reduction of uncertainty in the estimation of the covariance parameters, thereby following the same ultimate objective as designs optimal in the sense of (6.7). The fact that in most applications $\mathscr{X}$ is finite might reveal particularly useful for studying the convergence properties of such adaptive procedures, see [17] for such developments in the case of uncorrelated errors.

Contrasting with the uncorrelated case, non-additivity and nonconvexity are amongst the obstacles for constructing optimal designs for random fields, which have recently been reviewed in [14] (not to mention the potential existence of local extrema, see eg. Fig. 2 of [20]). Furthermore the concept of Fisher information is conveniently used as a basis for designing efficient experiments. However, if the output stems from correlated random fields as (6.1), the conditions under which

Fisher information may be suitable must be restated. For some small sample results see also [7].

A last point that requires investigation concerns the estimability of the random-field parameters σ^2 and γ. Under the infill design framework (i.e., when the design space is compact) typically not all parameters are estimable, only some of them being micro-ergodic, see [19]. However, it seems reasonable to consider that Jeffrey's law will apply and that parameters that are not estimable from the data $Y_t(\xi)$ should have little influence on predictions (interpolations/extrapolations) for the random field.

Acknowledgements The article was mainly written during the first author's research stay at the University of Nice-Sophia Antipolis and he wants to acknowledge its generous support. This work was also partially supported by a PHC Amadeus/OEAD Amadée grant FR11/2010. We are also grateful to an attentive referee for pointing out several omissions and typos.

References

1. Abt, M.: Estimating the prediction mean squared error in gaussian stochastic processes with exponential correlation structure. Scand. J. Stat. **26**(4), 563–578 (1999)
2. Baldi Antognini, A., Zagoraiou, M.: Exact optimal designs for computer experiments via kriging metamodelling. J. Stat. Plan. Infer. **140**(9), 2607–2617 (2010) URL http://dx.doi.org/10.1016/j.jspi.2010.03.027
3. Bursztyn, D., Steinberg, D.: Comparison of designs for computer experiments. J. Stat. Plan. Infer. **136**(3), 1103–1119 (2006) URL http://dx.doi.org/10.1016/j.jspi.2004.08.007
4. Cressie, N.A.C.: Statistics for Spatial Data (Wiley Series in Probability and Statistics), rev sub Edition. Wiley-Interscience, New York (1993)
5. Dette, H., Kunert, J., Pepelyshev, A.: Exact optimal designs for weighted least squares analysis with correlated errors. Stat. Sinica. **18** (1), 135–154 (2008)
6. Fedorov, V., Müller, W.: Optimum design for correlated processes via eigenfunction expansions. In: López-Fidalgo, J., Rodríguez-Díaz, J., Torsney, B. (Eds.), mODa'8 – Advances in Model–Oriented Design and Analysis, Proceedings of the 8th Int. Workshop, Almagro (Spain). Physica Verlag, Heidelberg, 57–66 (2007)
7. Ginsbourger, D., Dupuy, D., Badea, A., Carraro, L., Roustant, O.: A note on the choice and the estimation of kriging models for the analysis of deterministic computer experiments. Appl. Stoch. Model. Bus. **25**(2), 115–131 (2009) URL http://dx.doi.org/10.1002/asmb.741
8. Harman, R., Stulajter, F.: Optimal prediction designs in finite discrete spectrum linear regression models. Metrika, **72**, 281-294 (2009) URL http://dx.doi.org/10.1007/s00184-009-0253-4
9. Harville, D.A., Jeske, D.R.: Mean squared error of estimation or prediction under a general linear model. J. Am. Stat. Assoc. **87**(419), 724–731 (1992)
10. Kiefer, J., Wolfowitz, J.: The equivalence of two extremum problems. Canadian. J. Math. **12**, 363–366 (1960)
11. Kislák, J., Stehlík, M.: Equidistant and D-optimal designs for parameters of Ornstein-Uhlenbeck process. Stat. Probabil. Lett. **78**(12), 1388–1396 (2008)
12. Müller, W., Pázman, A.: Measures for designs in experiments with correlated errors. Biometrika. **90**(2), 423–434 (2003)
13. Müller, W.G.: Collecting Spatial Data, 3rd Edition. Springer, Heidelberg (2007)
14. Müller, W. G., Stehlík, M.: Issues in the optimal design of computer simulation experiments. Appl. Stoch. Model. Bus. **25**(2), 163–177 (2009)

15. Müller, W. G., Stehlík, M.: Compound optimal spatial designs. Environmetrics **21**(3-4), 354–364 (2009) URL http://dx.doi.org/10.1002/env.1009
16. Pronzato, L.: Optimal experimental design and some related control problems. Automatica **44**(2), 303–325 (2008)
17. Pronzato, L.: One-step ahead adaptive D-optimal design on a finite design space is asymptotically optimal. Metrika **71**(2), 219–238 (2010) URL http://dx.doi.org/10.1007/s00184-008-0227-y
18. Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and analysis of computer experiments. Stat. Sci. **4**(4), 409–423 (1989)
19. Stein, M.L.: Interpolation of Spatial Data: Some Theory for Kriging. (Springer Series in Statistics), 1st Edition. Springer (1999)
20. Torsney, B., Martín-Martín, R.: Multiplicative algorithms for computing optimum designs. J. Stat. Plan. Infer. **139**(12), 3947–3961 (2009) URL http://dx.doi.org/10.1016/j.jspi.2009.05.007
21. Ying, Z.: Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. J. Multivariate Anal. **36**, 280–296 (1991)
22. Zagoraiou, M., Antognini, A.B.: Optimal designs for parameter estimation of the Ornstein-Uhlenbeck process. Appl. Stoch. Model. Bus online (**5**), 583–600 (2009) URL http://dx.doi.org/10.1002/asmb.749
23. Zhu, Z., Stein, M.: Spatial sampling design for parameter estimation of the covariance function. J. Stat. Plan. Infer. **134**(2), 583–603 (2005)
24. Zhu, Z., Stein, M.L.: Spatial sampling design for prediction with estimated parameters. J Agr Biol Envir St **11**(1), 24–44 (2006)
25. Zhu, Z., Zhang, H.: Spatial sampling design under the infill asymptotic framework. Environmetrics **17**(4), 323–337 (2006)
26. Zimmerman, D.L.: Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics. **17**(6), 635–652 (2006)

Chapter 7
Modeling Spatial and Spatio-Temporal Non Gaussian Processes

Denis Allard

7.1 Introduction

The ubiquitous assumption of normality for modeling spatial and spatio-temporal data can be understood for many reasons. A major one is that the multivariate normal distribution is completely characterized by its first two moments. In addition, the stability of multivariate normal distribution under summation and conditioning offers tractability and simplicity. Gaussian spatial processes are well modeled and understood by the statistical and scientific communities, but for a wide range of environmental applications Gaussian spatial or spatio-temporal models cannot reasonably be fitted to the observations.

The purpose of this chapter is to offer a brief and partial review of some spatial or spatio-temporal models for data that are positive, skewed or long-tailed. In a word, models for data that do not fit the Gaussian assumption.

While the definition of a Gaussian spatial random process is precise, different approaches can be proposed for building non Gaussian spatial processes. We shall review two distinct approaches. A classical approach consists in transforming a spatial Gaussian process in a way that fits the data. Expressed the other way around, it consists in transforming the data into values that are compatible with a Gaussian assumption. Another strategy is to assume that the random field follows specific processes, such as χ^2 or t processes (see e.g. [1] or [2]) or Gamma processes, see [33].

This chapter is organized in three sections. In this introductory section, some definitions and properties of Gaussian spatial processes that will be needed in the rest of this chapter are first recalled; one example of a process with Gaussian univariate marginal which is not a Gaussian process is also given. In the second section, we shall present the class of the transformed Gaussian processes, which

D. Allard (✉)
Biostatistics and Spatial Processes (BioSP), INRA, Avignon, France
e-mail: allard@avignon.inra.fr

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_7,

consists in transforming each value of a Gaussian process with a fixed function. Besides the problem of choosing the adequate transform, there are a few theoretical pitfalls associated with this method, which will be presented. In the third section, we illustrate the second approach by presenting the spatial skew-normal processes.

We shall specifically focus on the modeling of the covariance function (and/or variogram) for these non Gaussian processes. We shall see that the question of how building valid classes of variogram is still an open problem for some of the models presented.

When possible, we shall provide illustrations with simulated or environmental data.

7.1.1 Gaussian Spatial Random Processes

Let us recall briefly the definition and some properties of Gaussian Random Processes. Let us consider a continuous domain of interest, $\mathscr{D} \in \mathbf{R}^d$, where typically $d = 2$ or $d = 3$ and let us denote $s_1, \ldots, s_n$ a set of locations in $\mathscr{D}$. The distribution of the process $\{Y(s) \ : \ s \in \mathscr{D}\}$ is given by the collection of all finite-dimensional joint distributions

$$F(y_1, \ldots, y_n; s_1, \ldots, s_n) = P(Y(s_1) \leq y_1, \ldots, Y(s_n) \leq y_n), \tag{7.1}$$

for all n and all collections $(s_1, \ldots, s_n)$ of sites in $\mathscr{D}$. The Kolmogorov existence theorem [7] states that the stochastic process model P is valid if the family of the finite dimensional joint distributions (7.1) is consistent under permutation and marginalization.

A spatial random process $\{Y(s) \ : \ s \in \mathscr{D}\}$ is a multi-Gaussian random process, or a *Gaussian random process* for short, if all finite dimensional distributions (7.1) are multivariate Gaussian, that is if the probability density function of the vector of values $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))'$ is

$$f(\mathbf{y}) = (2\pi)^{-n/2} \det(\mathbf{C})^{-1/2} \exp\{-0.5(\mathbf{y} - \boldsymbol{\mu})'\mathbf{C}^{-1}(\mathbf{y} - \boldsymbol{\mu})\}, \tag{7.2}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\mu} = (E[Y(s_1)], \ldots, E[Y(s_n)])'$ is the vector of expectations and $\mathbf{C}$ is the covariance matrix whose elements are $\mathbf{C}_{[i,j]} = \mathrm{Cov}(Y(s_i), Y(s_j))$. The elements of the covariance matrix are such that $\mathrm{Cov}(Y(s_i), Y(s_j)) = C(s_i, s_j)$ for some real valued function $C(\cdot, \cdot)$ on $\mathscr{D} \times \mathscr{D}$, called the covariance function.

A spatial random process is stationary if the finite dimensional joint distributions are invariant under translation. In the case of a Gaussian random process, this implies that

$$E[Y(s_i)] = \mu; \quad \mathrm{Cov}(Y(s_i), Y(s_j)) = C(s_j - s_i) \tag{7.3}$$

for all $s_i, s_j \in \mathscr{D}$. For non Gaussian random processes, conditions (7.3) refer to *second order stationarity*, also called *weak stationarity*.

The variogram of a spatial random process is defined as

$$\gamma(h) = 0.5\text{Var}\{Y(s+h) - Y(s)\}. \tag{7.4}$$

For weakly stationary random processes, straightforward calculations show that $\gamma(h) = C(0) - C(h)$. There is a less demanding assumption for stationarity than second order stationarity, that only requires the existence of the first and second moments of differences $Y(s+h) - Y(s)$, instead of the existence of first and second moments of $Y(\cdot)$. A random process $Y(\cdot)$ is said to be *intrinsically stationary* if

$$E[Y(s+h) - Y(s)] = 0 \text{ and } \text{Var}(Y(s+h) - Y(s)) = 2\gamma(h), \tag{7.5}$$

for any s and $s + h \in \mathscr{D}$, in which case

$$\gamma(h) = 0.5E[(Y(s+h) - Y(s))^2]. \tag{7.6}$$

It is very often preferable to use the variogram, $\gamma(h)$ instead of the covariance function because (1) the variogram exists for the less demanding assumption of *intrinsically stationarity*; (2) estimation based on variograms are more robust than those based on covariance functions. See [11] and [10] for more details on the estimation of both the variogram and the covariance function for general spatial random processes. See also [35] on how likelihood approaches can be used for Gaussian spatial random processes. From now on, except when explicitly stated otherwise, only stationary random processes for which (7.3) holds will be considered.

Since the covariance matrix $\mathbf{C}$ in (7.2) must be nonnegative definite, the associated covariance function must be positive definite. The classical Bochner theorem [8] states that a real-valued continuous function $C(h)$ is a positive definite function if and only if it is the Fourier transform of a symmetric, nonnegative and finite measure F on $\mathbf{R}^d$ such that $\int_{\mathbf{R}^d} dF(\omega) < \infty$:

$$C(h) = \int_{\mathbf{R}^d} \exp\{i < h, \omega >\} dF(\omega), \tag{7.7}$$

where $< \cdot, \cdot >$ designates the dot product in $\mathbf{R}^d$. $F(\cdot)$ is called the spectral decomposition of the covariance function. More details can be found in [30].

A spatial random process $Y(\cdot)$ in $\mathscr{D}$ is *mean square differentiable* if $E[(Y(s+h) - Y(s))^2] \to 0$ as $||h|| \to 0$. Since $0.5E[(Y(s+h) - Y(s))^2] = \gamma(h) = C(0) - C(h)$ for a second order stationary process, mean square differentiability is equivalent to the covariance function (or, equivalently, the variogram) being continuous at the origin. It can be easily shown [30] that a second order stationary process is p times mean square differentiable if its covariance function (or, equivalently, its variogram) is $2p$ times differentiable at the origin. The exponential covariance function $C(h) = \exp\{-||h||/a\}$ is continuous at the origin. The corresponding process $Y(\cdot)$ is thus mean square continuous but not mean square differentiable. The Gaussian covariance function $C(h) = \exp\{-(||h||/a)^2\}$ is infinitely differentiable

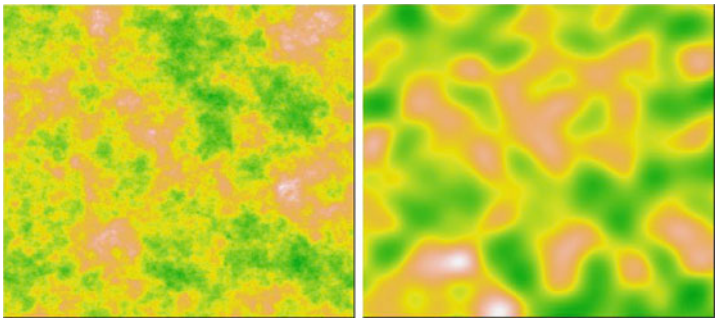

Fig. 7.1 Realizations of $(0, 1)$ Gaussian random processes. *Left*: with exponential covariance. *Right*: with Gaussian covariance

at the origin. Its associated Gaussian random process is thus infinitely mean square differentiable. A realization of both processes is shown in Fig. 7.1.

7.1.2 An Example of a Non Gaussian Random Process with Gaussian Marginals

All Gaussian spatial random processes have Gaussian univariate marginals, but the converse is obviously not true. As an example of a non Gaussian random process with Gaussian marginals, we shall consider the Poisson lines tessellation model with Gaussian random values in cells. A tessellation model randomly partitions the domain into non overlapping cells. To each cell of the tessellation is independently assigned a random value. Here, we shall draw from the standard Gaussian distribution. The Poisson tessellation model is generated by Poisson random lines. A line in $\mathbf{R}^2$ is fully specified by two parameters: a direction, $\alpha \in [0, 2\pi)$, and a distance to the origin, $d > 0$. Poisson random lines correspond to an homogeneous Poisson process for (α, d) in $[0, 2\pi) \times \mathbf{R}_+$ The number of lines intersecting the domain is drawn from a Poisson distribution with parameter $\lambda(\mathscr{D})$. Each line is then drawn with independent uniform distributions for α and d, subject to the condition that the line intersects the domain. The only parameter driving the tessellation is thus the parameter $\lambda(\mathscr{D})$: to large values of $\lambda(\mathscr{D})$ correspond small cells and vice versa. Regarding the random values in the cells, we shall consider here standard $(0, 1)$ Gaussian random variables. This hierarchical model is strictly stationary. It defines a random process $Y_M(\cdot)$ whose univariate marginals are Gaussian, but which is obviously not multi-Gaussian. [19] shows that the variogram of $Y_M(\cdot)$ has an exponential form

$$\gamma_M(h; a_M) = 1 - \exp\{-||h||/a_M\}, \quad a_M = L(\mathscr{D})/(2\lambda(\mathscr{D})), \tag{7.8}$$

where $L(\mathscr{D})$ is the perimeter of the domain $\mathscr{D}$.

As a consequence of (7.8), for a well chosen parameter $\lambda(\mathscr{D})$ it is possible to define a second order stationary process with same marginal univariate Gaussian and same exponential variogram as a (multi-) Gaussian process $Y_G(\cdot)$. These models are thus indistinguishable if we restrict ourselves to first and second order moments.

While expectation and covariance functions are similar, it is not the case of the first order variogram, also called *madogram* [10]. It is defined as half the expectation of the absolute difference between $Y(s+h)$ and $Y(s)$:

$$\gamma^1(h) = 0.5E[|Y(s+h) - Y(s)|]. \tag{7.9}$$

The madogram of a stationary Gaussian random process $Y_G(s)$ with variogram $\gamma_G(h; a_G)$ can easily be computed. The difference $Y_G(s+h; a_G) - Y_G(s; a_G)$ is a Gaussian random variable with 0 mean and variance equal to $2\gamma(h; a_G)$. Recalling that the expectation of the absolute value of a Gaussian random variable X with zero mean and variance τ^2 is $E[|X|] = \sqrt{2/\pi}\tau$, one gets that

$$\gamma^1_G(h; a_G) = \sqrt{\gamma_G(h; a_G)}/\sqrt{\pi}.$$

For the Poisson tessellation model, it can be shown (see below) that the first order variogram is

$$\gamma^1_M(h; a_M) = \gamma_M(h; a_M)/\sqrt{\pi}.$$

In short, the relationship between first order and second order variograms is quadratic for Gaussian random process, and it is linear for Poisson tessellation model. We thus have a powerful tool for discriminating between these two models.

The two models are now combined to define mixture models

$$Y_w(s) = wY_G(s; a_G) + (1 - w^2)^{1/2} Y_M(s; a_M), \quad 0 \le w \le 1, \tag{7.10}$$

which have same expectation and same exponential covariance function for all w if $a_G = a_M$. Note that if $a_G \neq a_M$ it is still not possible to attribute which part of the variogram is related to the tessellation model and which part is related to the multi-Gaussian model. Note that negative weights could also be considered, but because the Gaussian fields are centered we can set the parametrization (7.10) without loss of generality. Figure 7.2 shows a realization of a Poisson tessellation model, a multi-Gaussian random process, and a mixture model as defined in (7.10), with exactly same univariate $(0, 1)$ Gaussian distribution and same exponential covariance. The experimental histograms and variograms computed on these realizations, not shown here, are very similar.

It can be shown [14] that the first order variogram for the mixture model is

$$\gamma^1_w(h) = \Big\{ w(1 - \gamma_M(h; a_m))\sqrt{\gamma_G(h; a_G)} + \gamma_M(h; a_m)\sqrt{w^2\gamma_G(h; a_G) + (1 - w^2)} \Big\} / \sqrt{\pi}. \tag{7.11}$$

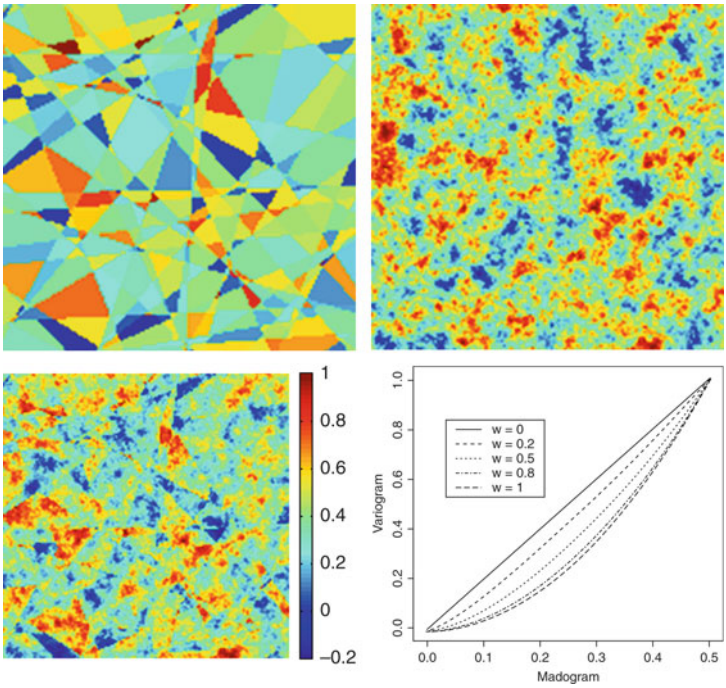

Fig. 7.2 *Top*: Realizations of a Poisson random lines tessellation process (*left*) and of a Gaussian random process (*right*) with same (0, 1) univariate marginal and same exponential model. *Bottom*: Left, mixture model with $w^2 = 0.5$ (*left*); Right, relationship between variogram and madogram of the mixture model

The proof of this result relies on the computation of $|Y_w(s+h) - Y_w(s)|$ conditional on the event $\{s$ and $s+h$ belong to the same cell$\}$. Equation (7.11) shows that even when $a_G = a_M$, the first order variogram depends on w. The relationship between the variogram and the madogram in (7.11), from linear for $w = 0$ to quadratic for $w = 1$, is shown in the bottom right panel of Fig. 7.2. Simultaneous computation of variograms and madograms can thus be used to estimate w, and consequently to discriminate between a multi-Gaussian structure and a mosaic structure.

This approach has been proposed in [14] to discriminate between natural and cultivated landscapes on remote sensing images of vegetation indices. Figure 7.3 shows a vegetation index measured from the SPOT-HRV satellite with a 20 m spatial resolution. The left-hand-side image is an agricultural site near Avignon, France. The right-hand-side image is a natural vegetation site near Montpellier, France. On crop sites the fraction $(1 - w^2)$ of the variance explained by the tessellation component of the mixture model is generally high or very high. The associated range a_M could be related to the mean size of the agricultural fields, even though the experimental variogram could be disturbed by the gathering of fields with similar vegetation index. On natural and forest vegetation sites, the fraction of variance w^2

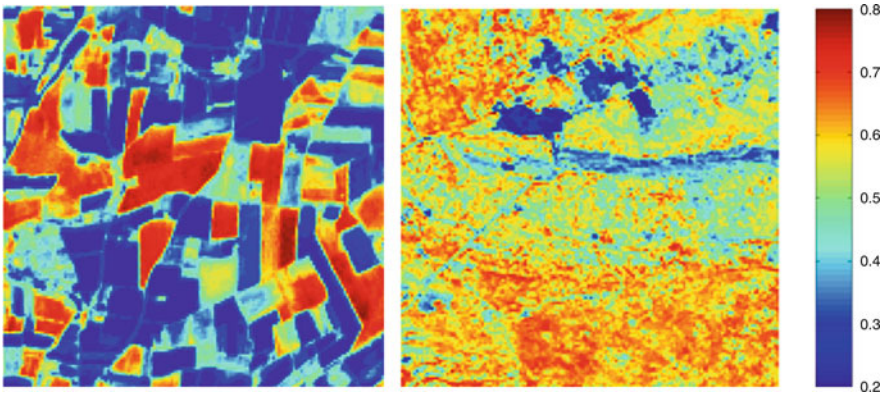

Fig. 7.3 Examples of landscape structures characterized from high spatial resolution NDVI images (SPOT-HRV sensor at 20 m spatial resolution). *Left*: agricultural site near Avignon, France. *Right*: natural vegetation site near Montpellier, France

explained by the multi-Gaussian model is, on the contrary, generally quite large. On the sites shown in Fig. 7.3, it was found that $w^2 = 0.07$ on the agricultural site and $w^2 = 0.42$ on the natural vegetation site.

7.2 Transformed Gaussian Random Process

Consider a $(0, 1)$ Gaussian random process $Y(\cdot)$ with correlation function $\rho(h) = C(h)/C(0)$ on $\mathscr{D}$. Let $\phi(y)$ be a function from $\mathbf{R}$ to $E \subset \mathbf{R}$. The random process defined by $Z(s) = \phi\{Y(s)\}$, $s \in \mathscr{D}$, is a *transformed* Gaussian random process. We shall first consider strictly monotonic transforms which are invertible. In the next subsection, we shall consider more general transforms that are not one-to-one correspondences.

7.2.1 Strictly Monotonic Transforms

When the transform function ϕ admits a proper inverse, the general approach for dealing with transformed spatial or spatio-temporal Gaussian data is thus the following: (1) estimate the parameters of ϕ and compute the normal score of the data, i.e. compute $Y(s_i) = \phi^{-1}\{Z(s_i)\}$ for $i = 1, \ldots, n$; (2) estimate the covariance function on the transformed data; (3) do spatial prediction $Y^p(\cdot)$ and/or stochastic simulations $Y^s(\cdot)$ conditional on the data $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))'$; (4) back-transform the predictions/simulations with ϕ. Care must be taken to the back-transform step. While $Y^s(\cdot)$ distribution is by construction the same as $\mathbf{Y}$, this is not the case for $Y^p(\cdot)$. For example, if prediction is made by simple kriging

(see e.g. [10] for details on the practice of kriging), $\mathrm{Cov}(Y^p(s), Y^p(s+h)) = C(h) - \mathbf{C}(s)'\mathbf{C}^{-1}\mathbf{C}(s+h)$, where $\mathbf{C}(s) = (C(s-s_1), \ldots, C(s-s_n))'$. The covariance function of the predicted process is thus different than the one of the original process. Ignoring the variance reduction in the back-transform step can lead to serious bias. It is thus necessary to take it explicitly into account, which can be done if one specifies the transform function ϕ.

Probably the most intuitive way of transforming the data is by applying a parametric transform. Many environmental data are positive and/or long-tailed. It is for example the case for soil humidity, pollutant grades, wind speed, radiation, etc. For these variables, transforms related to the power and to the exponential functions are quite common. In the geostatistical literature, a semi-parametric approach based on the orthogonal Hermite polynomials is also proposed, see e.g. [10], [32]. This approach will not be presented here. The reader is referred to the references above.

7.2.1.1 Lognormal Spatial Processes

A positive random process $Z(\cdot)$ is a *lognormal spatial process* on $\mathscr{D}$ if and only if there exists μ and $\sigma > 0$ such that

$$\ln Z(s) = \mu + \sigma Y(s), \quad \text{for all } s \in \mathscr{D}, \tag{7.12}$$

where $Y(\cdot)$ is a $(0,1)$ Gaussian random process with correlation function $\rho(h)$. In other words, the transform function is $\phi(y) = \exp\{\mu + \sigma y\}$. Let m, $C_Z(h)$ and $\gamma_Z(h)$ denote respectively the expectation, covariance function and variogram of the lognormal spatial process $Z(\cdot)$. Using the general result $E[e^{aX}] = e^{a^2/2}$ when X is $(0,1)$ Gaussian random variable, it is straightforward to show that

$$\begin{aligned} E[Z(s)] &= e^{\mu} e^{\sigma^2/2} = m \\ \mathrm{Cov}(Z(s), Z(s+h)) &= m^2 \left(e^{\sigma^2 \rho(h)} - 1 \right) = C_Z(h). \end{aligned} \tag{7.13}$$

For the variogram one gets from (7.13)

$$\gamma_Z(h) = m^2 e^{\sigma^2} \left(1 - e^{-\sigma^2 \gamma(h)} \right). \tag{7.14}$$

Denoting $\tau^2 = m^2 e^{\sigma^2}$ and $g(h) = e^{-\sigma^2 \gamma(h)}$, (7.14) yields

$$\gamma_Z(h) = \tau^2 (g(0) - g(h)),$$

which shows that $g(h)$ is a covariance function. This result is thus a partial proof by construction of the following theorem.

Theorem 7.1. *If $\gamma(h)$ is a variogram, $e^{-t\gamma(h)}$ is a covariance function for all $t > 0$.*

This theorem is a particular case of the more general Schoenberg theorem [29] which states that $\gamma(\cdot)$ is conditionally negative definite if and only if $\exp(-t\gamma(\cdot))$ is positive definite for all positive t. See also [27] and [28] on the use of Schoenberg theorem for building valid classes of space-time covariances.

The construction above assumed a weakly stationary random processes $Y(\cdot)$ It is in fact possible to extend the definition of lognormal processes to intrinsically stationary processes with unbounded variogram $\gamma(h)$. But in this case (7.12) cannot be applied since μ and σ do no longer exist. When conditioning $\{Y(s) : s \in \mathscr{D}\}$ to its average on a domain $V \supset \mathscr{D}$, the process $\{Z(s) = \exp\{Y(s)\} : s \in \mathscr{D}\}$ is a locally stationary lognormal random process on $\mathscr{D}$. In this framework it is shown in [20] that

$$\gamma_Z(h) = b(1 - \exp\{-\gamma(h)\}) \tag{7.15}$$

for some $b > 0$. Equation (7.15) indicates that the exponential variogram is a valid model for lognormal processes, if and only if the variogram of the underlying Gaussian process is intrinsically stationary, with a linear variogram $\gamma(h) = a||h||$ for some $a > 0$.

Equation (7.14) establishes the parametric form of the variogram of a lognormal spatial process. Any valid correlation function $\rho(h)$ (i.e. a positive definite function with $\rho(0) = 1$) will lead to a valid covariance function for the process $Z(\cdot)$ through (7.13). If we decide instead to directly model the covariance of $Z(\cdot)$, it must be of the specific form (7.14). In other words, not only $C_Z(h)$ but also $\ln(C_Z(h)/m^2 + 1)$ must be positive definite, which excludes certain models. [22] proves for example that the spherical model is not compatible with bi-lognormality[1] for certain values of the variance. It is further conjectured that the spherical covariance function is not a valid model for lognormal processes.

Let us go back to the problem of prediction. For Gaussian random processes, the conditional expectation, which is the optimal predictor, is linear with respect to the data. That is $Y^p(s_0) = E[Y(s_0) \mid \mathbf{Y}] = \boldsymbol{\lambda}'\mathbf{Y}$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$ is the solution of the kriging equation (see e.g. [11], [32]). Then the optimal predictor for $Z(s_0)$ is

$$Z^p(s_0) = E[Z(s_0) \mid \mathbf{Y}] = \exp\left\{ Y^p(s_0) + 0.5 \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j (1 - \rho(s_i - s_j)) \right\}, \tag{7.16}$$

and

$$\mathrm{Var}(Z^p(s_0) - Z(s_0)) = (Z^p(s_0))^2 \left[\exp\left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j (1 - \rho(s_i - s_j)) \right\} - 1 \right].$$

[1] The logarithm of a bi-lognormal process is a process for which all univariate and bivariate marginals are Gaussian. It is thus a weaker assumption than the lognormality considered here.

Equation (7.16) shows that $Z^p(s_0)$ is not simply the exponential transform of $Y^p(s_0)$. It does include a correction factor based on the prediction variance. As a first consequence, $Z^p(s_0)$ is not unbiased anymore. Working with quantiles rather than moments might be an interesting alternative because quantiles simply follow the transformation: since the distribution of the error $Y^p(s_0) - Y(s_0)$ is symmetric (Gaussian), its median coincides with its mean, and thus $Z^p(s_0)$ is a median unbiased estimator of $Z(s_0)$. As a second consequence, $Z^p(s_0)$ is more sensitive to the covariance parameters (sill and range) than a linear predictor built on $\mathbf{Z}$ directly. It is also very sensitive to the lognormal assumptions. It is thus not at all obvious that the mathematical optimality of (7.16) should be preferred to the robustness of a conventional kriging on the raw values $\mathbf{Z}$, which is only linearly optimal. A wise advise is probably to perform both kriging and to assess their performance via a cross-validation study.

Note that lognormal processes are also used as driving intensities for non homogeneous point processes, see e.g. [24].

7.2.1.2 Box-Cox Spatial Processes

A more general transformation for positive values Z is the Box-Cox transformation:

$$y = \phi_\lambda^{-1}(z) = \frac{z^\lambda - 1}{\lambda} \text{ if } \lambda \neq 0; \quad \phi_0^{-1}(z) = \ln z. \tag{7.17}$$

Obviously the case $\lambda = 0$ brings us back to the lognormal model above. For $\lambda \neq 0$, derivations similar to those described in the previous paragraph are possible. They involve technicalities that are beyond the scope of this presentation. In [12] this transformation has successfully been applied for the prediction of rainfall using a Bayesian approach.

7.2.2 Transformed Truncated Gaussian Processes

The parametric transform seen in the previous section relies on the assumption that the transform is a one-to-one correspondence. There are unfortunately many cases for which this assumption is not possible. The modeling of rainfall data is a very typical example. Rainfall data are characterized by a large proportion of values equal to 0, up to 90% of the data, depending on the location and on the time step of the measures. A one-to-one transform is thus not appropriate. Models need to be able to account explicitly for a large proportion of 0 values.

The truncated Gaussian model is a particular transformed Gaussian process, with $\phi(y) = \psi(y)I[y \geq v]$, where $I[\cdot]$ is the indicator function equal to 1 if the expression within brackets holds, and 0 otherwise, and where $\psi(y)$ is a strictly monotonic function. In other words, let $\{Y(s),\ s \in \mathscr{D}\}$ be a Gaussian process on $\mathscr{D}$

and define

$$Z(s) = \psi(Y(s)), \quad Y(s) \geq \nu, \tag{7.18}$$

and $Z(s) = 0$ otherwise. The process $Z(\cdot)$ is thus a non strictly monotonic transform of the latent Gaussian process $Y(\cdot)$. This model has been used in [4] for modeling spatio-temporal rainfall data with a quadratic form of the power transform of the data: $\psi(y) = \alpha_0 + \alpha_1 y^{\gamma} + \alpha_2 y^{2\gamma}$, for $y > 0$, which is monotonic for a large range of values of y when $(\alpha_0, \alpha_1, \alpha_2, \gamma)$ are estimated from the data, thus guaranteeing strict monotonicity on a large interval of y values. The data were spatial and temporal. A Gaussian Random Markov Field (GMRF) approach was used for modeling the latent Gaussian process, and estimation of the parameters was done by maximizing the log of the pairwise probabilities.

We shall take a different route. First, we shall consider an exponential transformation of the truncated Gaussian field, instead of a power-quadratic one, i.e. $\psi(y) = z_m + b(e^{a[Y-\nu]^c} - 1)$. It was found to offer an excellent fit to the data (see Fig. 7.4), and it offers the possibility of deriving explicit formulas in the case $c = 1$. We shall denote with $f(y)$ the probability density function (pdf) of a $(0, 1)$ Gaussian random variable and $F(y)$ its cumulative probability function (cpf). We shall further denote the complementary cpf $G(y) = 1 - F(y)$. The model is thus the following

$$Z(\cdot) = \begin{cases} 0, & Y(\cdot) < \nu \\ z_m + b(e^{a[Y(\cdot)-\nu]^c} - 1), & Y(\cdot) \geq \nu, \end{cases} \tag{7.19}$$

were z_m is the resolution of the rain gauge (i.e. it is the minimum quantity a rain gauge is capable to measure; typically $z_m = 0.5$ mm), ν is a parameter related to the rain frequency and a, b and c are parameters related to the rain intensity regime. When $c = 1$ explicit expressions can be derived for the expectation and the variance of rainfall, given that the day is not dry:

$$\begin{aligned} E[Z \mid Z > 0] &= E[z_m + b(e^{a[Y-\nu]} - 1) \mid Y \geq \nu] \\ &= z_m + b\left(e^{-a\nu} E[e^{aY} \mid Y \geq \nu] - 1\right) \\ &= z_m + b\left(e^{a^2/2 - a\nu} \frac{G(\nu - a)}{G(\nu)} - 1\right), \end{aligned} \tag{7.20}$$

and

$$\begin{aligned} \mathrm{Var}(Z \mid Z > 0) &= b^2 \mathrm{Var}(e^{a[Y-\nu]} \mid Y \geq \nu) \\ &= b^2 e^{a^2 - 2a\nu} \left[e^{a^2} \frac{G(\nu - 2a)}{G(\nu)} - \left(\frac{G(\nu - a)}{G(\nu)} \right)^2 \right]. \end{aligned} \tag{7.21}$$

From (7.20) and (7.21) a method of moments estimator for a and b can be derived. One can derive the variogram of $Y(\cdot)$, conditional on $Y(\cdot) \geq \nu$. First note that $E[Y \mid Y \geq \nu] = f(\nu)/G(\nu)$. Recall that $\rho(h)$ is the correlation function of the latent $(0, 1)$ Gaussian process. Then, following [31], technical but otherwise

straightforward computations lead to

$$E[Y(s)Y(s+h) \mid Y(s) \geq v, Y(s+h) \geq v] = \rho(h) + \frac{2\rho(h)vf(v)G(v^*_{\rho(h)})}{G_2(v,v;\rho(h))} + \frac{(1-\rho^2(h))f_2(v,v;\rho(h))}{G_2(v,v;\rho(h))} - \left(\frac{f(v)}{G(v)}\right)^2, \quad (7.22)$$

where $v^*_{\rho(h)} = v\sqrt{(1-\rho(h))/(1+\rho(h))}$, and f_2 and G_2 are respectively the pdf and the complementary cumulative function of a $(0,1)$ bivariate Gaussian vector with correlation $\rho(h)$. From (7.22) we can derive the expression of the variogram:

$$\gamma_{Y|Y\geq v}(h) = \gamma(h) + \frac{\gamma(h)^2}{G_2(v,v;\rho(h))}\left\{vf(v)G(v^*_{\rho(h)}) - [2-\gamma(h)]f_2(v,v;\rho(h))\right\}, \quad (7.23)$$

Note that (7.23) is a one-to-one mapping between the variogram of the latent Gaussian process, $\gamma(h)$, and the variogram of the truncated process, $\gamma_{Y|Y\geq v}$. To each definite positive covariance function of $Y(\cdot)$ corresponds a variogram of $Y_{Y|Y>v}$. Conversely, variograms of the truncated process must have the specific form (7.23).

This model has been used for modeling hourly precipitation values on 30 year long climatic series recorded at 12 locations in France, as part of the research project CLIMATOR [9]. The aim was to build a model for disaggregating daily values of large scale climatic models into hourly values used as input values for agricultural precipitation models. The disaggregation was done stochastically according to model (7.19), conditional on the daily value, with a temporal autocorrelation of the Gaussian process modeled as the sum of two exponential covariance functions. A maximum likelihood procedure for estimating their parameters would rely heavily on the multi-Gaussian assumption which cannot be checked (not even mentioning tested). An estimation by a weighted least square procedure was thus preferred, because it was less sensitive to departures from the model. We refer to [34] and [35] for an in-depth discussion about the estimation of variogram parameters. Figure 7.4 illustrates the fitting of the model on hourly data measured in Toulouse, France, from September 1st to November 30th. Since at the Toulouse climatic station, $z_m = 0.50$ mm, data are multiples of z_m. The top panels show the histogram and the cumulative probability function of both the data and the model. The experimental cpf is a step function, while the theoretical cpf is a continuous function. The bottom left panel shows the quantity $Q(t) = E[Z.I(Z > t)]$, where t is a rainfall threshold. For $t = z_m$, $Q(z_m) = E[Z]$. The bottom right panel shows the experimental variogram of $Y_{Z>0} = Y_{Y|Y\geq\hat{v}}$ along with its theoretical curve, as given by (7.23).

7.2.3 *Excursion Sets of Gaussian Random Processes*

Yet a more severe transformation than (7.18) is to consider that $\psi(y)$ is a constant function, say, $\psi(y) = 1$. Then, for a $(0,1)$ Gaussian random process $Y(\cdot)$ in $\mathscr{D}$,

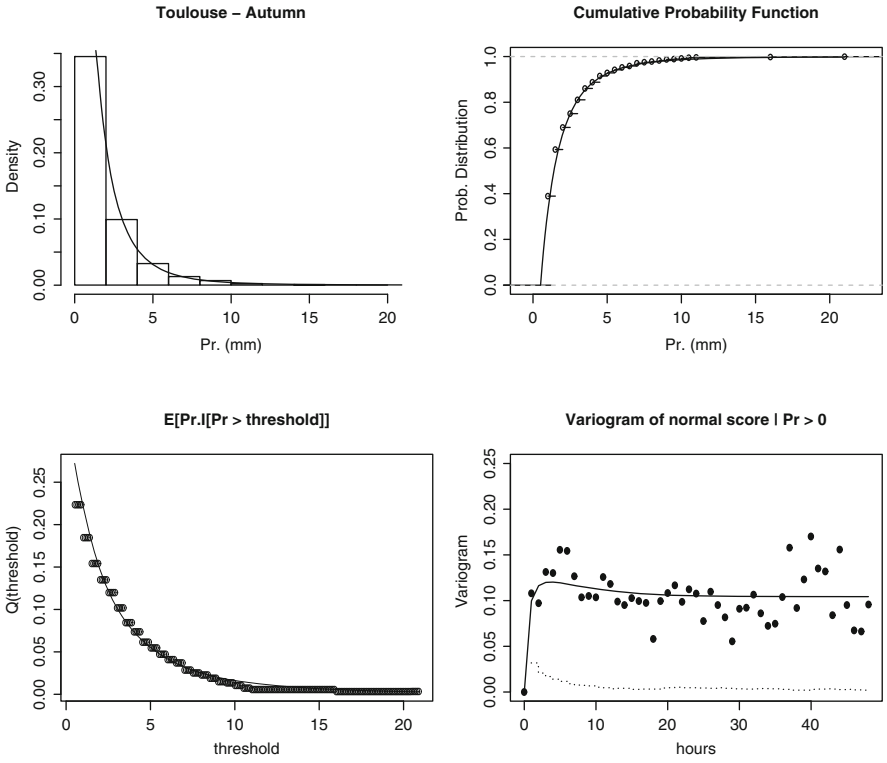

Fig. 7.4 Fitting of the model on Sept-Oct-Nov data measured in Toulouse, France. *Top left*: histogram and fitted pdf. *Top right*: experimental and fitted cpf. *Bottom left*: experimental and fitted quantity $Q(t)$. Bottom right experimental and fitted variogram. See text for details

with correlation function $\rho(h)$, the model (7.18) now becomes

$$Z_\nu(s) = 1, \;\; Y(s) \geq \nu, \;\; s \in \mathscr{D}, \tag{7.24}$$

and $Z_\nu(s) = 0$ otherwise. An alternative point of view is to consider the set

$$X_\nu = \{s \in \mathscr{D} : Y(s) \geq \nu\}. \tag{7.25}$$

The set X_ν is called the *excursion set* of the Gaussian random process $Y(\cdot)$. There is an obvious duality between excursion set and truncated Gaussian random process, since

$$Z_\nu(s) = 1 \Leftrightarrow s \in X_\nu. \tag{7.26}$$

In the usual geostatistical literature, (7.24) is referred to as the *truncated Gaussian random process*. It has been extensively used for modeling geological lithofacies for simulating petroleum reservoirs [10]. In [25] an interesting

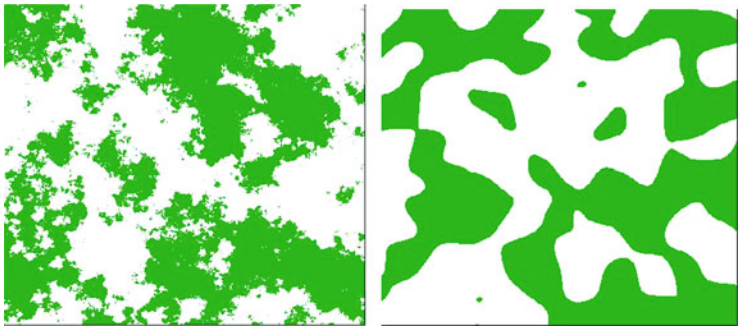

Fig. 7.5 Excursion sets of the realizations in Fig. 7.1. *Left*: with exponential covariance. *Right*: with Gaussian covariance. See text for details

connection with point processes is made. This work considers spatial Cox point processes where the random intensity is defined by the excursion set of a Gaussian random process, such that different point intensities appear in the two phases formed by the random set and its complement.

Figure 7.5 shows the excursion set at $\nu = 0$ of the realizations shown in Fig. 7.1. One can observe that their boundaries present very different roughness. The following theorem provides the relationship between the regularity of the Gaussian random process and the properties of the boundaries of the excursion set. Its proof can be found in [19]; a sketch is presented here for sake of completeness.

Theorem 7.2. *The variogram $\gamma_\nu(h)$ of $Z_\nu(\cdot)$ given in (7.24) is*

$$\gamma_\nu(h) = \frac{1}{2\pi} \int_{\rho(h)}^{1} \frac{1}{\sqrt{1-r^2}} e^{-\nu^2/(1+r)} dr.$$

Proof. First note that

$$\gamma_\nu(h) = P(Y(s) < \nu, Y(s+h) \geq \nu) = \int_{-\infty}^{\nu} \int_{\nu}^{\infty} g_{\rho(h)}(u,v) du\, dv. \tag{7.27}$$

Recall that the bivariate Gaussian pdf is

$$g_\rho(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{0.5(u^2 - 2\rho uv + v^2)}{1-\rho^2} \right\}.$$

Then, by direct computation one can check that $\partial g_\rho(u,v)/\partial\rho = \partial^2 g_\rho(u,v)/\partial u \partial v$, from which, using (7.27) we find

$$\frac{\partial \gamma_\nu}{\partial \rho}(h) = \frac{\partial}{\partial \rho} \int_{-\infty}^{\nu} \int_{\nu}^{\infty} g_{\rho(h)}(u,v) du\, dv = \int_{-\infty}^{\nu} \int_{\nu}^{\infty} \frac{\partial^2 g_\rho(u,v)}{\partial u \partial v} du\, dv = -g_\rho(\nu,\nu).$$

Hence

$$\gamma_\nu(h) = \int_{\rho(h)}^{1} g_r(\nu,\nu)dr = \frac{1}{2\pi}\int_{\rho(h)}^{1} \frac{1}{\sqrt{1-r^2}} e^{-\nu^2/(1+r)}dr.$$

□

From this result is deduced the regularity of the excursion set. We perform the change of variable $r = \cos(2t)$ in Theorem 7.2. Since $\gamma(h) = 1 - \rho(h)$:

$$\gamma_\nu(h) = \frac{1}{\pi}\int_0^{\arcsin\sqrt{\gamma(h)/2}} \exp\left\{-\frac{\nu^2}{2}(1+\tan^2 t)\right\} dt \tag{7.28}$$

Now, for $h \approx 0$, i.e., for $\gamma(h) \approx 0$

$$\gamma_\nu(h) \approx \frac{1}{\pi\sqrt{2}}\sqrt{\gamma(h)}e^{-\nu^2/2}. \tag{7.29}$$

Hence, from Equation (7.29),

- If $\gamma(h) \propto ||h||^2$ at the origin, which is for example the case for a Gaussian covariance function, X_ν has finite specific perimeter. In other words, Gaussian random processes that are at least mean square differentiable lead to excursion sets with finite specific perimeter. They will display very regular boundaries, as in the right panel of Fig. 7.5.
- If $\gamma(h) \propto ||h||^\alpha$, $0 < \alpha < 2$ near $h = 0$, which is for example the case for an exponential covariance function, X_ν has an infinite specific perimeter. Gaussian random processes that are not mean square differentiable lead to excursion sets with infinite specific perimeter. They display very erratic boundaries, as illustrated on the left panel of Fig. 7.5.

It must be noted that the expression in Theorem 7.2 is again a one-to one mapping between the class of valid variograms for Gaussian random processes to the class of valid variograms for excursion sets. This mapping is not a one-to-one correspondence, because covariance functions of random sets must verify more conditions than positive definiteness, as we shall see below. The question of establishing from (7.28) which variograms are associated with excursion sets of Gaussian random process is a difficult one. Let $\gamma_0(h)$ be the variogram of a random process taking only two values, 0 and 1. According to (7.29), $\gamma_0(h)$ is the variogram of an ($\nu = 0$) excursion set of Gaussian random process if the function $\gamma(h) = 2\sin^2\{\pi\gamma_0(h)\}$ is a variogram in $\mathbf{R}^d$. [19] proves that it is the case for the exponential family, i.e. for $\gamma_0(h) = e^{-a||h||}$.

7.2.4 Variograms of Random Sets

Excursion sets presented in the previous paragraph are one specific example of *randoms sets*. We recall briefly some important definitions and properties, but the

reader is referred to [21] and [19] for more detailed presentations. A random set is a stochastic process whose realizations are subsets of $\mathbf{R}^d$. Typical examples are sets obtained by thresholding a random process $Y(\cdot)$ at a given level ν: when $Y(\cdot)$ is a Gaussian random process, the associated random sets are the excursion sets defined in Sect. 7.2.3. When $Y(\cdot)$ is a Poisson line tessellation with Gaussian values, the associated random sets are union of disjoint Poisson polygons, as defined in Sect. 7.1.2.

Random closed sets are characterized by the hitting functional

$$T(K) = P(X \cap K \neq \emptyset), \tag{7.30}$$

where K is any compact set of $\mathbf{R}^d$ [21]. This characterization allows us to consider a wide variety of random closed sets, including spatial point processes. In the rest of this section, we shall only consider random regular sets, i.e., closed sets with non void interior. Second order characteristics of random regular sets can be derived from (7.30).

- Setting $K = \{s\}$ leads to the first order characteristic

$$p(s) = T(\{s\}) = P(s \in X).$$

- Setting $K = \{s, s + h\}$ leads to the second order characteristic

$$c(s, s + h) = T(\{s, s + h\}) = P(s \in X, s + h \in X).$$

Assuming stationarity yields

$$p(s) = p \text{ and } c(s, s + h) = c(h). \tag{7.31}$$

The quantity p is called the proportion of the random set and $c(h)$ is its non centered covariance function. Some immediate properties of $c(h)$ are $c(0) = p$ and $\lim_{||h||\to\infty} c(h) = p^2$. Contrarily to Gaussian random processes, first and second order characteristics of a random sets are thus tightly related.

As already seen in (7.26), any random set X can be characterized by its indicator function: $I_X(s) = 1$, if $x \in X$, and $I_X(s) = 0$ otherwise. The variogram of $I_X(\cdot)$ is

$$\begin{aligned}
\gamma_X(h) &= 0.5E[(I_X(s) - I_X(s + h))^2] \\
&= 0.5P(I_X(s) \neq I_X(s + h)) \\
&= P(s \in X) - P(s \in X, s + h \in X) \qquad (7.32) \\
&= p - c(h). \qquad (7.33)
\end{aligned}$$

The third equality uses the duality between indicator functions and random sets. Variograms of random sets verify $\gamma_X(0) = 0$ and $\lim_{||h||\to\infty} \gamma_X(h) = p - p^2 = p(1-p)$. Notice that variograms of indicators of random sets are necessarily bounded.

From (7.32) one gets $\lim_{||h||\to 0} \gamma_X(h) = \lim_{||h||\to 0} P(s \in X) - P(s \in X, s+h \in X)$. Hence, the behavior of the variogram at the origin conveys information about the boundary of X. Let us define $\sigma^{(d)}$ the specific $(d-1)$ volume of X; $\sigma^{(2)}$ is for example the specific perimeter of the random set X in $\mathbf{R}^2$. The next theorem states precisely the relationship between $\sigma^{(d)}$ and $\gamma_X(h)$.

Theorem 7.3 ([21]). *Let X be an isotropic random regular set in $\mathbf{R}^d$, and let $\gamma_X(h)$ denote its associated variogram. Then,*

$$\sigma^{(d)} = \frac{d\omega_d}{\omega_{d-1}} \gamma'_X(0),$$

where ω_d denotes the d-volume of the unit ball in $\mathbf{R}^d$. For example, for $d = 2$, one gets $\omega_d = \pi$, $\omega_{d-1} = 2$ and $\sigma^{(2)} = \pi\gamma'_X(0)$.

Thus,

- If $\gamma_x(h)$ has a linear behavior at the origin, $\sigma^{(d)}$ is finite and the boundary is regular.
- If $\gamma'_x(0)$ is infinite, the boundary of X has a non integer dimension D with $d-1 < D < d$.
- The case $\gamma'_x(0) = 0$ would correspond to a degenerate case, where $X = \mathbf{R}^d$ with probability p and $X = \emptyset$ with probability $1-p$.

If X is the excursion set of a Gaussian random process $Y(\cdot)$, the first case corresponds to mean square differentiable random processes while the second case corresponds to non mean square differentiable random processes.

There is yet another condition a variogram of a random set must verify. Since

$$\{I_X(s) - I_X(s+h)\}^2 = |I_X(s) - I_X(s+h)|$$

and using

$$|I_X(s) - I_X(s+h+h')| \le |I_X(s) - I_X(s+h)| + |I_X(s+h) - I_X(s+h+h')|,$$

it is clear that the variogram $\gamma_x(h)$ must satisfy the triangular inequality

$$\gamma_X(h+h') \le \gamma_X(h) + \gamma_X(h'). \tag{7.34}$$

Consider for example $\gamma(h) \approx ||h||^\alpha$ when $||h|| \approx 0$. Then choosing $h = h'$ in (7.34) yields $(2||h||)^\alpha \le 2||h||^\alpha$, hence $\alpha \le 1$. This again shows that regular variograms such as Gaussian variograms or Matérn variograms with $\kappa > 1/2$ cannot be variograms of random sets.

Clearly, from (7.34) and the other considerations seen so far, not all functions $\gamma_X(h) = C(0) - C(h)$, where $C(h)$ is a covariance function, can be the variogram of a random set. We have seen that $\gamma_X(h)$ must be bounded and that $\gamma_X(h)$ cannot be too regular. The natural question is therefore "Is there a general characterization of variograms of random sets ?" There is unfortunately no definitive answer to that question, which is still an open problem. The triangular inequality (7.34) is a necessary condition, but it is not sufficient. In [23] a conjecture is proposed to characterize the family of indicator variograms.

Conjecture 7.1 ([23]). Let $\gamma(h)$ be a variogram. It is the variogram of an indicator function (i.e., it is the variogram associated to a random set) if the inequality

$$\sum_{i,j=1}^{n} \epsilon_i \epsilon_j \gamma(s_i - s_j) \leq 0 \tag{7.35}$$

for all n, for all $(s_1, \ldots, s_n) \in \mathbf{R}^d$, and for all sequence of values $(\epsilon_1, \ldots, \epsilon_n)$ taking values in $\{-1, 0, 1\}$ such that $\sum_{i=1}^{n} \epsilon_i = 1$.

Matheron was able to prove that (7.35) is a necessary condition, but he could not prove that it is sufficient.

Even though a general characterization of the covariance function for random sets is still an open question, some partial results do exist. In [13] it is shown that the spherical variogram cannot be the variogram of a stationary mosaic random field in $\mathbf{R}^3$ and of a stationary Boolean random set in $\mathbf{R}^3$. It also cannot be the variogram of the excursion set of a stationary Gaussian random field in $\mathbf{R}^d$, $d \leq 3$. Similar conclusions hold for the circular variogram in $\mathbf{R}^2$. The variogram of a stationary mosaic random field in $\mathbf{R}$ can be spherical, circular or triangular. These results suggest that the spherical variogram is not a good candidate for random sets in general. Among the variograms with linear behaviour at the origin, we have seen that the exponential function is a valid covariance function for random sets. It is in particular the covariance function associated with the Poisson line tessellation (Sect. 7.1.2) and the covariance function of an excursion set of a particular Gaussian random process (Sect. 7.2.3).

7.3 Skew-Normal Spatial Random Processes

In the previous section, we have seen how to define transformed Gaussian random processes. In this section, we present a different approach, which consists in working with a larger class of distributions than the Gaussian one, namely the complete skew-normal distribution (CSN). The family of the CSN distributions is an extension of the Gaussian distribution which admits skewness while at the same time retaining most of the tractability offered by the Gaussian distribution: it is closed under summation, marginalization and conditioning ([5], [6], [16] and [15]).

In [18] a Bayesian approach is proposed for spatial prediction of rainfall using skew-normal processes. Their model was unfortunately based on the usual skew-normal distributions, which has been shown to be unable to impose a strong amount of skewness for spatial data [26]. CSN distributions have been used for Bayesian spatial regression which can handle covariates and missing observation [17]. Spatial prediction is made by a Metropolis-Hasting algorithm. The Bayesian approach is of course very appealing because it can handle hierarchical models and missing data. It is however time consuming and can be difficult to implement. In [3] a spatial model based on the complete skew-normal CSN distribution is proposed. It will be shown below that for this model, it is possible to perform a method of moment approach for estimating the parameters of the model, and in particular that it is possible to estimate the covariance function of the spatial random process.

7.3.1 The Multivariate Closed Skew-Normal Distribution

An n-dimensional random vector $\mathbf{Y}$ is said to have a multivariate closed skew-normal distribution, denoted by $\mathrm{CSN}_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$, if its density function is of the form

$$f(\mathbf{y}) = c_m \, \phi_n(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \Phi_m(\mathbf{D}'(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \boldsymbol{\Delta}), \text{ with } c_m^{-1} = \Phi_m(\mathbf{0}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D}), \tag{7.36}$$

where $\boldsymbol{\mu} \in \mathbf{R}^n$ and $\boldsymbol{\nu} \in \mathbf{R}^m$ are vectors of expectation, $\boldsymbol{\Sigma} \in \mathbf{R}^{n \times n}$ and $\boldsymbol{\Delta} \in \mathbf{R}^{m \times m}$ are covariance matrices, $\mathbf{D} \in \mathbf{R}^{n \times m}$ is a transformation matrix and $\phi_n(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_n(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ are the pdf and cdf, respectively, of the n-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathbf{D}'$ is the transpose of the matrix $\mathbf{D}$. If $\mathbf{D} = \mathbf{0}$, the density (7.36) reduces to the multivariate normal one.

The $\mathrm{CSN}_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$ distribution defined by (7.36) can be generated in the following way. Let $\mathbf{U}$ be a Gaussian vector of dimension m and consider the augmented Gaussian vector $(\mathbf{U}', \mathbf{Z}')'$ such that

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{Z} \end{pmatrix} \stackrel{\mathrm{d}}{=} N_{m+n}\left(\begin{pmatrix} \boldsymbol{\nu} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta} + \mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D} & -\mathbf{D}'\boldsymbol{\Sigma} \\ -\boldsymbol{\Sigma}\boldsymbol{D} & \boldsymbol{\Sigma} \end{pmatrix} \right), \tag{7.37}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. Then it is straightforward to show that, conditional on $\mathbf{U} \leq \mathbf{0}$, the random vector

$$\mathbf{Y} = \boldsymbol{\mu} + [\mathbf{Z} | \mathbf{U} \leq \mathbf{0}] \tag{7.38}$$

is distributed according to the $\mathrm{CSN}_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$ distribution defined in (7.36). Here the notation $\mathbf{U} \leq \mathbf{0}$ corresponds to $U_i \leq 0$, for all $i = 1, \ldots, m$. This construction offers a wide range of possible models depending on the choice of $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Sigma}$ and $\mathbf{D}$. The moment generating function (mgf) of $\mathbf{Y}$ is

$$M(\mathbf{t}) = c_m \Phi_m(\mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{t}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D}) \exp\left\{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\boldsymbol{t}\right\}. \tag{7.39}$$

It is the product of the mgf of a Gaussian vector, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and the cdf of the m dimensional normal distribution with mean $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Delta} + \mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D}$. From (7.39), it is easy to derive the moments of $\mathbf{Y}$, but these expressions are difficult to use in practice because the cumulative distribution Φ_m cannot be represented in analytical form. This fact raises an important computational hurdle if m is large. In a spatial context, it implies that only simple structures for $\boldsymbol{\Delta}$ and $\mathbf{D}$ can be investigated. The next section clarifies the forms of such matrices. The parameter $\boldsymbol{\nu}$ in (7.36) is difficult to estimate because it is a redundant quantity in the expression $\Phi_m(\mathbf{D}'(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \boldsymbol{\Delta})$. It is thus decided to set $\boldsymbol{\nu} = \mathbf{0}$ in the rest of this chapter.

7.3.2 Skew-Normal Spatial Processes

In order to define a skew-normal spatial random process, it is sufficient to define a proper distribution verifying the Kolmogorov consistency conditions for any finite vector $\mathbf{Y}$. Similarily to (7.38) we define a CSN random process $\{Y(s)\}$, $s \in \mathscr{D}$ as

$$Y(s) \overset{\mathrm{d}}{=} \mu(s) + [Z(s) \mid \mathbf{U} \leq \mathbf{0}].$$

Thus, for any n and any vector $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))'$ of $Z(\cdot)$, $\mathbf{Y} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + [\mathbf{Z} \mid \mathbf{U} \leq \mathbf{0}]$, where $\mathbf{U}$ and $\mathbf{Z}$ are distributed according to (7.37). The vector $\mathbf{Y}$ can also be expressed as a sum of two independent processes. Let us introduce

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \overset{\mathrm{d}}{=} N_{m+n}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta} + \mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n \end{pmatrix}\right),$$

where $\mathbf{I}_n$ represents the identity matrix of size n. Then the vector $\mathbf{Z}$ is set equal to $-\mathbf{F}\mathbf{U} + \mathbf{G}^{1/2}\mathbf{V}$ with $\mathbf{F} = \boldsymbol{\Sigma}\boldsymbol{D}(\boldsymbol{\Delta} + \mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D})^{-1}$ and $\mathbf{G} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{D}(\boldsymbol{\Delta} + \mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D})^{-1}\mathbf{D}'\boldsymbol{\Sigma}$. This is a multivariate Gaussian vector with zero-mean vector and covariance $\boldsymbol{\Sigma}$ and, more importantly, the bivariate couple $(\mathbf{U}', \mathbf{Z}')'$ satisfies (7.37). The independence of $\mathbf{U}$ and $\mathbf{V}$ allows us to write

$$\mathbf{Y} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{F}[\mathbf{U} \mid \mathbf{U} \geq \mathbf{0}] + \mathbf{G}^{1/2}\mathbf{V}. \tag{7.40}$$

This decomposition is useful when deriving moments and empirical variogram properties. The spatial skew-normal random process shall differ according to the values of the parameters. In [26] and [3], several possibilities are explored. In particular, the integer m is assumed to be a known quantity, not an extra parameter to be estimated. Moreover, it is shown that m must be large enough in order to impose sufficient skewness. As a consequence m is set to be equal to the number of data:

$m = n$. In the model proposed in [3], it is further imposed that $\boldsymbol{\Delta} = \boldsymbol{\Sigma}$ and that $\boldsymbol{\Sigma}$ corresponds to a covariance function $C(h)$ computed at all pairs of locations s_i, s_j, with $1 \leq i, j \leq n$. Such a construction corresponds to equating the hidden points which define implicitly the elements of $\mathbf{U}$ with the observed ones. This model is thus referred to as the *homotopic model*. To simplify the interpretation of the model, it is further assumed that $\mathbf{D} = \delta \mathbf{I}_n$, where $\delta \in \mathbf{R}$ is a single parameter controlling the skewness. When $\delta = 0$, $\mathbf{Y}$ is independent of $\mathbf{U}$, i.e. $\mathbf{Y}$ is a Gaussian vector with expectation $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. As δ^2 increases, the vector $\mathbf{Y}$ tends to a proper truncated Gaussian vector. With these choices (7.40) leads to

$$\mathbf{Y} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \frac{\delta}{1+\delta^2}[\mathbf{U} \mid \mathbf{U} \geq \mathbf{0}] + \frac{1}{\sqrt{1+\delta^2}} \boldsymbol{\Sigma}^{1/2} \mathbf{V}. \tag{7.41}$$

The vector $\mathbf{Y}$ is thus a weighted sum of a sample from a stationary process with covariance function proportional to $C(h)$ and a vector $\mathbf{U}$ conditional on $\mathbf{U} \geq \mathbf{0}$, whose covariance function is $(1 + \delta^2)C(h)$. For the sake of stationarity, we shall further impose that $\boldsymbol{\mu} = \mu\ (1, \ldots, 1)'$.

For computing the moments of $\mathbf{Y}$, we shall take advantage of the decomposition in (7.41) and use results concerning the moments of the truncated multivariate normal distribution given in [31]. Let us denote $\mathbf{R}$ the correlation matrix of $\mathbf{U}$, $R_{ij} = \mathbf{R}_{[i,j]}$ its elements and $\mathbf{R}_k$ its kth partial correlation matrix. Then, as shown in [3],

$$m_i = E[Y_i] - \mu = \frac{\delta^* \sigma}{\sqrt{2\pi}} \sum_{k=1}^{n} R_{ik} \frac{\Phi_{n-1}(\mathbf{0}; \mathbf{0}, \mathbf{R}_k)}{\Phi_n(\mathbf{0}; \mathbf{0}, \mathbf{R})}, \tag{7.42}$$

with $\delta^* = \delta(1 + \delta^2)^{-1/2}$. The parameter driving the skewness is δ^*, a quantity varying between -1 and 1. When $|\delta^*| = 1$, the skewness is maximum. It should be noted that although $\mathbf{U}$ and $\mathbf{V}$ are stationary processes, $\mathbf{Y}$ is not stationary.

As second moment, we shall use the experimental variogram, denoted $\hat{\gamma}_Y(h)$. It is established in [3] that

$$E[\hat{\gamma}_Y(h)] = \left[\frac{1}{N(h)} \sum_{\{(i,j):|x_i - x_j| \approx ||h||\}} \gamma_{ij} \right] + \frac{\delta^{*2} \sigma^2}{2\pi} \Gamma(h), \tag{7.43}$$

with

$$\Gamma(h) = \frac{1}{2N(h)} \sum_{\{(i,j):|x_i - x_j| \approx ||h||\}} (\Psi_{ii} + \Psi_{jj} - 2\Psi_{ij}) \tag{7.44}$$

and $\gamma_{ij} = \sigma^2(1 - R_{ij})$. In (7.44), the Ψ_{ij}s are quantities related to the second order moments of $\mathbf{U}$, namely

$$\Psi_{ij} = \sum_{k=1}^{n} R_{ik} \sum_{l \neq k} \frac{R_{jl} - R_{kl} R_{jk}}{\sqrt{1 - R_{kl}^2}} \frac{\Phi_{n-2}(\mathbf{0}; \mathbf{0}, \mathbf{R}_{kl})}{\Phi_n(\mathbf{0}; \mathbf{0}, \mathbf{R})}, \tag{7.45}$$

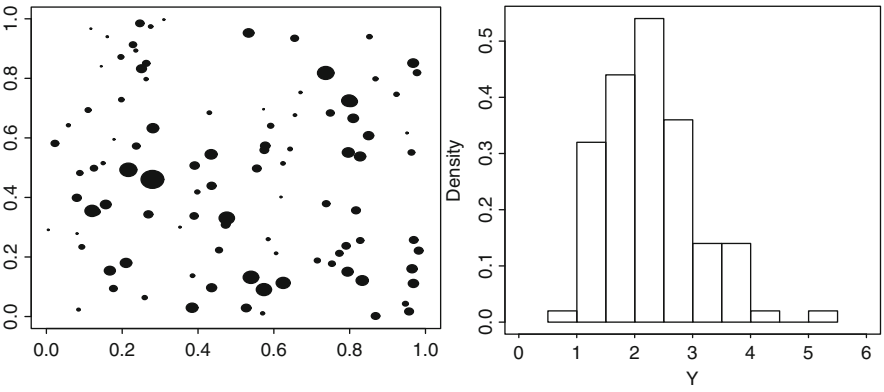

Fig. 7.6 CSN spatial process with $\mu = 1$, $\delta = 4$ and exponential covariance function $C(h) = \exp\{-10||h||\}$. Simulation at $n = 100$ locations and histogram (*right*)

with $\mathbf{R}_{kl}$ being the partial correlation matrix satisfying

$$\phi_n(x_s, x_k = u_k, x_l = u_l; \mathbf{R}) = \phi_2(u_k, u_r; R_{kl}) \;\; \phi_{n-2}\left(\frac{x_s - x_s^{CE}}{\sigma_s^{CE}}; \mathbf{R}_{kl}\right),$$

where x_s^{CE} and σ_s^{CE} are, respectively, the conditional expectation and conditional standard deviation of x_s given $(x_k, x_l) = (u_k, u_l)$. Note that in general $\Psi_{ij} \geq 0.5(\Psi_{ii} + \Psi_{jj})$ and thus that the quantity (7.44) is negative.

Based on these moments, it is thus possible to estimate the parameters of the model in (7.41). The method is described in details in [3]. It consists in: (1) estimating the parameters of the variogram for a given parametric family (e.g. Matérn class), based on the experimental variogram computed on the data; (2) computing the corresponding correlation matrix $\mathbf{R}$, the Ψ_{ij} and $\Gamma(h)$ using (7.44) and (7.45); (3) estimate δ^{*2} and σ^2 from (7.43); (4) estimate μ from (7.42).

This model and the estimation of its parameters are now illustrated. The left panel of Fig. 7.6 shows one realization at $n = 100$ sample locations within the unit square of the spatial homotopic model with $(\mu, \sigma^2, \delta^*) = (1, 1, 0.94)$ and an exponential covariance function $C(h) = \exp\{-10||h||\}$. The histogram, displayed on the right panel of Fig. 7.6 clearly indicates that the model is capable of generating a fair degree of skewness in the values taken by the process. As one would expect for a skewness parameter $\delta^* = 0.94$, the histogram is strongly positively skewed. The sample variogram is displayed in Fig. 7.7, Note that the sample variogram indicates a spatial structure up to a distance of around 0.1, in agreement with the simulated covariance function. However, the variance values corresponding to the sill of the variogram is around 0.7, considerably smaller than the value of 1 taken by σ, in agreement with (7.42). The estimated parameters are $(\hat{\mu}, \hat{\sigma}^2, \hat{\delta}^*, \hat{a}) = (1.30, 1.15, 0.92, 0.062)$. The resulting variogram is displayed in Fig. 7.7. The fit with the sample variogram is clear.

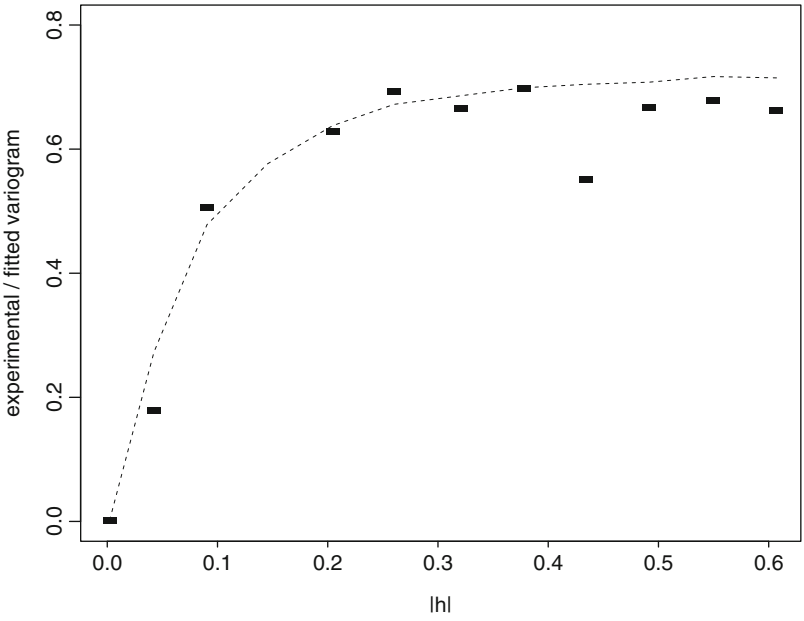

Fig. 7.7 Sample variogram and fitted variogram of the realization displayed Fig. 7.6

Acknowledgements The author wishes to acknowledge Frédéric Baret, Sébastien Garrigues and Philippe Naveau, co-authors of some of the papers cited here. The research presented in Sect. 7.2.2 was funded by the ANR CLIMATOR project.

References

1. Adler, R.: The Geometry of Random Fields. Wiley, New York (1981)
2. Adler, R., Taylor, J.: Random Fields and Geometry. Springer, Boston (2007)
3. Allard D., Naveau, P.: A new spatial skew-normal random field model. Commun. Stat. A-Theor. **36**, 1821–1834 (2007)
4. Allcroft, D., Glasbey, C.: A Latent Gaussian Markov Random-Field Model for Spatiotemporal Rainfall Disaggregation. J. Roy. Stat. Soc. C **55**, 1952–2005 (2003)
5. Azzalini A., Dalla-Valle, A.: The multivariate skew-normal distribution. Biometrika **83**, 715–726 (1996)
6. Azzalini, A.: The skew-normal distribution and related multivariate families. Scand. J. Stat. **32**, 159–188 (2005)
7. Billingsley, P.: Probability and Measure, 2nd edn. John Wiley & Sons, New York (1986)
8. Bochner, S.: Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse. Math. Ann. **108**, 378–410 (1933)
9. Brisson, N., Levrault, F. (Eds.): Changement climatique, agriculture et forêt en France: simulations d'impacts sur les principales espèces. Le Livre Vert du projet CLIMATOR (2007–2010). ADEME, Orléans (2010)

10. Chilès, J.P. and Delfiner, P.: Geostatistics: modeling spatial uncertainty. John Wiley & Sons, New York (1999)
11. Cressie, N.A.C.: Statistics for Spatial Data. John Wiley & Sons, New York (1993)
12. De Oliveira, V., Kedem, B., Short, D.A.: Bayesian prediction of transformed gaussian random fields. J. Am. Stat. Assoc. **92**, 1422–1433 (1997)
13. Emery, X.: On the Existence of Mosaic and Indicator Random Fields with Spherical, Circular and Triangular Variograms. Math. Geosci. **42**, 969–984 (2010)
14. Garrigues, S., Allard, D., Baret, F.: Using first and second order variograms for characterizing landscape spatial structures from remote sensing imagery. IEEE T. Geosci. Remote. **45**, 1823–1834 (2007)
15. Genton, M.: (Ed.) Skew-Elliptical Distributions and Their Applications. A Journey Beyond Normality. Chapman & Hall/CRC, Boca-Raton (2004)
16. Gupta, A., González-Farías, G., Domínguez-Molina, J.: A multivariate skew-normal distribution. J. Multivariate Anal. **89**, 181-190 (2004)
17. Karimi, O., Mohammadzadeh, M.: Bayesian spatial regression models with closed skew normal correlated errors and missing observations. Stat. Pap. DOI 10.1007/s00362-010-0329-2 (2010)
18. Kim, H.M., Mallick, B.K.: A Bayesian prediction using the skew Gaussian distribution. J. Stat. Plan. Infer. **120**, 85–101 (2004)
19. Lantuéjoul, C.: Geostatistical Simulations. Springer, Berlin (2002)
20. Matheron, G.: Effet proportionnel et lognormalité ou le retour du serpent de mer. Technical Report N-374, Centre de Géostatistique, Fontainebleau, France (1974)
21. Matheron, G.: Random Sets and Integral Geometry. Wiley, New York (1975)
22. Matheron, G.: Suffit-il, pour une covariance, d'être de type positif ? Sciences de la Terre, Série Informatique Géologique, **26**, 51–66 (1987)
23. Matheron, G.: Une conjecture sur la covariance d'un ensemble aléatoire. Cahiers de Géostatistique, Fasc. 3, 107–113. Ecole des Mines de Paris (1993)
24. Møller, J., Syversveen, A., Waagepetersen, R.: Log Gaussian Cox Processes. Scand. J. Stat. **25**, 451-482 (1998)
25. Myllymäki, M. and Penttinen, A.: Bayesian inference for Gaussian excursion set generated Cox processes with set marking. Stat. Comput., **20**, 305–315 (2010)
26. Naveau, P. Allard, D.: Modeling skewness in spatial data analysis without data transformation. In: Leuangthong, O., Deutsch, C., (Eds.) Proceedings of the Seventh International Geostatistics Congress, 929–938. Springer, Dordrecht (2004)
27. Porcu, E., Mateu, J., Christakos, G.: Quasi-arithmetic means of covariance functions with potential applications to space-time data. J. Multivariate Anal. **100**, 1830–1844 (2009)
28. Schlather, M.: Some covariance models based on normal scale mixtures. Bernoulli **16**, 780–797 (2010)
29. Schoenberg, I.J.: Metric spaces and completely montonic functions. Ann. Math. **39**, 811–841 (1938)
30. Stein, M.: Interpolation of Spatial Data: Some Theory for Kriging, New-York, Springer (1999)
31. Tallis, G.M.: The moment generating function of the truncated multi-normal Distribution. J. Roy. Stat. Soc. B **23**, 223–229 (1961)
32. Wackernagel, H.: Multivariate Geostatistics. An Introduction with Applications. 3rd Ed. Springer, Heidelberg (2003)
33. Wolpert, R. and Ickstadt, K.: Poisson/gamma random field models for spatial statistics, Biometrika, **85**, 251–267 (1998)
34. Zimmerman, D.L. and Stein, M.: Classical Geostatistical Methods. In: Gelfand, A.E., Diggle, P.J., Fuentes M., Guttorp, P. (Eds.) Handbook of Spatial Statistics, 29–44. Chapman & Hall/CRC, Boca-Raton (2010)
35. Zimmerman, D.L.: Likelihood-Based Methods. In: Gelfand, A.E., Diggle, P.J., Fuentes M., Guttorp, P. (Eds.) Handbook of Spatial Statistics, 45–56. Chapman & Hall/CRC, Boca-Raton (2010)

Chapter 8
Random Fields Arising in Chaotic Systems: Burgers Equation and Fractal Pseudodifferential Systems

Nikolai N. Leonenko and M. Dolores Ruiz-Medina

Abstract This chapter provides a general overview on the main results derived by the authors in relation to limit theory for the solution of linear and non-linear random evolution equations. Additionally, the local regularity properties of the solution to fractional pseudodifferential equations driven by random innovations are introduced. Specifically, limit results derived for the heat and Burgers equations with linear and quadratic external potentials are described in the first part of this chapter. In the second part, the local quadratic variation properties of the solution to fractional pseudodifferential equations on regular and fractal domains are formulated. The driven process can be a fractional Gaussian random field or Levy noise. The fractal domain case and multifractional pseudodifferential formulations are studied in the context of Gaussian white-noise innovations.

8.1 Introduction

Recent research (see [60]) has provided the evidence that stochastic anomalous diffusion process modeling constitutes an important topic in the analysis of geostatistical, geophysics and financial data (see [1]; [19]; [33]; [50]; [51]; [53]; [118], among others). In this context, *anomalous* means the presence of long-range dependence; self-similarity; non-linearity and multifractality. These features can be represented in terms of fractional or multifractional pseudodifferential operators and non-linear evolution equations with random initial conditions. This chapter

N.N. Leonenko (✉)
Cardiff School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF24 4AG
e-mail: LeonenkoN@cardiff.ac.uk

M.D. Ruiz-Medina
Faculty of Sciences, University of Granada, Campus Fuente Nueva s/n, 18071, Granada
e-mail: mruiz@ugr.es

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_8,

contains a review of the main contributions developed by the authors regarding the generation, probabilistic analysis and limit theory of the above-referred anomalous diffusion processes. New results are also derived to complete the contents of the sections of this chapter in relation to the description of the anomalous features (local singularity) characterizing the families of processes introduced.

Data in many fields of application display scaling, e.g. sample paths of fractional Brownian motion (see [85]), or even multiscaling, e.g. sample paths of multifractional Brownian motion (see [23]). A general formulation of multifractional random field models, extending multifractional Brownian motion, is provided in [99]. The effect of the homogeneous or heterogeneous fractal geometry of the domain is studied in [97] and [100]. A way to characterize the multiscaling/multifractal behaviour of a random process or field is via the non-trivial singularity spectrum of its sample paths ([58]). Multiplicative cascades and iterated function systems allow the generation of multifractals ([40]; [41]; [84]). Brownian motion in multifractal time and Lévy processes on fractal domains also display heterogeneous fractality ([58]; [91]). A very small sample of illustrative examples of application of multifractal analysis includes [31]; [33]; [46]; [86]; [91]. Recently, in the papers by [16] complementary methodologies are provided for generation of multifractal processes, based on the products of geometric Ornstein-Uhlenbeck type processes or geometric birth-death processes.

Non-linear dynamics are usually observed in complex systems. Burgers equation constitutes an important example of nonlinear partial differential equations studied in turbulence (see, for example, [30]; [32]; [41]; [48], and [103]).

Burgers equation with random initial conditions and external potential $\Phi = 0$ has been extensively studied in the last one and a half decade by [93], [2]; [29]; [28]; [45]; [37]; [42]; [67]; [75]; [87]; [26]; [101], and [95], among others. Books of [64], [118] and [21] contain a complete bibliography of the subject and expositions of some of the principal results of the theory of Burgers turbulence.

In particular, [2]; [15]; [29]; [28]; [36]; [42]; [67]; [74] obtained Gaussian and non-Gaussian scenarios for parabolically rescaled solutions of the Burgers equation with external potential $\Phi = 0$, and weakly dependent or strongly dependent random initial conditions. These scenarios are in some sense subordinated to the Gaussian white noise random measure.

For other approaches to Burgers' turbulence problem see [20], [93] and [108] (on asymptotic distributions of averages of solutions of Burgers equation with random initial data), [44], [107] (in the context of long memory shot noises), [26] and [101] (large deviation principle and statistics of shock waves), [87], [88], [115], [116] (on hyperbolic asymptotics), [104] (statistics of shocks and related topics), [106], [109] and their references.

In a sense, the Gaussian and non-Gaussian scenarios for a solution of Burgers equation with random initial condition result from the application of the central limit theorem or non-central limit theorems for non-linear functionals of Gaussian processes and fields. The results by [27]; [38]; [56] and [110] [111] contain a main principle for the study of limiting distributions of non-linear transformations of Gaussian processes with long-range dependence. These distributions are not always

Gaussian, but they are subordinated to Gaussian white noise random measures. For a recent development, see [5], [6]; [39]; [70]; [71]; [72]; [73]; [89]; [95], and references therein.

We consider the one-dimensional Burgers equation of the form

$$\frac{\partial}{\partial t}u + u\frac{\partial}{\partial x}u = \mu\frac{\partial^2}{\partial x^2}u + 2\mu\frac{\partial}{\partial x}\Phi, \tag{8.1}$$

subject to the initial condition

$$u(0,x) = u_0(x), \tag{8.2}$$

where $u = u(t,x)$, $t > 0$, $x \in \mathbf{R}$, is a velocity field, and

$$U = U(x) = \int_{-\infty}^{x} u_0(\xi)d\xi \tag{8.3}$$

is the velocity potential, $\Phi = \Phi(x)$, $x \in \mathbf{R}$, is the external potential, $\Phi_x = \frac{d}{dx}\Phi(x)$, $\mu > 0$, is the viscosity of the media, and the reciprocal $\mathrm{Re} = 1/\mu$ corresponds to the Reynolds number. The nonlinear equation (8.1) can be viewed as a simplified version of the Navier-Stokes equation. But the differences between Burgers' and Navier-Stokes' equations are as interesting as the similarities.

Our interest relies on the cases where the external potential Φ is a linear or quadratic function, and the initial condition

$$u_0(x) = \eta(x), \ x \in \mathbf{R}, \tag{8.4}$$

is a stationary random process with short- or long-range dependence, having sample paths that are integrable over all finite intervals. The initial-value problem (8.1)-(8.4) is known as Burgers turbulence problem.

In relation to linear anomalous diffusion processes, in the second part of this chapter, fractional and multifractional pseudodifferential equations driven by random innovations are formulated, and several results on their mean-quadratic local variation properties and sample-path properties (in the Gaussian case) are reviewed, as well as derived here. Specifically, different special cases of temporal, spatial and spatiotemporal pseudodifferential equations of the form

$$\widetilde{\mathscr{L}}X = \varepsilon$$

are considered, where $\widetilde{\mathscr{L}}$ can be a fractional or multifractional temporal, spatial or spatiotemporal pseudodifferential operator, and ε is a zero-mean, second-order temporal, spatial or spatiotemporal random process, satisfying suitable mean-quadratic local variation conditions. In the spatial case, $\widetilde{\mathscr{L}} = \mathscr{L}$ will be assumed to be elliptic, and in the spatiotemporal case, $\mathscr{L}$ is assumed to be of the form

$$\widetilde{\mathscr{L}} = \frac{\partial}{\partial t} + \mathscr{L},$$

where $\mathscr{L}$ can be a fractional or multifractional elliptic pseudodifferential operator.

The outline of the present chapter is the following:

Random Partial Differential Equations (with random initial conditions)

- Heat equation
- Fractional heat equation and fractional wave equation
- Non-linear equations: Burgers without and with linear and quadratic potentials

Fractional Differential Equations Driven by Random Innovations

- Temporal fractional pseudodifferential models
- Elliptic fractional pseudodifferential models on regular and fractal domains
- Fractional pseudodifferential evolution equations
- Multifractional pseudodifferential equations

8.2 Random Partial Differential Equations

In the present section, we study the limiting random fields obtained from the rescaled version of a spatiotemporal random field $u(t,x),\ t>0,\ x\in\mathbf{R}^n$, arising as solution of linear PDE or fractional PDE depending on various types of random initial conditions in the stationary class.

8.2.1 Random Heat Equation

The heat equation with random initial conditions is a classical subject in both mathematics and physical literature. An introduction of the rigorous probabilistic tools into the subject can be found in [59], and [92]. More recently, several researchers investigated solutions of the heat equation with random initial conditions (see [75], [5], [17], [70], [71], [72], [73]).

Let us consider the random field solution

$$u(t,x),\ t>0,\ x\in\mathbf{R}^n,$$

of the heat equation:

$$\frac{\partial u}{\partial t} = \mu\Delta u, \quad \mu>0, \tag{8.5}$$

subject to the stationary random initial condition with zero mean:

$$u_0(x) = \eta(x) = \int_{\mathbf{R}^n} e^{i\langle\lambda,x\rangle} Z_\eta(d\lambda),\ E\left(Z_\eta(d\lambda)\right)^2 = F_\eta(d\lambda), \tag{8.6}$$

$$E\eta(x)\eta(y) = B_\eta(x-y) = \int_{\mathbf{R}^n} e^{i\langle\lambda,x-y\rangle} F_\eta(d\lambda). \tag{8.7}$$

Then, the mean-square solution is

$$u = u(t,x) = \int_{\mathbf{R}^n} \frac{1}{(4\pi\mu t)^{n/2}} e^{-\frac{|x-y|^2}{4\mu t}} \eta(y)dy =$$

$$= \int_{\mathbf{R}^n} e^{i\langle\lambda,x\rangle - \mu t|\lambda|^2} Z_\eta(d\lambda), \quad t > 0,\ x \in \mathbf{R}^n. \tag{8.8}$$

The following assumptions are made on the class of random initial conditions η, defined as non-linear functionals of a Gaussian process ξ, i.e., $\eta(x) = G(\xi(x)),\ x \in \mathbf{R}^n$. Specifically, ξ, G and η satisfy the following conditions:

(A) The process $\xi(x)$, $x \in \mathbf{R}^n$, is a real, measurable, separable, mean-square continuous, stationary and Gaussian random field with $E\xi(x) = 0$, $E\xi^2(x) = 1$, and covariance function $B(x) = B_\xi(x) = E\left[\xi(0)\xi(x)\right]$, $x \in \mathbf{R}^n$, such that

$$B_\xi(x) = \int_{\mathbf{R}^n} e^{i\lambda x} F_\xi(d\lambda) = \int_{\mathbf{R}^n} e^{i\lambda x} f_\xi(\lambda)d\lambda,$$

that is $f_\xi(\lambda)$ is the spectral density of $\xi(x)$, $x \in \mathbf{R}^n$.
Under the condition **(A)**

$$\xi(x) = \int_{\mathbf{R}} e^{i\lambda x} \sqrt{f_\xi(\lambda)} W(d\lambda),\ x \in \mathbf{R}^n, \tag{8.9}$$

where W is the complex Gaussian white noise random measure.

(B) The function $G(u)$, $u \in \mathbf{R}$, is real, measurable, and such that $EG^2(\xi(0)) < \infty$. The (non-linear) function G admits a series expansion in terms of orthogonal Hermite polynomials given by

$$G(u) = \sum_{k=0}^{\infty} \frac{C_k H_k(u)}{k!}, \quad C_k = \int_{\mathbf{R}} G(u)H_k(u)\varphi(u)du,$$

where

$$H_k(u) = (-1)^k \left[\varphi(u)\right]^{-1} \frac{d^k}{du^k}\varphi(u), \quad k = 0, 1, 2, \ldots,$$

with

$$\varphi(u) = \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{u^2}{2}\right\}, \quad u \in \mathbf{R}.$$

(C) There exists an integer $m \geq 1$ such that $C_0 = 0, C_1 = \cdots = C_{m-1} = 0$, $C_m \neq 0$.

Under conditions (A), (B) and (C), the stationary process $\eta(x) = G(\xi(x)),\ x \in \mathbf{R}^n$, admits the spectral representation in form of L_2-stochastic integral:

$$\eta(x) = G(\xi(x)) = \sum_{k=m}^{\infty} \frac{C_k H_k(\xi(x))}{k!} = \int_{\mathbf{R}^n} e^{i\lambda x} Z_\eta(d\lambda), \tag{8.10}$$

with the associated spectral representation of its covariance function

$$B_\eta(x) = cov(\eta(0), \eta(x)) = \int_{\mathbf{R}^n} e^{i\lambda x} F_\eta(d\lambda),$$

in terms of the bounded and positive spectral measure F_η, which defines the second-order structure of the complex-valued orthogonally scattered random measure $Z_\eta(\cdot)$ such that $E\left|Z_\eta(d\lambda)\right|^2 = F_\eta(d\lambda)$.

(D) A weak-dependent scenario is assumed for the random initial condition η, that is,

$$\int_{\mathbf{R}^n} |B_\eta(x)|^m dx < \infty, \quad \int_{\mathbf{R}^n} B_\eta(x) dx \neq 0.$$

Theorem 8.1. *([76]). Suppose that the condition* ***(D)*** *of the short range dependence (SRD) is satisfied by η in (8.10). Then, the finite-dimensional distributions of the random fields*

$$U_\varepsilon(t,x) = \frac{1}{\varepsilon^{1/4}} u\left(\frac{t}{\varepsilon}, \frac{x}{\varepsilon^{1/2}}\right)$$

converge weakly ($\xrightarrow{D}$), as $\varepsilon \to 0$, to the finite-dimensional distributions of the Gaussian random field $U(t,x)$ with $EU(t,x) = 0$ and the following covariance function:

$$EU(t,x)U(t',x') = \left[\int_{\mathbf{R}^n} \sum_{k=m}^{\infty} \frac{C_k^2}{k!} B_\xi^k(x) dx\right] \exp\left\{-\frac{\|x-x'\|^2}{4\mu(t+t')}\right\} \frac{1}{\sqrt{4\pi\mu(t+t')}}.$$

Remark 8.1. Up to the constant the spectral density of the limiting field has the form:

$$g(\lambda) = const \times \frac{1}{\sqrt{4\mu(t+t')}} e^{-\|\lambda\|^2 \times 4\mu(t+t')}, \ \lambda \in \mathbf{R}^n.$$

For a long-range dependent (LRD) random initial condition we assume conditions **(A)–(C)**, and the following condition:

(E) Random field ξ defining $\eta(x) = G(\xi(x)),\ x \in \mathbf{R}^n$, satisfies

$$E\xi(x)\xi(y) = B_\xi(x-y) = \frac{L(\|x-y\|)}{\|x-y\|^{\varkappa}}, \quad 0 < \varkappa < n, \quad \|x-y\| \to \infty, \tag{8.11}$$

where L is a slowly varying function at ∞, bounded at each bounded interval.

By the Tauberian theorem (see [64]), condition **(E)** implies that, as $\|\lambda\| \to 0$,

$$F_\xi(d\lambda) = f_\xi(\lambda)d\lambda, \quad f_\xi(\lambda) \sim \frac{C_T}{\|\lambda\|^{n-\varkappa}}, \quad |\lambda| \to 0, \quad C_T = C_T(n,\varkappa)$$
$$= \frac{\Gamma(\frac{n-\varkappa}{2})}{2^\varkappa \pi^{\frac{n}{2}} \Gamma\left(\frac{\varkappa}{2}\right)}. \tag{8.12}$$

Theorem 8.2. *([75], [5]). Suppose that the condition **(E)** on the LRD is satisfied for η in (8.10). Then the random fields*

$$U_\varepsilon(t,x) = \frac{1}{\varepsilon^{m\varkappa/4}} u\left(\frac{t}{\varepsilon}, \frac{x}{\varepsilon^{1/2}}\right) \xrightarrow{D} U_m(t,x),\ \varepsilon \to 0,$$

where

$$U_m(t,x) = \frac{C_T^m C_m}{\sqrt{m!}} \int\limits_{\mathbf{R}^{nm}}{}' e^{i\langle x-y, \lambda_1+\ldots+\lambda_m\rangle - \mu t\|\lambda_1+\ldots+\lambda_m\|^2} \frac{W(d\lambda_1)\cdot\cdot W(d\lambda_m)}{(\|\lambda_1\|\cdots\|\lambda_m\|)^{(n-\varkappa)/2}}, \tag{8.13}$$

with C_T being the Tauberian constant defined in (8.12), C_m being defined as in condition **(C)**, *and the multiple stochastic integral (8.13) is taken with the respect to the complex Gaussian white noise random measure defined in (8.9), with diagonal hyperplanes $\lambda_i = \pm\lambda_j, i, j = 1, \ldots m, i \neq j$, excluded.*

Note that, in (8.13), the random field $U_1(t,x)$ is Gaussian, but $U_m(t,x)$ are non-Gaussian for $m \geq 2$. Thus, Theorem 8.2 can be considered as special form of non-central limit theorem, see [111], [38], among others.

Some spectral properties of random fields (8.13) have been studied by [76] and [15]. In particular, for $m = 1$, and each $t > 0$, the limit solution has spatial spectral density

$$g(\lambda) = C_T^2 C_1 e^{-\mu t\|\lambda\|^2} \frac{1}{\|\lambda\|^{n-\varkappa}}.$$

8.2.2 Linear Korteweg-de Vries Equation or Airy Equation

We consider the so-called Linear Korteweg-de Vries equation or Airy equation:

$$\frac{\partial u}{\partial t} = -\frac{\partial^3 u}{\partial x^3}, \quad t > 0, \quad x \in \mathbf{R}, \tag{8.14}$$

subject to random initial condition $u(0,x) = \eta(x),\ x \in \mathbf{R}$, a zero mean stationary stochastic process with covariance function $B_\eta(x) = cov(\eta(0), \eta(x)),\ x \in \mathbf{R}$, satifying conditions **(A)–(D)** for n = 1.

Then (see [22]), the finite-dimensional distributions of the random fields

$$U_\varepsilon(t,x) = \frac{1}{\varepsilon^{1/6}} u\left(\frac{t}{\varepsilon}, \frac{x}{\varepsilon^{1/3}}\right),\ t > 0,\ x \in \mathbf{R},$$

converge weakly, as $\varepsilon \to 0$, to the finite-dimensional distributions of the Gaussian random field $U(t,x)$ with $EU(t,x) = 0$ and the following covariance function:

$$EU(t,x)U(t',x') = \begin{cases} \frac{\sigma^2}{\sqrt{\pi}} \frac{1}{\sqrt[3]{3(t-t')}} Ai\left(\frac{x-x'}{\sqrt[3]{3(t-t')}}\right), & t > t' \\ \frac{\sigma^2}{\sqrt{\pi}} \frac{1}{\sqrt[3]{3(t-t')}} Ai\left(\frac{x-x'}{\sqrt[3]{3(t-t')}}\right), & t < t' \\ B(t-t'), & t = t', \end{cases} \tag{8.15}$$

where

$$Ai(x) = \frac{1}{\sqrt{\pi}} \int_0^\infty \cos\left(\alpha x + \frac{\alpha^3}{3}\right) dx, \quad x \in \mathbf{R},$$

is the Airy function of the first kind and

$$\sigma^2 = \sum_{k=m}^{\infty} \frac{C_k^2}{k!} \int_{\mathbf{R}} B_\xi^k(x) dx.$$

The above limiting covariance structure (8.15) is difficult to predict without using the rescaling limit theorems. It is interesting that the limiting random field is stationary both in space and time.

8.2.3 *Fractional Kinetic Systems or Riesz-Bessel Motion: Micro- and Macro-scalings*

Theorems 8.1 and 8.2 can be generalized to the following so-called fractional kinetic equation or Riesz-Bessel motion:

$$\frac{\partial^\beta u}{\partial t^\beta} = -\mu(I - \Delta)^{\gamma/2}(-\Delta)^{\alpha/2} u, \quad \mu > 0,\ \gamma \geq 0, \quad \alpha \in (0,2],\ \beta \in (0,2], \tag{8.16}$$

with $u = u(t,x),\ t > 0,\ x \in \mathbf{R}^n$, and subject to the stationary random initial condition

$$u_0(x) = \eta(x) = \int_{\mathbf{R}^n} e^{i\langle \lambda, x\rangle} Z_\eta(d\lambda), \quad E\left(Z_\eta(d\lambda)\right)^2 = F_\eta(d\lambda). \tag{8.17}$$

The time derivative of order β is defined in Caputo-Djrbashian sense:

$$\frac{\partial^\beta u}{\partial t^\beta}(t,x) = \frac{1}{\Gamma(1-\beta)}\left[\frac{\partial}{\partial t}\int_0^t (t-\tau)^{-\beta} u(\tau,x)d\tau - \frac{u(0,x)}{t^\beta}\right],$$

and the operators $(-\Delta)^{\alpha/2}$ and $(I-\Delta)^{\gamma/2}$ are interpreted as the inverses of the Riesz and Bessel potentials, respectively, that is, as the inverses of the integral operators, whose kernels have Fourier transforms

$$(2\pi)^{-d/2}\|\lambda\|^{-\alpha},\ \lambda \in R^d,$$

and

$$(2\pi)^{-d/2}\left(1+\|\lambda\|^2\right)^{-\gamma/2},\ \lambda \in R^d,$$

respectively (see, for more details, [4] and [7], [8], [11]). The case of classical diffusion or heat equation with random data corresponds to

$$(\beta,\alpha,\gamma) = (1,2,0),$$

while the wave equation corresponds to

$$(\beta,\alpha,\gamma) = (2,2,0).$$

Thus, fractional equation (8.16) interpolates the heat and wave equations. The time-fractional index β means subdiffusion if $\beta < 1$, diffusion if $\beta = 1$, and superdiffusion if $\beta > 1$. The spatial-fractional Riesz index α means the jumps of the evolution, the Bessel index γ means the tempering of large jumps.

Let us now formulate some results of [7], [8], [9], [10] and [81], [82], [83]. Note that the Fourier transform of the Green function of (8.16) takes the form (see [7]):

$$\hat{G}(t,\xi) = E_\beta\left(-\mu t^\beta \|\xi\|^\alpha (1+\|\xi\|^2)^{\gamma/2}\right),$$

where the Mittag-Leffler function E_β is given by

$$E_\beta(-x) = \sum_{j=0}^{\infty}\frac{(-1)^j x^j}{\Gamma(\beta j+1)},\ x \geq 0.$$

The Green function itself is of the form (assuming that the Fourier transform exists)

$$\begin{aligned} G(t,x) &= p(t,x;\alpha,\gamma,\mu) \\ &= \frac{1}{(2\pi)^n}\int_{\mathbf{R}^n} e^{i\langle\lambda,x\rangle} E_\beta\left(-\mu t^\beta \|\lambda\|^\alpha (1+\|\lambda\|^2)^{\gamma/2}\right) d\lambda. \end{aligned}$$

In particular, the Gaussian kernel is obtained as

$$p(t,x;2,0,\mu)=\frac{1}{(4\pi\mu t)^{n/2}}e^{-\frac{x\|x\|^2}{4\mu t}},$$

and the Cauchy kernel (density) also constitutes a particular case given by

$$p(t,x;0,1,\mu)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n-1}{2}}}\frac{\mu t}{\left[(\mu t)^2+\|x\|^2\right]^{\frac{n-1}{2}}},$$

for $\gamma=0,\alpha\in[0,2]$, the symmetric stable kernel

$$p\left(\frac{t}{2},x;\alpha,0,1\right)=\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}e^{i\langle\lambda,x\rangle-\frac{t}{2}\|x\|^\alpha}d\lambda$$

is defined. For $\gamma=0$, we obtain the Green function in terms of H-function of Fox (see [7], and the references therein):

$$G(t,x)=\frac{\pi^{-n/2}}{\|x\|^{2n}}H_{2,3}^{2,1}\left(\frac{\|x\|^{2\alpha}}{2^\alpha\mu t^\beta}\Big|^{(1,1)\ (1,\beta)}_{(\frac{n}{2},\frac{\alpha}{2})\ (1,1)\ (1,\frac{\alpha}{2})}\right).$$

Then, the solution to the Cauchy problem is

$$\begin{aligned}u(t,x)&=\int_{\mathbf{R}^n}\eta(y)G(t,x-y)dy\\&=\int_{\mathbf{R}^n}e^{i\langle\lambda,x\rangle}E_\beta\left(-\mu t^\beta\|\lambda\|^\alpha(1+\|\lambda\|^2)^{\gamma/2}\right)Z_\eta(d\lambda)\end{aligned}$$

with covariance

$$\begin{aligned}&\mathrm{Cov}(u(t,x),u(s,y))\\&=\int_{\mathbf{R}^n}e^{i\langle\lambda,x-y\rangle}E_\beta\left(-\mu t^\beta\|\lambda\|^\alpha(1+\|\lambda\|^2)^{\beta/2}\right)\\&\quad\times E_\beta\left(-\mu s^\beta\|\lambda\|^\alpha(1+\|\lambda\|^2)^{\beta/2}\right)F_\eta(d\lambda)\end{aligned}$$

and, in the case of LRD Gaussian random initial conditions, Theorem 8.2 can be generalized as follows.

Theorem 8.3. *(see [7] on microscaling) Suppose that condition* ***(E)*** *on the LRD structure is satisfied for the initial random condition in equation (8.10), and* $m\varkappa<\min\{2\alpha,n\}$. *Then, the following convergence in distribution holds*

$$U_\varepsilon(t,x) = \frac{1}{\varepsilon^{m\varkappa\beta/2\alpha}} u\left(\frac{t}{\varepsilon}, \frac{x}{\varepsilon^{\beta/\alpha}}\right) \stackrel{D}{\to} U_m(t,x), \quad \varepsilon \to 0,$$

where

$$U_m(t,x) = \frac{C_T^m C_m}{\sqrt{m!}} \int\limits_{\mathbf{R}^{nm}}{}' e^{i\langle x-y, \lambda_1+\ldots+\lambda_m\rangle} E_\beta\left(-\mu t^\beta \|\lambda_1+\ldots+\lambda_m\|^\alpha\right)$$
$$\times \frac{W(d\lambda_1)\cdot\cdot W(d\lambda_m)}{(\|\lambda_1\|\cdots\|\lambda_m\|)^{(n-\varkappa)/2}}, \tag{8.18}$$

with C_T being the Tauberian constant, defined in (8.12). Constant C_m is given as in condition $(\boldsymbol{C})$*, and the multiple stochastic integral (8.18) is taken with the respect to the complex Gaussian white noise random measure in (8.9), considering the diagonal hyperplanes $\lambda_i = \pm\lambda_j$, $i, j = 1, \ldots m, i \neq j$, excluded.*

Note that in (8.18) the random field $U_1(t,x)$ is Gaussian, but $U_m(t,x)$, $m \geq 2$, are non-Gaussian. Thus, for $m = 1$,

$$U_\varepsilon(t,x) = \frac{1}{\varepsilon^{\varkappa\beta/2\alpha}} u\left(\frac{t}{\varepsilon}, \frac{x}{\varepsilon^{\beta/\alpha}}\right) \stackrel{D}{\to} U_1(t,x), \quad \varepsilon \to 0,$$

where $U_1(t,x)$ is Gaussian field with

$$cov\,(U_1(t,x), U_1(s,y))$$
$$= \int\limits_{\mathbf{R}^n} e^{i\langle x-y,\lambda\rangle} E_\beta\left(-\mu t^\beta \|\lambda\|^\alpha\right) E_\beta\left(-\mu s^\beta \|\lambda\|^\alpha\right) \frac{d\lambda}{\|\lambda\|^{n-\varkappa}},$$

$$\varkappa < \min(2\alpha, n).$$

The spatial spectral density has bifractal form:

$$g(\lambda) = C_T E_\beta^2\left(-\mu t^\beta \|\lambda\|^\alpha\right) \frac{1}{\|\lambda\|^{n-\varkappa}}$$

$$g(\lambda) \sim \frac{const}{\|\lambda\|^{n-\varkappa}}, \quad \|\lambda\| \to 0,$$

$$g(\lambda) \sim \frac{const}{\|\lambda\|^{n-2\alpha-\varkappa}}, \quad \|\lambda\| \to \infty.$$

This is a macro scaling: to "heat up" the initial data in order to have a nontrivial limit. Note that the Riesz index α is the unique parameter that in the limiting distribution plays different roles when macro-scaling procedures are considered.

In contrast, under the so-called micro-scaling ($t \to t\varepsilon\,, \varepsilon \to 0$), both Bessel parameter γ and Riesz parameter α play different roles in the limiting distribution.

We need to "freeze down" both time t and space x (again $L = 1$) if in the random initial condition we consider a rescaling of the form:

$$\eta\left(x\frac{1}{\varepsilon^{\frac{1}{\alpha+\gamma}-\chi}}\right),$$

for a fixed parameter $\chi > 0$. Then, for $m\varkappa < \min\{2(\alpha+\gamma), n\}$ we have the following convergence of finite-dimensional distributions (see, for details, [83]):

$$U_\varepsilon(t,x) = \frac{1}{\varepsilon^{m\varkappa\chi/2}} u\left(t\varepsilon, x\varepsilon^{\frac{\beta}{\alpha+\gamma}}\right) \xrightarrow{D} U_m(t,x), \quad \varepsilon \to 0,$$

where

$$U_m(t,x) = \frac{C_T^m C_m}{\sqrt{m!}} \int\limits_{\mathbf{R}^{nm}} e^{i\langle x-y, \lambda_1+\ldots+\lambda_m\rangle} E_\beta\left(-\mu t^\beta \|\lambda_1+\ldots+\lambda_m\|^{\alpha+\gamma}\right)$$
$$\times \frac{W(d\lambda_1)\cdot\cdot W(d\lambda_m)}{(\|\lambda_1\|\cdots\|\lambda_m\|)^{(n-\varkappa)/2}}.$$

In the above formulae, we assume that fractional parameters are such that the above integral exists (see, for a details, [83]). As before, the random field $U_1(t,x)$ is Gaussian, but $U_m(t,x)$, $m \geq 2$, are non-Gaussian. For $m = 1$, the spatial spectral density has bifractal form

$$g(\lambda) = C_T E_\beta^2\left(-\mu t^\beta \|\lambda\|^{\alpha+\gamma}\right) \frac{1}{\|\lambda\|^{n-\varkappa}}$$

$$g(\lambda) \sim \frac{const}{\|\lambda\|^{n-\varkappa}}, \quad \|\lambda\| \to 0,$$

$$g(\lambda) \sim \frac{const}{\|\lambda\|^{n-2(\alpha+\gamma)-\varkappa}}, \quad \|\lambda\| \to \infty.$$

8.3 Cole-Hopf Solutions of the Burgers Equation

The Burgers equation (8.1) can be reduced to a parabolic type equation by the Hopf-Cole transformation

$$u(t,x) = -2\mu\frac{\partial}{\partial x}\log h(t,x) \tag{8.19}$$

(see e.g. [117]; [93]; [48]; [2], or [25]), which reduces (8.1)-(8.3) to the equation

$$h_t = \mu h_{xx} - \Phi h \tag{8.20}$$

subject to the random initial condition

$$h(0,x) = h_0(x) = \exp\left\{-\frac{U(x)}{2\mu}\right\} = \exp\left\{-\frac{1}{2\mu}\int_{-\infty}^{x} u_0(\xi)d\xi\right\}, \tag{8.21}$$

where, in equation (8.20), $h = h(t,x)$, $t > 0$, $x \in \mathbf{R}$.

The parabolic type equation (8.20) reduces, for $\Phi \equiv 0$, to the classical heat equation

$$h_t = \mu h_{xx}, \tag{8.22}$$

which has, as given before, a fundamental solution of the following form

$$g(t,x) = \frac{1}{\sqrt{4\pi\mu t}}\exp\left\{-\frac{x^2}{4\mu t}\right\}, \quad t > 0,\ x \in \mathbf{R}. \tag{8.23}$$

Thus, the field

$$u(t,x) = \frac{\int_R \frac{x-y}{t} g(t,x-y)e^{-\frac{U(y)}{2\mu}}dy}{\int_R g(t,x-y)e^{-\frac{U(y)}{2\mu}}dy} = \frac{I(t,x)}{J(t,x)} \tag{8.24}$$

solves the initial-value problem (8.1)-(8.3) in the case where the external potential $\Phi \equiv 0$.

We recall Hopf's formulation of this result (see [52]).

Theorem 8.4. *Let $u(t,x)$ be the solution of equation (8.1) with locally integrable initial condition $u(0,x) = u_0(x)$ and external potential $\Phi \equiv 0$. If*

$$\int_0^x u_0(\xi)d\xi = o(x^2), \tag{8.25}$$

as $|x| \to \infty$, then,

$$u(t,x) = \frac{\int_R \frac{x-y}{t}\exp\left\{-\frac{1}{2\mu}F(x,y,t)\right\}dy}{\int_R \exp\left\{-\frac{1}{2\mu}F(x,y,t)\right\}dy}, \tag{8.26}$$

where

$$F(x,y,t) = \frac{(x-y)^2}{2t} + \int_0^y u_0(\xi)d\xi = \frac{(x-y)^2}{2t} + U(y) \tag{8.27}$$

is a solution of (8.1), with $\Phi \equiv 0$, that satisfies

$$\int_0^x u(y,t)dy \to \int_0^a u_0(y)dy \quad as \quad x \to a, \quad t \to 0,$$

for every a. If, in addition, $u_0(x)$ is continuous at $x = a$, then,

$$u(x,t) \to u_0(a) \quad as \quad x \to a, \quad t \to 0.$$

We note that the equation (8.20) with initial condition

$$h(0,x) = h_0(x) \tag{8.28}$$

can be solved explicitly in a number of cases (see, for example, [63], or [66]). In particular, for the linear external potential

$$\Phi(x) = a + bx, \tag{8.29}$$

the field

$$h(t,x) = \exp\left\{-t(a+bx) + \frac{\mu b^2 t^3}{3}\right\} \int_R \frac{e^{-\frac{(x-y-b\mu t^2)^2}{4\mu t}}}{\sqrt{4\pi\mu t}} h_0(y)dy \tag{8.30}$$

solves the initial-value problem (8.20)–(8.21), and, under suitable conditions, the field

$$u(t,x) = 2\mu bt + \frac{\int_R \frac{x-y-b\mu t^2}{t} e^{-\frac{(x-y-b\mu t^2)^2}{4\mu t}} e^{-\frac{U(y)}{2\mu}} dy}{\int_R e^{-\frac{(x-y-b\mu t^2)^2}{4\mu t}} e^{-\frac{U(y)}{2\mu}} dy} \tag{8.31}$$

solves the initial-value problem (8.1)–(8.3) with the external potential (8.29).

For the quadratic external potential

$$\Phi(x) = a + bx^2, \quad b > 0, \tag{8.32}$$

the field

$$\begin{aligned} h(t,x) &= \exp\left\{-at - \sqrt{b}x^2 \tanh(\omega t)/\sqrt{4\mu}\right\} \\ &\times \int_R \frac{\exp\left\{-\frac{[x-y\ \cosh(\omega t)]^2}{\sqrt{\mu/b}\ \sinh(2\omega t)}\right\} h_0(y)dy}{\left[2\pi\sqrt{\mu/b}\ \sinh(\omega t)\right]^{1/2}} \end{aligned} \tag{8.33}$$

solves the initial-value problem (8.20)–(8.21), where

$$\omega = 2\sqrt{\mu b}. \tag{8.34}$$

Note that for $a = 0$ and $b = 1$, the above solution formula can be reduced to Mehler's formula. (For information on Mehler's formula see, for example, [34], pp.286-287 or [24], p.154).

Hence, under suitable conditions, the solution of (8.1)-(8.3) with external potential (8.32) can be written as

$$u(t,x) = 2\sqrt{\mu b}x\ \tanh(\omega t) + 4\mu \frac{\int_R \frac{x-y\cosh(\omega t)}{\sqrt{\mu/b}\sinh(2\omega t)} \exp\left\{-\frac{[x-y\cosh(\omega t)]^2}{\sqrt{\mu/b}\sinh(2\omega t)} - \frac{U(y)}{2\mu}\right\} dy}{\int_R \exp\left\{-\frac{[x-y\cosh(\omega t)]^2}{\sqrt{\mu/b}\sinh(2\omega t)} - \frac{U(y)}{2\mu}\right\} dy}. \tag{8.35}$$

Remark 8.2. For an external potential Φ of general (possibly random) form the solution of Burgers initial-value problem can be expressed by the Feynman-Kac formula (see, for example [88]; [118]; [64], among others), or by the functional integrals (see, [102]).

8.3.1 Burgers Equation

As commented before, Burgers equation with random initial conditions and external potential $\Phi = 0$ has been extensively studied in the last one and a half decade (see references provided in the Introduction). We begin with the one-dimensional Burgers equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \mu \frac{\partial^2 u}{\partial x^2}, \quad t > 0,\ x \in \mathbf{R},$$

subject to the random initial conditions

$$u(0,x) = \frac{d}{dx}\xi(x), \quad x \in \mathbf{R},$$

where $\xi(x)$, $x \in \mathbf{R}$, is a zero mean stationary stochastic process with covariance function $B(x)$. (See also [18] in relation to the Burgers equation with random boundary conditions).

Theorem 8.5. *([2], and [75], [77]). Suppose that condition* ***(D)*** *on the short range dependence (SRD) structure is satisfied for the random initial condition in (8.10). Then, the finite-dimensional distributions of the random fields*

$$U_\varepsilon(t,x) = \frac{1}{\varepsilon^{3/4}} u\left(\frac{t}{\varepsilon}, \frac{x}{\sqrt{\varepsilon}}\right),\ t > 0,\ x \in \mathbf{R},$$

converge weakly ($\xrightarrow{D}$), *as* $\varepsilon \to 0$, *to the finite-dimensional distributions of the Gaussian random field* $U(t,x)$, *with* $EU(t,x) = 0$ *and the following covariance function:*

$$EU(t,x)U(t',x') = const\times$$

$$\times \frac{1}{(t+t')^{3/2}} \left(1 - \frac{(x-x')^2}{2\mu(t+t')}\right)^2 \exp\left\{-\frac{(x-x')^2}{2\mu(t+t')}\right\},$$

where the constant is given in [75] and [77].

Note that the spectral density of the above limiting random field is of the form

$$g(\lambda) = const \times \lambda^2 e^{-\lambda^2 \mu(t+t')}, \ \lambda \in \mathbf{R}.$$

This random field is stationary in space, but not in time.

Multidimensional generalization of these results can be found in [2], [75] [77]), and [13], [15]). In the above papers, one can find also the investigation of limiting distributions for strongly dependent initial conditions, in particular, for convergence to non-Gaussian distributions, see [65], [67], [69]. The limiting distributions in the so-called KPZ turbulence is investigated in [15].

8.3.2 Non-Gaussian Scenarios in Burgers Turbulence and their Spectra

In this chapter, parabolically rescaled solutions of Burgers equations are studied. These parabolically rescaled solutions are in fact approximations to the hyperbolically rescaled solutions. We present the second- and higher-order spectral densities of homogeneous (in space) random fields arising as rescaled solutions of the Burgers equations with singular non-Gaussian initial conditions. This subject is a continuation of those by [75], [76], [77]), in which the second-order spectral densities were studied for the Burgers turbulence problem with non-Gaussian singular data, and [13], in which second- and higher-order spectral densities were given for fractional random fields arising as rescaled solutions of the heat and fractional heat equations with singular random data (for further details on these equations, see [5], [6] and [8]).

In a sense, non-Gaussian scenarios are more realistic models of zero viscosity than Gaussian scenarios. Furthermore, to provide a full description of singularity, we have to consider higher-order spectral densities and their singular properties. But even for the second order, our results for the spectral density in one dimension can be compared with the results of [49]. Indeed the singular property of the energy spectrum of the initial condition (8.42) below is transformed by the Burgers equation into the singular property, for $n = 1$, $l = k = 1$, and up to a constant, as $|\lambda|^a e^{-2\mu t \lambda^2}$, with $a = 2\varkappa + 1$, $0 < \varkappa < 1/2$. This result is exactly the same as formula (122) of [49]. However, we can see that these singular properties depend on the dimension n, and the results change dramatically starting from dimension $n \geq 3$. In our opinion, both non-Gaussian scenarios in parabolically rescaled Burgers and KPZ equations and singular properties of higher-order spectral densities provide a

description of Burgers and KPZ turbulence complementary to that of [49] and [47], via vanishing viscosity together with a power-law investigation of the solutions.

The closed-form expressions of higher-order spectral densities in turn will play an essential role in the statistical estimation of these random fields. In fact, in the presence of possible long-range dependence, non-Gaussianity and non-linearity inherent in the formulated models, particularly in a situation where useful information is contained in higher orders rather than the second order, an estimation theory using information in higher-order spectral densities is more viable. Some components of such a theory are provided in [14] and [12], based on the minimum contrast principle.

Consider the n-dimensional Burgers equation

$$\frac{\partial u}{\partial t} + (u, \nabla)u = \mu \Delta u, \quad \mu > 0, \tag{8.36}$$

subject to the random initial conditions in potential form:

$$u(0, x) = \nabla U(x), \quad x \in \mathbf{R}^n, \tag{8.37}$$

where Δ denotes the n-dimensional Laplacian, and ∇ the gradient operator in $\mathbf{R}^n$. Equation (8.36) describes the time evolution of the velocity field

$$u(t, x) = (u_1(t, x), ..., u_n(t, x)), \ (t, x) \in (0, \infty) \times \mathbf{R}^n, \ n \geq 1.$$

We will assume that the initial velocity potential $U(x)$ is a scalar random field of the form described in Condition **F** below.

Equation (8.36) is a parabolic equation with quadratic, inertial nonlinearity, which can be viewed as a simplified version of the Navier-Stokes equation with the pressure term ∇p omitted, and with the viscosity coefficient μ corresponding to the inverse of the Reynolds number (see [93], p.152). As commented before, via the Cole-Hopf transformation,

$$u(t, x) = -2\mu \nabla \log h(t, x), \tag{8.38}$$

the Burgers problem (8.36)-(8.37) is reduced to the parabolic-type equation

$$\frac{\partial h}{\partial t} = \mu \Delta h, \quad t > 0, \quad x \in \mathbf{R}^n, \tag{8.39}$$

subject to the random initial condition

$$h(0, x) = h_0(x) = \exp\left\{-\frac{U(x)}{2\mu}\right\} \tag{8.40}$$

(see e.g. [117], and [48]).

We introduce the following condition concerning the initial velocity potential.

F. The initial velocity potential $U(x)$ is a random field of the form

$$U(x) = \xi^2(x) - 1, \quad x \in \mathbf{R}^n,$$

where the random field $\xi(x)$ is a real measurable homogeneous and isotropic Gaussian field with $E\xi(x) = 0$, $E\xi^2(x) = 1$ and covariance function of the form

$$B_\xi(x) = \|x\|^{-\varkappa} L(\|x\|), \quad , 0 < \varkappa < n, \quad \text{as} \quad \|x\| \to \infty, \tag{8.41}$$

where the function $L(t)$, $t > 0$, is slowly varying at infinity and is bounded on each bounded interval. Furthermore, the spectral density $f_\xi(\lambda)$, $\lambda \in \mathbf{R}^n$, of the field $\xi(x)$ exists, is decreasing for $\|\lambda\| \geq \lambda_0 > 0$ and continuous for all $\lambda \neq 0$.

Note that the random field $\xi(x)$ of Condition A can be represented as

$$\xi(x) = \int_{\mathbf{R}^n} e^{i(\lambda,x)} \sqrt{f_\xi(\lambda)} W(d\lambda),$$

where $W(\cdot)$ is a Gaussian white noise, and, as commented before, from the Tauberian theorem for Hankel type transform (see, for instance, [64], Theorem 1.1.4), we obtain that the spectral density $f_\xi(\lambda)$ satisfies

$$f_\xi(\|\lambda\|) \sim \|\lambda\|^{\varkappa - n} L\left(\frac{1}{\|\lambda\|}\right) c(n, \varkappa), \quad 0 < \varkappa < n, \quad \|\lambda\| \to 0, \tag{8.42}$$

where $c(n, \varkappa)$ is the Tauberian constant

$$C_T(n, \varkappa) = \frac{\Gamma\left(\frac{n-\varkappa}{2}\right)}{2^\varkappa \pi^{n/2} \Gamma(\varkappa/2)}. \tag{8.43}$$

Identity (8.42) means that the random initial condition under consideration displays a singular property; in fact, the random field $\xi(x)$ will then have long-range dependence.

We will study the spectral properties of the limit distributions of the rescaled solutions, namely, with parabolic scaling, of the Burgers equation (8.36) with initial data (8.37) satisfying Condition **F**. These parabolic scaling limits of the solution can be described in terms of their multiple stochastic integral representation as stated in the following theorem (see [69], [68] or [64]).

Theorem 8.6. *Let $u(t, x)$, $(t, x) \in (0, \infty) \times \mathbf{R}^n$, be a solution of the initial value problem (8.36)-(8.37) with ξ defining the initial velocity potential $U(x) = \xi^2(x) - 1$ satisfying Condition **F**, and $\alpha \in (0, n/2)$. Then, the finite-dimensional distributions of the random fields*

$$Z_\varepsilon(t,x)=\frac{\varepsilon^{-(1+\varkappa)/2}}{L(1/\sqrt{\varepsilon})}u\left(t/\varepsilon,x/\sqrt{\varepsilon}\right),\quad (t,\,x)\in(0,\infty)\times\mathbf{R}^n,\ 0<\varkappa<n/2,$$

converge weakly, as $\varepsilon\longrightarrow 0$*, to the finite-dimensional distributions of the vector field* $Z(t,x)$, $(t,\,x)\in(0,\infty)\times\mathbf{R}^n$, *homogeneous in* x*, with the following multiple stochastic integral representation:*

$$Z(t,x)=C(\mu)C_T(n,\varkappa)\int_{\mathbf{R}^{2n}}^{\prime}\frac{e^{i(x,\lambda_1+\lambda_2)-\mu t\|\lambda_1+\lambda_2\|^2}(\lambda_1+\lambda_2)}{(\|\lambda_1\|\,\|\lambda_2\|)^{(n-\varkappa)/2}}W(d\lambda_1)\,W(d\lambda_2),\tag{8.44}$$

$(t,\,x)\in(0,\infty)\times\mathbf{R}^n$, *where the constant* $C_T(n,\varkappa)$ *is given by (8.43),*

$$C(\mu)=\frac{\mu^2 i}{1+\mu},\tag{8.45}$$

and the double stochastic integral $\int^{\prime}\ldots$ *is evaluated with respect to the Gaussian complex white noise measure* $W(\cdot)$ *in* $\mathbf{R}^n$*, with the diagonal hyperplanes* $\lambda_1=\pm\lambda_2$ *being excluded from the domain of the integration.*

The second-order spectral densities $f(\lambda)=(f_{lk}(\lambda))_{l,k=1}^n$ of the non-Gaussian vector random field $Z(t,x)$ representing the limit of the parabolically rescaled solution of problem (8.36)-(8.37) are given as follows:

$$\begin{aligned}f_{lk}(\lambda)&=const\times e^{-2\mu t\|\lambda\|^2}\|\lambda\|^{2\varkappa-n}\lambda^{(l)}\lambda^{(k)},\\ l,k&=1,\ldots,n,\quad \lambda=\left(\lambda^{(1)},\ldots,\lambda^{(n)}\right)\in\mathbf{R}^n.\end{aligned}\tag{8.46}$$

Note that different non-Gaussian scenarios are also given in [65], [68], [42].

8.3.3 Heat and Burgers Equations with Quadratic External Potential: Multidimensional Case

The consideration of quadratic external potentials in Burgers equation with random initial data involves a non-periodic behavior, in general, a non-compact situation, in the model representing the external force acting on the fluid, according, for example, with the large-scale forcing in the hydrodynamics setting. In this section, we consider this external forcing scenario under strongly dependent random initial conditions. The strong dependence parameter of the initial velocity potential characterizes the limiting distribution. In the heat equation, this parameter also defines the strong dependence properties of the limiting random field. However, in the Burger equation, the limiting random field does not display strong dependence,

but it is a fractal random field displaying a local singular behavior characterized by the LRD parameter of the initial velocity potential.

We have considered, with $\omega = 2\sqrt{\mu b}$ and $\varepsilon \to 0$, the rescaling procedure

$$t \mapsto \frac{1}{\omega}\ln\frac{t}{\varepsilon};\ \ x \mapsto \frac{xt}{\varepsilon^2};\ 0 < \varepsilon < t.$$

Our rescaling procedures differ significantly from the rescaling procedure

$$t \mapsto \frac{t}{\varepsilon};\ \ x \mapsto \frac{x}{\sqrt{\varepsilon}};$$

of the heat and Burgers equations with external potential $\Phi = 0$ (see [2], [75], [76], [118], [64], [5], [8]), [95] and references therein) or the rescaling procedure

$$t \mapsto \frac{t}{\varepsilon};\ \ x \mapsto \frac{x}{\sqrt{\varepsilon}} + b\mu\frac{t^2}{\varepsilon^2};$$

of the heat and Burgers equations with a linear external potential $\Phi(x) = a + bx$ (see [66]).

In this section, we review the results by [70] on the limit behavior, in the sense of the convergence of finite-dimensional distributions, of a scaling version of the solution to the initial-value problem

$$\begin{aligned}
&\frac{\partial}{\partial t}\mathbf{u} + \langle \mathbf{u}, \nabla\rangle\,\mathbf{u} = \mu\Delta\mathbf{u} + 2\mu\nabla\Phi \\
&\mathbf{u}(0,\mathbf{x}) = \mathbf{u}_0(\mathbf{x}) = \nabla U(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R}^n,
\end{aligned} \tag{8.47}$$

where $\mathbf{u} = \mathbf{u}(t,\mathbf{x}),\ \mathbf{x} \in \mathbf{R}^n$, is a vector random field, the scalar field U represents the initial velocity potential. Here, as usual, ∇ denotes the gradient and Δ denotes the Laplacian. We consider the case of quadratic external potentials, that is, the case where

$$\Phi = a + b\|\mathbf{x}\|^2 = a + b\left(\sum_{i=1}^{n} x_i^2\right), \quad b > 0.$$

The solution $\mathbf{u}$ to problem (8.47) can be defined as the Hopf-Cole transformation

$$\mathbf{u}(t,\mathbf{x}) = -2\mu\nabla\ln h(t,\mathbf{x}) \tag{8.48}$$

of the solution h to the following heat equation:

$$\begin{aligned}
&\frac{\partial}{\partial t}h = \mu\Delta h - \Phi h \\
&h(0,\mathbf{x}) = h_0(\mathbf{x}) = G(u(\mathbf{x})), \quad \mathbf{x} \in \mathbf{R}^n
\end{aligned} \tag{8.49}$$

(see, for example, [104] and [25]). The initial velocity potential U is assumed to be given by an homogeneous and isotropic random field ξ with spectral density

$$f_\xi(\boldsymbol{\lambda}) = \frac{(2\pi)^\gamma}{\varphi(\gamma)} \|\boldsymbol{\lambda}\|^{-n+2\gamma}(1+\|\boldsymbol{\lambda}\|^2)^{-\alpha}, \quad \gamma > 0,\ \alpha - \gamma > 0, \tag{8.50}$$

where

$$\varphi(\gamma) = \frac{\pi^{n/2} 2^\gamma \Gamma(\frac{\gamma}{2})}{\Gamma\left(\frac{n}{2} - \frac{\gamma}{2}\right)}.$$

Hence, ξ admits a spectral representation in the L_2-stochastic integral form given by:

$$\xi(\mathbf{x}) = \int_{\mathbf{R}} \exp\{i\ \langle \boldsymbol{\lambda}, \mathbf{x}\rangle\}\ Z_\xi(d\boldsymbol{\lambda}) = \int_{\mathbf{R}^n} \exp\{i\ \langle \boldsymbol{\lambda}, \mathbf{x}\rangle\}\ f_\xi^{1/2}(\boldsymbol{\lambda}) W(d\boldsymbol{\lambda}), \tag{8.51}$$

with the associated spectral representation of its covariance function

$$B_\xi(\mathbf{x}) = \mathrm{cov}(\xi(\mathbf{0}), \xi(\mathbf{x})) = \int_{\mathbf{R}^n} \exp\{i\ \langle \boldsymbol{\lambda}, \mathbf{x}\rangle\}\ F_\xi(d\boldsymbol{\lambda}),$$

in terms of the bounded and positive spectral measure F_ξ, which defines the second-order structure of the complex-valued orthogonally scattered random measure $Z_\xi(\cdot)$ such that $E\left|Z_\xi(d\boldsymbol{\lambda})\right|^2 = F_\xi(d\boldsymbol{\lambda})$, and with

$$F_\xi(d\boldsymbol{\lambda}) = f_\xi(\boldsymbol{\lambda}) d\boldsymbol{\lambda}.$$

Here, W represents a Gaussian white noise measure. For $\gamma < n/2$, this model displays long-range dependence (see, for example, [4]).

As before, we consider the Chebyshev-Hermite expansion of function

$$G(u) = \exp\left\{-\frac{u}{2\mu}\right\}, \quad u \in \mathbf{R}. \tag{8.52}$$

Thus,

$$h_0(\mathbf{x}) = \exp\left\{-\frac{\xi(\mathbf{x})}{2\mu}\right\} = \sum_{k=0}^{\infty} \frac{C_k}{k!} H_m(\xi(\mathbf{x})),$$

where

$$S(\mathbf{x}) = \mathrm{cov}(h_0(\mathbf{0}), h_0(\mathbf{x})) = B_{h_0}(\mathbf{x}) = \sum_{k=1}^{\infty} \frac{C_k^2}{k!} B_\xi^k(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R}^n. \tag{8.53}$$

The stationary process $h_0(\mathbf{x}) = \exp\left\{-\frac{\xi(\mathbf{x})}{2\mu}\right\}$, $\mathbf{x} \in \mathbf{R}^n$, then admits the spectral representation in the form of L_2-stochastic integral:

$$h_0(\mathbf{x}) = \exp\left\{-\frac{\xi(\mathbf{x})}{2\mu}\right\} = \exp\left\{\frac{1}{8\mu^2}\right\} + \int_{\mathbf{R}^n} \exp\{i\,\langle \boldsymbol{\lambda}, \mathbf{x}\rangle\}\, Z_{h_0}(d\boldsymbol{\lambda}), \tag{8.54}$$

with the associated spectral representation of its covariance function

$$B_{h_0}(\mathbf{x}) = \operatorname{cov}(h_0(\mathbf{0}), h_0(\mathbf{x})) = \int_{\mathbf{R}^n} \exp\{i\,\langle \boldsymbol{\lambda}, \mathbf{x}\rangle\}\, F_{h_0}(d\boldsymbol{\lambda}),$$

in terms of the bounded and positive spectral measure F_{h_0}, which defines the second-order structure of the complex-valued orthogonally scattered random measure $Z_{h_0}(\cdot)$ such that $E\left|Z_{h_0}(d\boldsymbol{\lambda})\right|^2 = F_{h_0}(d\boldsymbol{\lambda})$.

8.3.3.1 Main Results

The solution h to equation (8.49) is defined as

$$\begin{aligned} h(t,\mathbf{x}) &= \frac{\exp\{-at\}}{\left(\cosh(\omega t) 2\pi \left(\frac{\mu}{b}\right)^{1/2} \tanh(\omega t)\right)^{n/2}} \exp\left\{-1/2 \left(\frac{b}{\mu}\right)^{1/2} \tanh(\omega t)\|\mathbf{x}\|^2\right\} \\ &\times \int_{\mathbf{R}^n} \exp\left\{-\frac{\|\frac{\mathbf{x}}{\cosh(\omega t)} - \mathbf{x}'\|^2}{2\left(\frac{\mu}{b}\right)^{1/2} \tanh(\omega t)}\right\} h(0,\mathbf{x}')d\mathbf{x}' \\ &= \frac{\exp\{-at\}}{(\cosh(\omega t))^{n/2}} \exp\left\{-1/2 \left(\frac{b}{\mu}\right)^{1/2} \tanh(\omega t)\|\mathbf{x}\|^2\right\} \\ &\times \int_{\mathbf{R}^n} \exp\left\{i\left\langle \boldsymbol{\lambda}, \frac{\mathbf{x}}{\cosh(\omega t)}\right\rangle\right\} \exp\left\{-\left(\frac{\mu}{b}\right)^{1/2} \frac{\tanh(\omega t)}{2}\|\boldsymbol{\lambda}\|^2\right\} Z_{h_0}(d\boldsymbol{\lambda}), \end{aligned} \tag{8.55}$$

where Z_{h_0} is given as in equation (8.54), $\|\boldsymbol{\lambda}\|^2 = \sum_{i=1}^{n} \lambda_i^2$, and

$$\omega = 2(\mu b)^{1/2}.$$

The following result provides the limiting Gaussian random field distribution of a rescaling of the random field solution (8.55).

Theorem 8.7. *Let h be the solution (8.55) to problem (8.49). Then, the finite-dimensional distributions of the generalized random field $\widetilde{\mathscr{H}}_\varepsilon$ defined by the ordinary random field*

$$H_\varepsilon(t,\mathbf{x}) = \frac{\exp\left\{\frac{1}{2}\left(\frac{b}{\mu}\right)^{1/2}\|\frac{\mathbf{x}t}{\varepsilon^2}\|^2\left[\frac{\varepsilon^2-t^2}{\varepsilon^2+t^2}\right]\right\}}{\varepsilon^{a/\omega+n/2+\gamma}} h\left(\frac{1}{\omega}\ln\left(\frac{t}{\varepsilon}\right), \frac{\mathbf{x}t}{\varepsilon^2}\right),$$
$$\omega = 2(\mu b)^{1/2},\ 0<\varepsilon<t, \tag{8.56}$$

converge, when ε goes to zero, to the finite-dimensional distributions of the Gaussian generalized random field $\widetilde{\mathscr{H}}$ with zero mean and covariance function

$$E\left[\widetilde{\mathscr{H}}(t_1,\phi)\widetilde{\mathscr{H}}(t_2,\varphi)\right] = R(t_1,\phi,t_2,\varphi) =$$
$$= (t_1t_2)^{-(a/\omega+n/2)}2^{1-2\gamma}\int_{\mathbf{R}}\int_{\mathbf{R}^n}\phi(\mathbf{x}_1)\|\mathbf{x}_1-\mathbf{x}_2\|^{-2\gamma}\varphi(\mathbf{x}_2)d\mathbf{x}_1d\mathbf{x}_2, \tag{8.57}$$

for $0<\gamma<n/2$, and $\phi,\varphi\in H^{\gamma-n/2}(\mathbf{R}^n)$.

Let us now consider the Burgers problem (8.47). Its solution is defined as

$$\mathbf{u}(t,\mathbf{x}) = -2\mu\nabla\ln\left(h(t,\mathbf{x})\right) = -2\mu\nabla\left[-\frac{1}{2}\mathbf{x}\mathbf{W}\mathbf{x}^T\right]$$
$$-2\mu\frac{\nabla\left[\exp\left\{\mathbf{x}\mathbf{V}\nabla\right\}\exp\left\{\frac{1}{2}\nabla\mathbf{Z}\nabla^T\right\}h_0\right](t,\mathbf{x})}{\left[\exp\left\{\mathbf{x}\mathbf{V}\nabla\right\}\exp\left\{\frac{1}{2}\nabla\mathbf{Z}\nabla^T\right\}h_0\right](t,\mathbf{x})}$$
$$= \sqrt{\mu b}\tanh(\omega t)\nabla\|\mathbf{x}\|^2 - 2\mu\frac{\nabla\left[\exp\left\{\frac{1}{2}\nabla\mathbf{Z}\nabla^T\right\}h_0\left(t,\frac{\mathbf{x}}{\cosh(\omega t)}\right)\right]}{\left[\exp\left\{\frac{1}{2}\nabla\mathbf{Z}\nabla^T\right\}h_0\right]\left(t,\frac{\mathbf{x}}{\cosh(\omega t)}\right)}. \tag{8.58}$$

In particular, we have

$$u_j(t,\mathbf{x}) = 2\sqrt{\mu b}x_j\ \tanh(\omega t)$$
$$+4\mu\frac{\int_{\mathbf{R}^n}\frac{x_j-y_j\cosh(\omega t)}{\sqrt{\mu/b}\sinh(2\omega t)}\exp\left\{-\frac{[\mathbf{x}-\mathbf{y}\cosh(\omega t)]^2}{\sqrt{\mu/b}\sinh(2\omega t)}-\frac{U(\mathbf{y})}{2\mu}\right\}d\mathbf{y}}{\int_{\mathbf{R}^n}\exp\left\{-\frac{[\mathbf{x}-\mathbf{y}\cosh(\omega t)]^2}{\sqrt{\mu/b}\sinh(2\omega t)}-\frac{U(\mathbf{y})}{2\mu}\right\}d\mathbf{y}},\ j=1,\ldots,n. \tag{8.59}$$

The following result provides the limiting random field distribution, under LRD random initial conditions, for the Burgers problem (8.47).

Theorem 8.8. *Let $\mathbf{u}$ be the solution (8.58)–(8.59) to the Burgers equation (8.47) with quadratic external potential, considering random initial velocity potential $U=\xi$ with spectral density (8.50), and with G being defined as in (8.52). Then,*

the finite-dimensional distributions of the vector generalized random field $\mathscr{U}^{\varepsilon}$ associated with the ordinary random field

$$\mathbf{U}^{\varepsilon}(t,\mathbf{x}) = \frac{1}{\varepsilon^{n/2+2+\gamma}} \left(\frac{(2\pi)^{\gamma}}{\varphi(\gamma)}\right)^{-1/2} \left[\mathbf{u}\left(\frac{1}{\omega}\ln\left[\frac{t}{\varepsilon}\right], \frac{\mathbf{x}t}{\varepsilon^2}\right) - \mathbf{F}\left(\frac{1}{\omega}\ln\left[\frac{t}{\varepsilon}\right], \frac{\mathbf{x}t}{\varepsilon^2}\right)\right], \quad 0 < \varepsilon < t, \tag{8.60}$$

converge to the finite-dimensional distributions of the Gaussian vector generalized random field $\mathscr{U}$ with covariance matrix given by, for $i, j = 1, \ldots, n$,

$$E\left[\mathscr{U}_i(t_1,\phi)\mathscr{U}_i(t_2,\varphi)\right] = \frac{C_1^2 2^{n+2-2\gamma}\mu^2}{C_0^2 (t_1 t_2)^{n/2+1}} \times \int_{\mathbf{R}^n}\int_{\mathbf{R}^n} \phi(\mathbf{x}_1)\left[\frac{-\partial^2}{\partial(x_1^i - x_2^i)^2}\|\mathbf{x}_1 - \mathbf{x}_2\|^{-2\gamma}\right]\varphi(\mathbf{x}_2)\, d\mathbf{x}_1 d\mathbf{x}_2, \tag{8.61}$$

$$E\left[\mathscr{U}_i(t_1,\phi)\mathscr{U}_j(t_2,\varphi)\right] = \frac{C_1^2 2^{n+2-2\gamma}\mu^2}{C_0^2 (t_1 t_2)^{n/2+1}} \times \int_{\mathbf{R}^n}\int_{\mathbf{R}^n} \phi(\mathbf{x}_1)\left[\frac{-\partial^2}{\partial(x_1^j - x_2^j)\partial(x_1^i - x_2^i)}\|\mathbf{x}_1 - \mathbf{x}_2\|^{-2\gamma}\right] \times \varphi(\mathbf{x}_2)\, d\mathbf{x}_1 d\mathbf{x}_2, \tag{8.62}$$

where $0 < \gamma < n/2$ and $\phi, \varphi \in H^{\gamma - n/2 + 1/n}(\mathbf{R}^n)$, with

$$\begin{aligned}\mathbf{F}(t,\mathbf{x}) &= (F_1(t,\mathbf{x}), \ldots, F_n(t,\mathbf{x})) \\ &= \left(2(\mu b)^{1/2}\tanh(\omega t)x_1, \ldots, 2(\mu b)^{1/2}\tanh(\omega t)x_n\right) \\ &= (\omega\tanh(\omega t)x_1, \ldots, \omega\tanh(\omega t)x_n)\,.\end{aligned} \tag{8.63}$$

(The proof of the above assertions can be found in [70]).

8.4 Other Possible Scenarios: Weak-Dependent Random Initial Conditions

In this section, we provide an alternative scaling procedure which allows to derive the limiting Gaussian distributions of the solutions to the heat and Burgers equations with quadratic external potentials, under weak-dependent scenarios. The results reviewed here are obtained in [73]. Let us then consider the solution u to

$$\frac{\partial}{\partial t}u(t,x)+u(t,x)\frac{\partial}{\partial x}u(t,x)=\mu\frac{\partial^2}{\partial x^2}u(t,x)+2\mu\frac{\partial}{\partial x}\Phi(x), \tag{8.64}$$

subject to the initial condition

$$u(0,x)=u_0(x), \tag{8.65}$$

where, as before, $u=u(t,x)$, $t>0$, $x\in\mathbf{R}$, is a velocity field,

$$U=U(x)=\int_{-\infty}^{x}u_0(y)dy \tag{8.66}$$

is the initial velocity potential, $\Phi=\Phi(x)$, $x\in\mathbf{R}$, is the external potential, $\mu>0$ is the viscosity of the medium and the reciprocal $\mathscr{R}=1/\mu$ corresponds to the Reynolds number. As in the previous section, $\Phi=a+bx^2$, $b\neq 0$. However, our interest, in this section, relies on the case where the initial velocity potential $U=\xi$ is a stationary random process displaying short range dependence. Thus, we assume that conditions **(A)**-**(D)** hold.

The scaling procedure established, as $t\to\infty$, is the following

$$\begin{aligned} t &\longrightarrow \frac{1}{\omega}\ln t\\ x &\longrightarrow \frac{x}{\widetilde{\rho}(t)}\\ \xi(x) &\longrightarrow \xi(\rho(t)x), \end{aligned} \tag{8.67}$$

where $\widetilde{\rho}$ and ρ satisfy suitable asymptotic conditions formulated below, in order to apply the Central Limit Theorem derived in [54]. In particular, we consider

$$\widetilde{\rho}(t)=\frac{\rho(t)}{t}.$$

Additionally, the random field

$$H(t,x)=H(t,x,\xi_\varphi)=\int_{-\psi(t)}^{\psi(t)}\frac{1}{\nu_x(t)}h\left(\frac{x-y}{w(t)}\right)G\left(\xi\left(\frac{y}{\varphi(t)}\right)\right)dy \tag{8.68}$$

is also considered for the proof of the results below (see [73]), and the following conditions are also assumed:

(G) Function $G\in\left(L^2(\mathbf{R});\phi(z)dz\right)$, *with* $\phi(z)$ *being the standard Gaussian measure.*

In our definition of function G we have assumed Conditions **(B)**-**(C)**, which, under condition **(A)**, is equivalent to Condition **(G)**.

(H) *Function ψ in equation (8.68) must be positive, tending to infinity when t goes to infinity, and satisfying*

$$\lim_{t\longrightarrow\infty}\frac{\psi(t)}{w(t)}=\infty, \tag{8.69}$$

$$\lim_{t\longrightarrow\infty}\frac{\psi(t)}{w^{2+\delta}(t)}=0, \quad \text{for any } \delta>0, \tag{8.70}$$

$$\lim_{t\longrightarrow\infty}\frac{\psi(t)}{\varphi(t)}=\infty. \tag{8.71}$$

Equations (8.69) and (8.70) hold if and only if ρ is chosen such that

$$\lim_{t\longrightarrow\infty}\frac{\psi(t)}{\rho(t)}=\infty, \tag{8.72}$$

$$\lim_{t\longrightarrow\infty}\frac{\psi(t)}{\rho^{2+\delta}(t)}=0, \quad \text{for any } \delta>0. \tag{8.73}$$

Remark 8.3. In our definition of the scaling procedure we have considered scaling (8.67) with $\widetilde{\rho}(t)=\frac{\rho(t)}{t}$. Thus, the following time-dependent linear transformation of the spatial component

$$x\longrightarrow\frac{xt}{\rho(t)}$$

is considered. Other valid choice of function $\widetilde{\rho}$ could be

$$\widetilde{\rho}=\rho.$$

However, in this case, equations (8.72) and (8.73) must be reformulated in terms of the quotient between function ψ and function $\rho(t)\sqrt{\sinh(2\ln t)}$.

Equation (8.71) holds, since $\psi(t)\longrightarrow\infty$ and $\varphi(t)\longrightarrow 1/2$, when $t\longrightarrow\infty$.

(I) *Function h must be in $L^2(\mathbf{R})$, and for $0<\beta<1$,*

$$\int_0^\infty \frac{h*\bar{h}}{s^\beta}ds<\infty,$$

where $\bar{h}=h$.

For our selection of function $h(z)=\exp\left\{-|z|^2\right\}$, we have that such function is square integrable and since

$$\frac{h*\bar{h}(s)}{s^\beta}=\frac{\exp\left\{-\frac{|s|^2}{2}\right\}}{s^\beta},$$

then,

$$\int_0^\infty \frac{h * \bar{h}}{s^\beta} ds < \infty, \quad 0 < \beta < 1.$$

Thus, condition **(I)** holds.

(J) *Functions v_x and w, defining the scaling of h, must satisfy*

$$\lim_{t \longrightarrow \infty} \frac{v_x^2(t)}{w(t)} = \infty.$$

With the choice made of functions v and w, the following identities hold

$$\begin{aligned}
\frac{v_x^2(t)}{w(t)} &= \frac{\left[\frac{\rho(t)[2\pi\sqrt{\mu/b}\sinh(\ln(t))]^{1/2} t^{a/\omega-1}\cosh(\ln(t))}{\exp\left\{\left(\frac{b}{\mu}\right)^{1/2}\left(\frac{xt}{\rho(t)}\right)^2[1-\tanh(\ln(t))]\right\}}\right]^2}{t^{-1}\rho(t)\left[\sqrt{\frac{\mu}{b}}\sinh(2\ln t)\right]^{1/2}} \\
&= \frac{[\rho(t)]^2(2\pi\sqrt{\mu/b}\sinh(\ln(t)))t^{2a/\omega-2}[\cosh(\ln(t))]^2}{\exp\left\{\left(\frac{b}{\mu}\right)^{1/2}\left(\frac{xt}{\rho(t)}\right)^2[1-\tanh(\ln(t))]\right\} t^{-1}\rho(t)\sqrt{(\mu/b)^{1/2}\sinh(2\ln t)}} \\
&= \frac{\rho(t)2\pi\sqrt{\mu/b}(t^2-1)t^{2a/\omega+2}(t^2+1)^2}{4t^5\exp\left\{\left(\frac{b}{\mu}\right)^{1/2}\left(\frac{xt}{\rho(t)}\right)^2\left[\frac{2}{t^2+1}\right]\right\}(t^4-1)^{1/2}(\mu/b)^{1/4}} \longrightarrow \infty \quad t \longrightarrow \infty.
\end{aligned} \tag{8.74}$$

Function B_ξ is also assumed to be such that

$$\int_{\mathbf{R}} |S(z)| dz < \infty, \int_{\mathbf{R}} S(z) dz \neq 0, \tag{8.75}$$

which implies

$$\int_{\mathbf{R}} |S_\varphi(z)| dz < \infty,$$

where

$$S_\varphi(z) = \sum_{k=m}^{\infty} \frac{C_k^2}{k!} B_{\xi_\varphi}^k(z),$$

in the case where the Hermite rank is $m \geq 1$, with B_{ξ_φ} being the covariance function of ξ_φ.

Under the above setting of conditions, the following results hold (see [73]):

Proposition 8.1. *If B_ξ satisfies (8.75), then*

$$(i) \quad \sigma_1^2(G) = \lim_{t \longrightarrow \infty} \mathrm{Var}\left[\frac{v_x(t)}{\sqrt{w(t)}} H(t, x, \xi_\varphi)\right] = \|h\|_2^2 \|S\|_1,$$

*where $H(t,x,\xi_\varphi)$, v_x, w, h and S are defined as before, satisfying conditions **(G)-(J)**.*

$$(ii) \quad \lim_{t\longrightarrow\infty} \mathrm{Var}\left[\frac{v_x(t)}{\sqrt{w(t)}}\left(\widetilde{H}(t,x,\xi_\varphi)-H(t,x,\xi_\varphi)\right)\right]=0, \tag{8.76}$$

where

$$\widetilde{H}(t,x,\xi_\varphi)=\int_{-\infty}^{\infty}\frac{1}{v_x(t)}h\left(\frac{x-y}{w(t)}\right)G\left(\xi\left(\frac{y}{\varphi(t)}\right)\right)dy. \tag{8.77}$$

Theorem 8.9. *Under condition (8.75), when $t\to\infty$,*

$$\frac{v_x(t)}{\sqrt{w(t)}}\left[\widetilde{H}\left(\frac{1}{\omega}\ln t,xt\right)-\mathbf{E}\left(\widetilde{H}\left(\frac{1}{\omega}\ln t,xt\right)\right)\right]\xrightarrow{w} \quad H(\xi_{\varphi_\infty})\sim N(0,\sigma_1^2(G)), \tag{8.78}$$

where, as before, $\widetilde{H}$ is defined as in equation (8.77).

For the limiting distribution of the random field solution to the Burgers equation the following computations and result are derived in [73]:

Considering first the function

$$F(t,x)=\omega x\tanh(\omega t),\ \omega=2\sqrt{\mu b},$$

we have

$$\begin{aligned}u(t,x)-F(t,x)&=4\mu\frac{\displaystyle\int_{\mathbf{R}}\frac{x-y\cosh(\omega t)}{\sqrt{\mu/b}\sinh(2\omega t)}\exp\left\{-\frac{[x-y\cosh(\omega t)]^2}{\sqrt{\mu/b}\sinh(2\omega t)}\right\}h_0(y)dy}{\displaystyle\int_{\mathbf{R}}\exp\left\{-\frac{[x-y\cosh(\omega t)]^2}{\sqrt{\mu/b}\sinh(2\omega t)}\right\}h_0(y)dy}\\&=4\mu\frac{\displaystyle\int_{\mathbf{R}}\frac{x-z}{\sqrt{\mu/b}\sinh(2\omega t)}\exp\left\{-\frac{[x-z]^2}{\sqrt{\mu/b}\sinh(2\omega t)}\right\}h_0\left(\frac{z}{\cosh(\omega t)}\right)\frac{dz}{\cosh(\omega t)}}{\displaystyle\int_{\mathbf{R}}\exp\left\{-\frac{[x-z]^2}{\sqrt{\mu/b}\sinh(2\omega t)}\right\}h_0\left(\frac{z}{\cosh(\omega t)}\right)\frac{dz}{\cosh(\omega t)}}.\end{aligned} \tag{8.79}$$

The scaling

$$\begin{aligned}t&\longrightarrow\frac{1}{\omega}\ln t\\x&\longrightarrow\frac{xt}{\rho(t)}\\\xi(x)&\longrightarrow\xi(\rho(t)x),\end{aligned}$$

for ρ, as before, satisfying condition **(H)**, then leads to the following expression

$$
\begin{aligned}
&u\left(\frac{1}{\omega}\ln t, \frac{xt}{\rho(t)}, \xi_\varphi\left(\rho(t)\cdot\right)\right) - F\left(\frac{1}{\omega}\ln t, \frac{xt}{\rho(t)}\right) \\
&\quad = 4\mu \frac{\displaystyle\int_{\mathbf{R}} \frac{\frac{xt}{\rho(t)} - z}{\sqrt{\mu/b}\sinh(2\ln t)} \exp\left\{-\frac{[\frac{xt}{\rho(t)} - z]^2}{\sqrt{\mu/b}\sinh(2\ln t)}\right\} h_0\left(\frac{z\rho(t)}{\cosh(\ln t)}\right)\frac{dz}{\cosh(\ln t)}}{\displaystyle\int_{\mathbf{R}} \exp\left\{-\frac{[\frac{xt}{\rho(t)} - z]^2}{\sqrt{\mu/b}\sinh(2\ln t)}\right\} h_0\left(\frac{\rho(t)z}{\cosh(\ln t)}\right)\frac{dz}{\cosh(\ln t)}} \\
&= 4\mu \frac{\displaystyle\int_{\mathbf{R}} \frac{\frac{x}{\rho(t)} - y}{\sqrt{\mu/b}(1/t)\sinh(2\ln t)} \exp\left\{-\frac{[\frac{x}{\rho(t)} - y]^2}{\sqrt{\mu/b}(1/t)^2\sinh(2\ln t)}\right\} h_0\left(\frac{ty\rho(t)}{\cosh(\ln t)}\right)\frac{tdy}{\cosh(\ln t)}}{\displaystyle\int_{\mathbf{R}} \exp\left\{-\frac{[\frac{x}{\rho(t)} - y]^2}{\sqrt{\mu/b}(1/t)^2\sinh(2\ln t)}\right\} h_0\left(\frac{ty\rho(t)}{\cosh(\ln t)}\right)\frac{tdy}{\cosh(\ln t)}} \\
&= 4\mu \frac{\displaystyle\int_{\mathbf{R}} \frac{x-u}{\varphi(t)\sqrt{\mu/b}(1/t)\sinh(2\ln t)[\rho(t)]^2} \exp\left\{-\frac{[x-u]^2}{\sqrt{\mu/b}(1/t)^2\sinh(2\ln t)[\rho(t)]^2}\right\} G\left(\xi\left(\frac{u}{\varphi(t)}\right)\right)du}{\displaystyle\int_{\mathbf{R}} \frac{1}{\varphi(t)\rho(t)} \exp\left\{-\frac{[x-u]^2}{\sqrt{\mu/b}(1/t)^2\sinh(2\ln t)[\rho(t)]^2}\right\} G\left(\xi\left(\frac{u}{\varphi(t)}\right)\right)du} \\
&\quad = \frac{I(t,x,\xi_\varphi)}{J(t,x,\xi_\varphi)} = U(t,x,\xi_\varphi), \qquad (8.80)
\end{aligned}
$$

with

$$\varphi = \frac{\cosh(\ln t)}{t}.$$

Lemma 8.1. *The random field $J(t,x,\xi_\varphi)$ defined in equation (8.80) converges in probability to the constant C_0, when $t \longrightarrow \infty$.*

Theorem 8.10. *Assume that condition (8.75) holds. Then, when t goes to infinity,*

$$(i) \qquad \frac{\nu(t)}{\sqrt{w(t)}}\left[I(t,x,\xi_\varphi) - E[I(t,x,\xi_\varphi)]\right] \underset{\longrightarrow}{w} U(\xi_{\varphi_\infty}) \sim N(0,\sigma_2^2(G)) \qquad (8.81)$$

$$
\begin{aligned}
(ii) \qquad &\frac{\nu(t)}{\sqrt{w(t)}}\left[u\left(\frac{1}{\omega}\ln t, \frac{xt}{\rho(t)}, \xi_\varphi\left(\rho(t)\cdot\right)\right) - F\left(\frac{1}{\omega}\ln t, \frac{xt}{\rho(t)}\right)\right] \underset{\longrightarrow}{w} U(\xi_{\varphi_\infty}) \\
&\sim N\left(C_0, \frac{\sigma_2^2(G)}{C_0^2}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\sigma_2^2(G) &= \|h_u\|_2^2 \|S\|_1 \\
\nu(t) &= \varphi(t)\sinh(2\ln t)(1/t)[\rho(t)]^2 \\
w(t) &= \frac{\rho(t)\sqrt{\sinh(2\ln t)}}{t}, \qquad (8.82)
\end{aligned}
$$

with

$$h_u(y) = \frac{y}{\sqrt{\mu/b}} \exp\left\{ -\frac{y^2}{\sqrt{\mu/b}} \right\}.$$

8.5 Some Comments on the Introduced Results

In the above section, the limiting distributions of the solutions to the heat and Burgers equations, under weak-dependent scenarios, are derived from the application of [54] results. Specifically, the scaling proposed is formulated within the general setup established in [73], which can be applied for defining different versions of Central Limit Theorems in the context of Heat and Burgers equations. For example, in [72], the function ρ defining the scaling of the spatial component and random initial condition is the function $t/\ln t$. This choice of such a function allows the definition of a nice and simple class of admissible functions ψ and ν needed in the application of the mentioned [54] results. Alternative scaling procedures can be investigated, but the weak-convergence to the Gaussian distribution does not hold as it is can be seen applying different methodologies such as diagram formulae and [89] results, among others.

Finally, we remark other possible lines of investigation such as the ones related with limit results for other external potential functional forms like the exponential forms, that is, for

$$\Phi(x) = \frac{\lambda^2}{2} e^{2x}.$$

In this case, the fundamental solution $q(t, x, y)$ of the heat equation

$$\frac{\partial}{\partial t} h(t, x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} h(t, x) - \frac{\lambda^2}{2} e^{2x} h(t, x), \tag{8.83}$$

$$h(0, x) = h_0(x) \tag{8.84}$$

was obtained by [55] in the following form:

$$q(t, x, y) = \int_0^\infty \exp\left(-\frac{\lambda^2}{2} u - \frac{e^{2x} + e^{2y}}{2u} \right) \theta\left(\frac{e^{x+y}}{u}, t \right) \frac{du}{u}, \tag{8.85}$$

where

$$\theta(r, t) = \frac{r}{(2\pi^3 t)^{1/2}} \int_0^\infty e^{\frac{\pi^2 - z^2}{2t} - r\cosh z} \sinh z \sin \frac{\pi z}{t} dz.$$

Thus, the solution of the initial value problem (8.83)-(8.84) takes the form

$$h(t, x) = \int q(t, x, y) h_0(y) dy =$$

$$= \int_{-\infty}^{\infty} \left(\int_0^{\infty} \exp\left(-\frac{\lambda^2}{2} u - \frac{e^{2x} + e^{2y}}{2u} \right) \left[\frac{\frac{e^{x+y}}{u}}{(2\pi^3 t)^{1/2}} \right.\right.$$
$$\left.\left. \times \int_0^{\infty} e^{\frac{\pi^2 - z^2}{2t} - \frac{e^{x+y}}{u} \cosh z} \sinh z \sin \frac{\pi z}{t} dz \right] \frac{du}{u} \right) h_0(y) dy.$$

The solution of the Burgers equation with exponential external potential

$$\frac{\partial}{\partial t} u(t,x) + u(t,x) \frac{\partial}{\partial x} u(t,x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} u(t,x) + \lambda^2 e^{2x},$$

subject to the initial condition

$$u(0,x) = u_0(x), \quad U(x) = \int_{-\infty}^{x} u_0(\xi) d\xi \tag{8.86}$$

is then given by

$$u(t,x) = \frac{\int_R \frac{\partial}{\partial x} q(t,x,y) e^{-\frac{U(y)}{2\mu}} dy}{\int_R q(t,x,y) e^{-\frac{U(y)}{2\mu}} dy}, \tag{8.87}$$

where

$$\frac{\partial}{\partial x} q(t,x,y) = \int_0^{\infty} \int_0^{\infty} F(u,z) dz du,$$

and

$$F(u,z) = \frac{d}{dx} [\exp\left(-\frac{\lambda^2}{2} u - \frac{e^{2x} + e^{2y}}{2u} \right) \frac{\frac{e^{x+y}}{u}}{(2\pi^3 t)^{1/2}} e^{\frac{\pi^2 - z^2}{2t} - \frac{e^{x+y}}{u} \cosh z} \sinh z \sin \frac{\pi z}{t} \frac{1}{u}] =$$

$$= \left[\frac{1}{u} (\sinh z) \frac{e^{x+y}}{\sqrt{2\pi^3 t}} \left(\sin \frac{\pi}{tu} z \right) \exp\left(\frac{1}{2t} \left(\pi^2 - z^2 \right) - \frac{1}{u} (\cosh z) e^{x+y} \right) \right.$$
$$\times \exp\left(-\frac{1}{2} u\lambda^2 - \frac{1}{2u} \left(e^{2x} + e^{2y} \right) \right)$$
$$- \frac{1}{u^2} (\sinh z) e^{x+y} \frac{e^{2x}}{\sqrt{2\pi^3 t}} \left(\sin \frac{\pi}{tu} z \right) \exp\left(\frac{1}{2t} \left(\pi^2 - z^2 \right) - \frac{1}{u} (\cosh z) e^{x+y} \right)$$
$$\times \exp\left(-\frac{1}{2} u\lambda^2 - \frac{1}{2u} \left(e^{2x} + e^{2y} \right) \right)$$
$$- \frac{1}{u^2} (\cosh z \sinh z) \frac{e^{2(x+y)}}{\sqrt{2\pi^3 t}} \left(\sin \frac{\pi}{tu} z \right) \exp\left(\frac{1}{2t} \left(\pi^2 - z^2 \right) - \frac{1}{u} (\cosh z) e^{x+y} \right)$$
$$\left. \times \exp\left(-\frac{1}{2} u\lambda^2 - \frac{1}{2u} \left(e^{2x} + e^{2y} \right) \right) \right]. \tag{8.88}$$

In this context, the Laplace transform can be useful to obtain suitable scalings for deriving limit results, since the Laplace transform of the fundamental solution is given by

$$\int_0^\infty e^{-\alpha^2 t/2} q(t,x,y) = 2I_0(\lambda e^x) K_\alpha(\lambda e^y), x < y,$$

where $I_0(z)$ and $K_\alpha(w)$ are the usual modified Bessel functions, for which we have

$$I_\alpha(a)K_\alpha(b) = \frac{1}{2}\int_0^\infty \exp\left(-\frac{1}{2}u - \frac{a^2+b^2}{2u}\right) I_\alpha\left(\frac{ab}{u}\right)\frac{1}{u}du.$$

8.6 Fractional Differential Equations Driven by Random Innovations

In this section, we review some results on mean quadratic local properties of the solution to fractional pseudodifferential equations driven by random innovations. The temporal and spatial cases are first studied. In the spatial case, pseudodifferential problems are defined on a bounded domain of $\mathbf{R}^n$. Specifically, $C^\infty-$ bounded domains and fractal domains are considered. An overview of some recent results on fractional pseudodifferential evolution equations driven by random innovations is secondly provided for the introduction of new classes of fractal and strong-dependence spatiotemporal random fields.

8.6.1 Temporal Fractional Pseudodifferential Models

Several extended formulations of [43] equation, as well as the weak-sense definition of their solutions are studied in the fractional pseudodifferential framework in [60]. Specifically, the following family of temporal pseudodifferential models is considered

$$(D+\alpha)^\nu X_{-\nu}(t) = \varepsilon(t), \quad t \in \mathbf{R}, \tag{8.89}$$

in terms of the operator $D = D_t = \frac{d}{dt}$, $\nu > 0$, $\alpha \in \mathbf{R}$, where $\{\varepsilon(t),\ t \in \mathbf{R}\}$ is a zero-mean white noise. The local regularity properties of the sample-path of the strong-sense solution of the above equation are characterized, in the Gaussian case, in the following result from [60].

Proposition 8.2. *For $\nu > 1/2$, the solution $\mathscr{X}_{-\nu}$ to equation (8.89) is Hölder continuous, in the mean-square sense, of order $\nu - 1/2$, i.e., the following inequality holds for certain positive constant C,*

$$E[X_{-\nu}(x+h) - X_{-\nu}(x)]^2 \leq C|h|^{2\nu-1}, \quad h \longrightarrow 0.$$

In the Gaussian case, the sample paths of $\mathscr{X}_{-\nu}$ have modulus of continuity $\Lambda(\delta)$ of order $\delta^{\nu-1/2}(|\ln\delta|)^{-1}$, and satisfy with probability one

$$\dim_H\left(\text{image}(X_{-\nu}(\cdot))\right) = 1 \wedge \left(\frac{1}{\nu - 1/2}\right)$$
$$\dim_H\left(\text{graph}(X_{-\nu}(\cdot))\right) = \left(\frac{1}{\nu - 1/2}\right) \wedge (2 - (\nu - 1/2)). \tag{8.90}$$

In the spatial case, the fractional elliptic pseudodifferential model

$$\left(\alpha^2 I - \Delta\right)^{\nu/2} X_\nu(x) = \varepsilon(x), \quad x \in \mathbf{R}^n,\ n \geq 2,\ \nu > 0, \tag{8.91}$$

provides the extension to $\mathbf{R}^n$ of the stochastic Laplace or stochastic Helmholtz equation of [43]. Here, operator $\left((-\Delta) + \alpha^2\right)^{\nu/2}$ is defined by

$$\left((-\Delta) + \alpha^2\right)^{\nu/2} = (-1)^{\nu/2} \sum_{j=0}^{\infty} \binom{\nu/2}{j} (-\Delta)^j \alpha^{2(\nu/2-j)}.$$

The homogeneous isotropic solution X_ν to this equation has spectral density

$$f(\lambda) = \frac{\sigma^2}{(2\pi)^d} \frac{1}{(|\lambda|^2 + \alpha^2)^\nu}, \quad \lambda \in \mathbf{R}^n. \tag{8.92}$$

The mean-quadratic local variation properties of X_ν are now derived in the following result.

Proposition 8.3. *The solution X_ν to equation (8.91) satisfies the following inequality:*

$$E[X_\nu(x+h) - X_\nu(x)]^2 \leq C\,|h|^{2\nu-n}, \quad h \longrightarrow 0, \tag{8.93}$$

for certain positive constant C.

Proof. The proof follows from the asymptotic order 2ν of the spectral density (8.92), which means that the functions in the Reproducing Kernel Hilbert space (RKHS) of X_ν admit weak-sense fractional derivatives up to order ν, i.e., from classical Embedding Theorems of fractional Sobolev spaces into Hölder-Zygmund spaces (see, for instance, [112]), this assertion is equivalent to the fact that they are Hölder continuous of fractional order $\nu - n/2$. That is, random field X_ν is Hölder continuous of the same order, in the mean-square sense. Hence, equation (8.93) holds.

Remark 8.4. An interesting covariance function family in relation to the above introduced spatial fractional pseudodifferential model class is the following (see [119])

$$B(r) = (2\pi)^n \int_0^\infty (ru)^{(2-n)/2} J_{(n-2)/2}(ru) u^{n-1} \frac{\sigma^2}{(2\pi)^n} \frac{1}{(u^2+\alpha^2)^\nu} du, \quad r > 0,$$
$$\nu > n/2,$$

which belongs to Matérn class (see [105], pp. 49-51), and takes the form

$$B(r) = \frac{\pi^{n/2}}{\alpha^{2\nu-n} 2^{\nu-n/2-1} \Gamma(\nu)} \frac{\sigma^2}{(2\pi)^n} K_{\nu-n/2}(r\alpha)(r\alpha)^{\nu-n/2}, \quad r > 0,$$
$$\alpha > 0,\ \nu > n/2,$$

where K_λ is the Macdonald function with index $\lambda \in \mathbf{R}$, and J_μ is the usual Bessel function of order μ.

8.6.2 *Elliptic Fractional Pseudodifferential Models on Regular and Fractal Domains*

Spatial fractional elliptic pseudodifferential models have been introduced, for example, in [96], [97]), and in [99], in the generalized random field framework. In this section, we establish some results related to the strong-sense covariance factorization, and the local variation properties of the solution to extended formulations of the above models, in terms of fractal Gaussian random innovations.

In a general abstract setting, we can consider the spatial fractional pseudodifferential model

$$\mathscr{L}_s X(x) = \varepsilon(x), \quad x \in D \subset \mathbf{R}^n, \tag{8.94}$$

where $\mathscr{L}$ is an elliptic fractional pseudodifferential operator of order $s > 0$, and ε is a zero-mean second-order innovation random field with covariance function $B_\varepsilon(x, y) = E[\varepsilon(x)\varepsilon(y)]$. The following result provides the covariance factorization and local regularity properties of the strong-sense solution X. In the Gaussian case, the sample-path properties are also characterized.

Proposition 8.4. *Assume that the RKHS of ε is isomorphic to the fractional Sobolev space $\bar{H}^\beta(D)$ of order β, for certain $\beta > 0$, constituted by functions with compact support contained in D, where D is a compact C^∞ domain. Then, the following assertions hold: For $\beta + s > n/2$,*

(1) The solution X admits the covariance factorization

$$B_X(x, y) = E[X(x)X(y)] = \int_{D^3} l_s(x, z) t_\varepsilon(z, v) t_\varepsilon(v, u) l_s(u, y) dz dv du, \tag{8.95}$$

where l_s is the kernel of the inverse operator $\mathscr{L}_s^{-1}$ of $\mathscr{L}_s$, i.e.,

$$\mathscr{L}_s^{-1}(f)(x) = \int_D l_s(x,y) f(y) dy, \quad \forall f \in \mathscr{D}(\mathscr{L}_s^{-1}),$$

with $\mathscr{D}(\mathscr{L}_s^{-1})$ denoting the domain of operator $\mathscr{L}_s^{-1}$. Here, t_ε is the kernel of operator $\mathscr{T}_\varepsilon$ satisfying that $R_\varepsilon = \mathscr{T}_\varepsilon \mathscr{T}_\varepsilon^*$, where R_ε is the covariance operator of ε with kernel its covariance function B_ε.

(2) *Random field X satisfies*

$$E[X(x+h) - X(x)]^2 \leq C|h|^{2(\beta+s)-n},$$

for certain positive constant C.

(3) *In the Gaussian case, random field X has Hölder continuous sample paths of order $\beta + s - n/2$.*

Remark 8.5. Note that the above covariance factorization of the Gaussian innovation process ε holds, in terms of isomorphisms between $L^2(D)$ and $\bar{H}^\beta(D)$, and between $H^{-\beta}(D)$ and $L^2(D)$, under the assumption that the RKHS of ε is isomorphic to the space $\bar{H}^\beta(D)$ (see [96]).

Proof. (i) The proof follows from the formal definition of random field X as

$$X = \mathscr{L}_s^{-1}\varepsilon,$$

and the assumption of equivalence of norms between the RKHS of ε and $\bar{H}^\beta(D)$. Specifically, from this assumption, and since $\mathscr{L}_s$ is elliptic and bounded, the covariance operator

$$R_X = \mathscr{L}_s^{-1} \mathscr{T} \mathscr{T}^* [\mathscr{L}_s^{-1}]^*$$

of X defines an isomorphism between the fractional Sobolev spaces $H^{-(\beta+s)}(D)$ and $\bar{H}^{\beta+s}(D)$. Therefore, X admits the covariance factorization (8.95), since its covariance operator is in the trace class for $s + \beta > n/2$, as follows from the isomorphic relationship between the RKHS of X and the fractional Sobolev $\bar{H}^{s+\beta}(D)$, as well as from the asymptotic order of its eigenvalues

$$|\lambda_k(R_X)| \leq C k^{-2(s+\beta)/n}, \quad k \in \mathbf{N},$$

obtained from the following entropy number inequality of the embedding $H^{-(s+\beta)}(D) \longrightarrow H^{s+\beta}(D)$,

$$e_k\left(id : H^{-(s+\beta)}(D) \longrightarrow H^{s+\beta}(D)\right) \leq C k^{-2(s+\beta)/n}, \quad k \in \mathbf{N},$$

and from

$$|\lambda_k(id)| \leq \sqrt{2} e_k(id), \quad k \in \mathbf{N}$$

(see, for example, [112]).

(ii) Random field X is Hölder continuous of order $\beta+s-n/2$, from the isomorphic identification of the RKHS of X with the fractional Sobolev space $\tilde{H}^{s+\beta}(D)$, as well as from the embedding $\tilde{H}^{s+\beta}(D) \hookrightarrow \mathscr{H}^{s+\beta-n/2}(D)$, with $\mathscr{H}^{s+\beta-n/2}(D)$ being the space of Hölder continuous functions of order $s+\beta-n/2$ on D.
(iii) The assertion follows from [1]'s results on $\alpha-$index random fields. Specifically, random field X is from (ii) a $\beta+s-n/2-$index random field, that is, it satisfies, for any $\epsilon>0$,

$$\sup_{|h|<\delta} |X(x+h)-X(x)| \leq Y\delta^{\beta+s-n/2-\epsilon}, \quad \delta \to 0, \quad \forall x \in \mathbf{R}^n,$$

where Y is an almost surely finite random variable.

Consider now model (8.94) in the case where $D=\Gamma$, with Γ being a fractal compact $d-$set as given in the following definition (see [113]):

Definition 8.1. Let Γ be a set in $\mathbf{R}^n$. Then, Γ is called a $d-$set, with $0 \leq d \leq n$, if there exists a Borel measure μ_Γ in $\mathbf{R}^n$ with the following two properties:

(i) Supp $\mu_\Gamma = \Gamma$
(ii) there are two constants $c_1>0$, and $c_2>0$, such that, for all $\gamma \in \Gamma$, and all r with $0<r<1$,

$$c_1 r^d \leq \mu_\Gamma\left(B(\gamma,r) \cap \Gamma\right) \leq c_2 r^d,$$

where $B(\gamma,r)$ denotes the closed ball in $\mathbf{R}^n$ centered at γ and of radius r.

Remark 8.6. Operator $\mathscr{L}_s$, in the $d-$set context, is defined as a fractal pseudodifferential operator, with inverse having pure point spectrum constituted by eigenvalues $\{\lambda_k(\mathscr{L}_s^{-1}),\ k \in \mathbf{N}\}$, that satisfy, for certain positive constant C,

$$|\lambda_k(\mathscr{L}_s^{-1})| \leq C k^{-s/d}, \quad k \in \mathbf{N}.$$

(For a more general treatment in the stochastic sense see also [35]).

Proposition 8.5. *Assume that the RKHS of ε is isomorphic to the fractional Sobolev space $\tilde{H}^{\alpha}(\Gamma)$ of order α, for certain $\alpha>0$, constituted by functions which coincide with the trace on $d-$set Γ of distributions in the space $H^{\alpha+\frac{n-d}{2}}(\mathbf{R}^n)$ (see [113]). Then, the following assertions hold: For $s+\alpha>d/2$,*

(i) The solution X admits the covariance factorization

$$B_X(x,y) = E[X(x)X(y)] = \int_{\Gamma^3} l_s(x,z)t_\varepsilon(z,v)t_\varepsilon(v,u)l_s(u,y)\mu_\Gamma(dz) \times \mu_\Gamma(dv)\mu_\Gamma(du), \tag{8.96}$$

where l_s is the kernel of the inverse operator $\mathscr{L}_s^{-1}$ of $\mathscr{L}_s$, i.e.,

$$\mathscr{L}_s^{-1}(f)(x) = \int_\Gamma l_s(x,y) f(y) \mu_\Gamma(dy), \quad \forall f \in \mathscr{D}(\mathscr{L}_s^{-1}),$$

with $\mathscr{D}(\mathscr{L}_s^{-1})$ denoting the domain of the inverse $\mathscr{L}_s^{-1}$ of the fractal pseudodifferential operator $\mathscr{L}_s$, which under the assumptions made coincides with a trace space on Γ. Here, t_ε is the kernel of operator $\mathscr{T}_\varepsilon$ satisfying that $R_\varepsilon = \mathscr{T}_\varepsilon \mathscr{T}_\varepsilon^*$, where R_ε is the covariance operator of ε with kernel its covariance function B_ε.

(ii) Random field X satisfies

$$E[X(x+h) - X(x)]^2 \leq C\,|h|^{2(s+\alpha)-d},$$

for certain positive constant C.

(iii) In the Gaussian case, random field X has Hölder continuous sample paths of order $s + \alpha - d/2$

Proof. The proof of (i)-(iii) follows in a similar way to the proof of Proposition 8.4, applying the Trace and Fractal Embedding results given, for example, in [113] for fractional Sobolev spaces on compact $d-$sets. Also Remark 8.6 provides the spectral properties of $\mathscr{L}_s$ as a fractal pseudodifferential operator on Γ. These properties are also applied in the derivation of (i–iii) as in the proof of previous Proposition 8.4.

8.7 Fractional Pseudodifferential Evolution Equations

Some results on the weak-sense definition of the random solutions to fractional pseudodifferential evolution equations of the type studied in [60] are now reviewed. Also, the mean quadratic local variation properties of the corresponding temporal and spatial increment processes are derived in Sect. 8.7.1.

Let $\{X(t,x),\ t > 0,\ x \in \mathbf{R}\}$ be the solution of the following fully-fledged stochastic partial differential equation (see Theorem 8.11 and Remark 8.7 below)

$$\left[\frac{\partial}{\partial t} + \gamma - c^2 \frac{\partial^2}{\partial x^2}\right]^\nu X(t,x) = \varepsilon(t,x), \tag{8.97}$$

$$t > 0, \quad x \in \mathbf{R}, \quad c \geq 0, \quad \nu > 0, \quad \gamma > 0, \quad c \in \mathbf{R},$$

where $\varepsilon = \{\varepsilon(t,x),\ t > 0,\ x \in \mathbf{R}\}$ is a white noise random field both in time and in space, i.e., ε is a (generalized) zero-mean random field with covariance function

$$\mathbf{E}[\varepsilon(f)\varepsilon(g)] = \sigma^2 \langle f, g \rangle_{L^2(\mathbf{R}_+ \times \mathbf{R})}, \quad \sigma^2 > 0,$$

for $f, g \in L^2(\mathbf{R}_+ \times \mathbf{R})$, the space of square-integrable function on $\mathbf{R}_+ \times \mathbf{R}$.

Heuristically, the second-order density of the stationary both in time and space m.s. solution to the equation (8.97) takes the form

$$f_{\gamma,\nu}(\mu,\lambda)=\frac{\sigma^2}{(2\pi)^2}\frac{1}{(\mu^2+(\gamma+c^2\lambda^2)^2)^{\nu}},\quad \mu,\lambda\in\mathbf{R},\quad \gamma>0,\quad \nu>1, \tag{8.98}$$

where μ and λ are, respectively, conjugates to t and x.

Note that for $\nu=1$ the spectral density (8.98) was obtained by [114]. Moreover, in [33], p. 225, a set of covariances in $(0,\infty)\times\mathbf{R}^d$ having the spatiotemporal spectral density

$$f(\mu,\lambda)=\frac{\sigma_1^2}{\left[c_1^2\mu^2+(a^2+|\lambda|^2)^{2p}\right]},\quad \mu\in\mathbf{R},\ \lambda\in\mathbf{R}^n,\ n\geq 1,$$

is introduced in a geostatistical context. In the case $p=1,d=1$ and $c_1^2=1$, this formula reduces to spectral density (8.98) with $\nu=1$, $\sigma_1^2=\frac{\sigma^2}{(2\pi)^2}$, $c_1^2=1$ and $a^2=\gamma$.

For $\gamma>0$, $\nu>1$, the random field with spectral density (8.98) exhibits a short-range dependence both in the time and space, that is,

$$\lim_{\max\{\lambda,\mu\}\to 0} f_{\gamma,\nu}(\mu,\lambda)=\frac{\sigma^2}{(2\pi)^2}\frac{1}{\gamma^{2\nu}}>0$$

and in time or in space separately, that is, for a fix $\lambda\in\mathbf{R}$,

$$\lim_{\mu\to 0} f_{\gamma,\nu}(\mu,\lambda)=\frac{\sigma^2}{(2\pi)^2}\frac{1}{(\gamma+c^2\lambda^2)^{2\nu}}>0,$$

or, for a fixed $\mu>0$,

$$\lim_{\lambda\to 0} f_{\gamma,\nu}(\mu,\lambda)=\frac{\sigma^2}{(2\pi)^2}\frac{1}{(\mu^2+\gamma^2)^{\nu}}>0.$$

However, the random field with spectral density (8.98) displays the following fractal behavior at infinity

$$f_{\gamma,\nu}(\mu,\lambda)=O\left((\min\left\{\mu,\lambda^2\right\})^{-2\nu}\right)$$

as $\min\{\mu,\lambda\}\to\infty$; and in time or in space separately, that is, for a fixed $\lambda\in\mathbf{R}$,

$$f_{\gamma,\nu}(\mu,\lambda)=O\left(\mu^{-2\nu}\right),\quad \mu\to\infty,$$

or, for a fixed $\mu\in\mathbf{R}$,

$$f_{\gamma,\nu}(\mu,\lambda)=O\left(\lambda^{-4\nu}\right),\quad \lambda\to\infty.$$

This random field also displays long-range dependence, both in time and in space, in the particular case $\gamma = 0$ as

$$\lim_{\max\{\lambda,\mu\}\to 0} f_{0,\gamma}(\mu,\lambda) = \infty.$$

However, for $\gamma = 0$, $0 < \nu < 1$, the spectral density in (8.98) does not correspond to a second-order stationary random field, but rather to a self-similar random field with homogeneous increments. In particular, the second-order moments of the random field $X(ta, x\sqrt{a})$ coincide with that of $a^{\nu-\frac{3}{4}}X(t,x)$ for every $a > 0$ and $\nu \in \left(\frac{3}{4}, \frac{7}{4}\right)$.

If $\gamma = 0$ and $\nu \in \left(0, \frac{1}{4}\right)$, the random field

$$v(t,x) = \int_t^{t+1}\int_x^{x+1} X(s,y)dsdy, \quad t > 0,\ x \in \mathbf{R},$$

has spectral density

$$h(\mu,\lambda) = \frac{\sigma^2}{(2\pi)^2}\frac{\sin^2\frac{\mu}{2}}{(\frac{\mu}{2})^2}\frac{\sin^2\frac{\lambda}{2}}{(\frac{\lambda}{2})^2}\frac{1}{(\mu^2+c^4\lambda^4)^\nu}, \quad \mu,\lambda \in \mathbf{R}.$$

So, it is stationary , and displays long-range dependence both in time and in space, that is

$$h(\mu,\lambda) \sim \frac{\sigma^2}{(2\pi)^2}\frac{1}{(\mu^2+c^4\lambda^4)^\nu}$$

as $\max\{\lambda,\mu\} \to 0$, and

$$h(\mu,\lambda) = O\left(\frac{1}{\mu^2\lambda^2(\mu^2+c^4\lambda^4)^\nu}\right)$$

as $\min\{\lambda,\mu\} \to \infty$.

In order to make rigorous the heuristically derivation of (8.98) from (8.97), we will use again the theory of GRFs on fractional Sobolev spaces.

Operator H_ν is a function of the elliptic self-adjoint differential operator $(\mathscr{L}_t, \mathscr{L}_x) = \left(-i\frac{\partial}{\partial t}, -i\frac{\partial}{\partial x}\right)$, densely defined on the separable Hilbert space $\left(L^2(\mathbf{R}_+\times\mathbf{R}), \mathbf{C}\right)$. Specifically,

$$H_\nu = \left(i\mathscr{L}_t + \gamma + c^2\mathscr{L}_x^2\right)^\nu. \tag{8.99}$$

Hence, H_ν admits the spectral representation, for each $\varphi \in \mathscr{D}(H_\nu)$,

$$\begin{aligned} H_\nu(\varphi)(g) &= \int_{\Lambda_t\times\Lambda_x}\left[i\mu+\gamma+c^2\lambda^2\right]^\nu d\left(E_{(\mu,\lambda)}(\varphi),g\right) \\ &= \left(\frac{\partial}{\partial t}+\gamma-c^2\frac{\partial^2}{\partial x^2}\right)^\nu(\varphi)(g), \quad \forall g \in L^2(\mathbf{R}_+\times\mathbf{R}), \end{aligned} \tag{8.100}$$

where $\Lambda_t \times \Lambda_x$ stands for the continuous spectrum of operator $(\mathscr{L}_t, \mathscr{L}_x)$, and $\{E_{(\mu,\lambda)} : (\mu,\lambda) \in \Lambda_t \times \Lambda_x\}$ denotes its spectral family.

The formal adjoint H_ν^* of H_ν is then represented as

$$\begin{aligned} H_\nu^*(\varphi)(g) &= \int_{\Lambda_t \times \Lambda_x} \left[-i\mu + \gamma + c^2\lambda^2\right]^\nu d\left(E_{(\mu,\lambda)}(\varphi), g\right) \\ &= \left(-i\mathscr{L}_t + \gamma + c^2\mathscr{L}_x^2\right)^\nu (\varphi)(g) \\ &= \left(-\frac{\partial}{\partial t} + \gamma - c^2\frac{\partial^2}{\partial x^2}\right)^\nu (\varphi)(g), \quad \forall g \in L^2(\mathbf{R}_+ \times \mathbf{R}), \quad (8.101) \end{aligned}$$

for each $\varphi \in \mathscr{D}(H_\nu^*) = \mathscr{D}(H_\nu)$, where we have used the fact that $-i\mathscr{L}_t = -\frac{\partial}{\partial t}$ is the formal adjoint of $i\mathscr{L}_t = \frac{\partial}{\partial t}$.

In terms of the spectral representations of H_ν and H_ν^*, their common domain $\mathscr{D}(H_\nu) = \mathscr{D}(H_\nu^*)$ is defined as

$$\mathscr{D}(H_\nu) = \left\{ f \in L^2(\mathbf{R}_+ \times \mathbf{R}) : \int_{\Lambda_t \times \Lambda_x} \left(\mu^2 + (\gamma + c^2\lambda^2)^2\right)^\nu d\left(E_{(\mu,\lambda)}(f), f\right) < \infty \right\}.$$

Note that, for each $\varphi \in \mathscr{D}(H_\nu^*) = \mathscr{D}(H_\nu)$,

$$H_\nu H_\nu^*(\varphi)(\psi) = \int_{\Lambda_t \times \Lambda_x} \left(\mu^2 + (\gamma + c^2\lambda^2)^2\right)^\nu d\left(E_{(\mu,\lambda)}(\varphi), \psi\right), \quad \forall \psi \in \mathscr{D}(H_\nu^*).$$

The inverse operator H_ν^{-1} of H_ν admits a similar spectral representation: for each $f \in \mathscr{D}(H_\nu^{-1})$,

$$\begin{aligned} H_\nu^{-1}(f)(g) &= \int_{\Lambda_t \times \Lambda_x} \left(i\mu + \gamma + c^2\lambda^2\right)^{-\nu} d\left(E_{(\mu,\lambda)}(f), g\right) \\ &= \left(i\mathscr{L}_t + \gamma + c^2\mathscr{L}_x^2\right)^{-\nu} (f)(g), \quad \forall g \in L^2(\mathbf{R}_+ \times \mathbf{R}). \quad (8.102) \end{aligned}$$

As the operator $(\mathscr{L}_t, \mathscr{L}_x)$ satisfies conditions given in [90] (pp. 145–148), the projection operators $E_{(\mu,\lambda)}$, $(\mu,\lambda) \in \Lambda_t \times \Lambda_x$, defining its spectral family, admit an integral representation in terms of a kernel given by

$$E_{(\mu,\lambda)}(t,s;x,y;\mu,\lambda) = \int_{-\infty}^{\lambda}\int_{-\infty}^{\mu} \Phi(t,s;x,y;\xi,\omega) d\rho(\xi,\omega),$$

where $\Phi(t,s;x,y;\xi,\omega) = (2\pi)^{-2}\exp\left(i\left\langle (t-s, x-y), (\xi,\omega)\right\rangle\right)$ represents the spectral kernel and $d\rho(\xi,\omega) = d\xi d\omega$ is the spectral measure of operator $(\mathscr{L}_t, \mathscr{L}_x)$. The above spectral representations of operators $H_\nu H_\nu^*$ and H_ν^{-1} can then be expressed in terms of Φ and ρ.

Theorem 8.11. *Let $X_{-\nu}$ be a GRF defined as*

$$X_{-\nu}(h) \underset{\text{m.s.}}{=} \int_{\mathbf{R}_+\times\mathbf{R}} \int_{\mathbf{R}_+\times\mathbf{R}} \bar{h}(t,x) l(t,x;s,y) d\varepsilon(s,y) dx dt,$$
$$\underset{\text{m.s.}}{=} \int_{\mathbf{R}_+\times\mathbf{R}} \bar{h}(t,x) \mathscr{X}_{-\nu}(t,x) dx dt, \tag{8.103}$$

for all $h \in L^2(\mathbf{R}_+ \times \mathbf{R})$, where ε is the generalized white noise as in (8.97), and

$$l(t,x;s,y) = (2\pi)^{-2} \int_{\Lambda_t\times\Lambda_x} \exp\left(i \left\langle (t-s, x-y)\,,\ (\mu,\lambda)\right\rangle\right) \left(i\mu + \gamma + c^2\lambda^2\right)^{-\nu} d\lambda d\mu. \tag{8.104}$$

Then, for $\nu \le 2/3$, $X_{-\nu}$ defines, in the weak-sense, a m.s. solution $\mathscr{X}_{-\nu}$ to equation (8.97), and for $\nu > 2/3$, $X_{-\nu}$ defines, in the strong-sense, the unique mean-square continuous solution $\mathscr{X}_{-\nu}$ to equation (8.97).

The GRF $X_{-\nu}$ has RKHS $\mathscr{H}(X_{-\nu}) =$

$$\left\{ \phi \in L^2(\mathbf{R}_+ \times \mathbf{R}) : \int_{\Lambda_t\times\Lambda_x} \left(\mu^2 + (\gamma + c^2\lambda^2)^2\right)^{\nu} \left|\hat{\phi}(\mu,\lambda)\right|^2 d\lambda d\mu < \infty \right\}, \tag{8.105}$$

with the inner product

$$\langle \phi, \varphi \rangle_{\mathscr{H}(X_{-\nu})} = \int_{\Lambda_t\times\Lambda_x} \left(\mu^2 + (\gamma + c^2\lambda^2)^2\right)^{\nu} \hat{\phi}(\mu,\lambda)\bar{\hat{\varphi}}(\mu,\lambda) d\lambda d\mu. \tag{8.106}$$

Remark 8.7. The mean-square fractional regularity order of the GRF $X_{-\nu}$ defined in equation (8.103), that is, the weak-sense regularity order of the functions of its RKHS, is $\frac{3\nu}{2} = \frac{\nu+2\nu}{2}$, with the fractional regularity order ν in time and the fractional regularity order 2ν in space ($\nu > 0$). Its minimum mean-square fractional singularity order, that is, the weak-sense regularity order of the test functions defining its domain, is then $-(3/2)\nu$. Although, in the formulation of Theorem 8.11, we consider the separable Hilbert space $L^2(\mathbf{R}_+ \times \mathbf{R})$. However, $X_{-\nu}$ can be defined on a larger function space according to its regularity order.

Proof. Random field $X_{-\nu}$ defined in equation (8.103) satisfies the following mean-square identity:

$$X_{-\nu}(h) \underset{\text{m.s}}{=} \int_{\mathbf{R}_+\times\mathbf{R}} \overline{\left[\int_{\mathbf{R}_+\times\mathbf{R}} \widetilde{l}(s,y;t,x) h(t,x) dx dt\right]} d\varepsilon(s,y)$$
$$\underset{\text{m.s}}{=} \varepsilon\left[\left(-\frac{\partial}{\partial t} + \gamma - c^2\frac{\partial^2}{\partial x^2}\right)^{-\nu} h\right], \quad \forall h \in L^2(\mathbf{R}_+ \times \mathbf{R}),$$

where

$$\widetilde{l}(t,x;s,y)=\int_{\Lambda_t\times\Lambda_x}\frac{\exp\left(i\left\langle(t-s,x-y)\,,\,(\mu,\lambda)\right\rangle\right)}{(2\pi)^2}\left(-i\mu+\gamma+c^2\lambda^2\right)^{-\nu}d\lambda d\mu,$$

is the kernel of the integral operator $\left(-\frac{\partial}{\partial t}+\gamma-c^2\frac{\partial^2}{\partial x^2}\right)^{-\nu}$.

Therefore,

$$\begin{aligned}X_{-\nu}\left(H_\nu^*\varphi\right)&\underset{\text{m.s}}{=}\varepsilon\left[\left(-\frac{\partial}{\partial t}+\gamma-c^2\frac{\partial^2}{\partial x^2}\right)^{-\nu}\left(-\frac{\partial}{\partial t}+\gamma-c^2\frac{\partial^2}{\partial x^2}\right)^{\nu}\varphi\right]\\&\underset{\text{m.s}}{=}\varepsilon(\varphi),\quad\forall\varphi\in\mathscr{D}(H_\nu^*).\end{aligned}$$

That is, $X_{-\nu}$ satisfies, in the m.s. sense, the generalized equation on $\mathscr{D}(H_\nu^*)$ associated with (8.97).

From Embedding Theorems between fractional Besov spaces, for $\nu>2/3$,

$$\mathscr{X}_{-\nu}(t,x)\underset{\text{m.s}}{=}\int_{\mathbf{R}_+\times\mathbf{R}}l(t,x;s,y)d\varepsilon(s,y)\tag{8.107}$$

defines the unique mean-square continuous ordinary solution to equation (8.97), with l defined as in equation (8.104). This implies that the spectral density of $\mathscr{X}_{-\nu}$ is given by (8.98).

The definition of the reproducing kernel Hilbert space $\mathscr{H}(X_{-\nu})$ of $X_{-\nu}$ is as in the theory of GRFs on fractional Sobolev spaces developed in [96], and motivated by the following definition of the dual random field $\widetilde{X}_{-\nu}$ of $X_{-\nu}$:

$$\widetilde{X}_{-\nu}(\psi)=\varepsilon\left[\left(\frac{\partial}{\partial t}+\gamma-c^2\frac{\partial^2}{\partial x^2}\right)^{\nu}\psi\right],\quad\forall\psi\in\mathscr{D}(H_\nu).$$

Hence,

$$\begin{aligned}\langle\phi,\varphi\rangle_{\mathscr{H}(X_{-\nu})}&=\left\langle\widetilde{X}_{-\nu}(\phi),\widetilde{X}_{-\nu}(\varphi)\right\rangle_{H(X_{-\nu})=H(\widetilde{X}_{-\nu})}\\&=\int_{\Lambda_t\times\Lambda_x}\left(i\mu+\gamma+c^2\lambda^2\right)^{\nu}\hat{\phi}(\mu,\lambda)\overline{\left(i\mu+\gamma+c^2\lambda^2\right)^{\nu}\hat{\varphi}(\mu,\lambda)}d\lambda d\mu\\&=\int_{\Lambda_t\times\Lambda_x}\left(\mu^2+(\gamma+c^2\lambda^2)^2\right)^{\nu}\hat{\phi}(\mu,\lambda)\bar{\hat{\phi}}(\mu,\lambda)d\lambda d\mu.\end{aligned}$$

The definition of $\widetilde{X}_{-\nu}$ guarantees the bicontinuity of the covariance operator $R_{X_{-\nu}}$ and the closeness of such an operator. Thus, the RKHS of $X_{-\nu}$ is a Hilbert (closed) space with the norm generated by the inner product (8.106), and it is a dense subspace of $L^2(\mathbf{R}_+\times\mathbf{R})$.

The extension of Theorem 8.11 to the $n-$multidimensional space case is straightforward using the relation

$$H_\nu = \left(i\mathscr{L}_t + \gamma + c^2\mathscr{L}_{\mathbf{x}}^2\right)^\nu,$$

with $\mathscr{L}_{\mathbf{x}}^2 = (-\Delta)$, and being Δ, as before, the Laplacian operator on $\mathbf{R}^n$. Hence, H_ν admits the spectral representation

$$H_\nu(\varphi)(g) = \int_{\Lambda_t\times\Lambda_{\mathbf{x}}} \left[i\mu + \gamma + c^2|\lambda|^2\right]^\nu d\left(E_{(\mu,\lambda)}(\varphi), g\right), \quad \forall g \in L^2(\mathbf{R}_+\times\mathbf{R}^n),$$

for each $\varphi \in \mathscr{D}(H_\nu)$, where $\Lambda_t \times \Lambda_{\mathbf{x}}$ stands for the continuous spectrum of the self-adjoint operator $(\mathscr{L}_t, \mathscr{L}_{\mathbf{x}}) = (-i\,\partial/\partial t, -i\,\partial/\partial x_1, \ldots, -i\,\partial/\partial x_n)$ on $L^2(\mathbf{R}_+ \times \mathbf{R}^n)$, and $\left\{E_{(\mu,\lambda)} : (\mu,\lambda) \in \Lambda_t \times \Lambda_{\mathbf{x}}\right\}$ is its spectral family.

In a similar manner one can view many useful spatiotemporal fractional pseudodifferential models as fractional versions of the heat equation.

Let $\mathscr{X}_{-1,-p}$ be the mean-square solution to

$$\left[c\frac{\partial}{\partial t} + \left((-\Delta) + a^2\right)^p\right]\mathscr{X}_{-1,-p} \underset{\text{m.s.}}{=} \varepsilon, \tag{8.108}$$

where Δ represents the Laplacian on $\mathbf{R}^n$, and ε spatio-temporal white noise with intensity σ. For $p > (n^2 + 2n - 1)/4n$, $\mathscr{X}_{-1,-p}$ is defined in the strong sense as

$$\mathscr{X}_{-1,-p}(t,x) \underset{\text{m.s.}}{=} \int_{T\times\mathbf{R}^n} l(t,x;s,y)d\varepsilon(s,y), \quad \forall (t,x) \in L^2(T\times\mathbf{R}^n),$$

where

$$l(t,x;s,y) = (2\pi)^{-n-1}\int_{\Lambda_t\times\Lambda_{\mathbf{x}}} \exp\left(i\,\langle(t-s,x-y),(\mu,\lambda)\rangle\right)\frac{1}{ic\mu + (|\lambda|^2+a^2)^p}d\lambda\,d\mu,$$

with associated spectral density

$$f_{1,p}(\mu,\lambda) = \frac{1}{c^2\mu^2 + (|\lambda|^2+a^2)^{2p}}$$

(cf. [33]).

For $p \le (n^2 + 2n - 1)/4n$, $\mathscr{X}_{-1,-p}$ is defined in terms of a generalized spatio-temporal random field

$$X_{-1,-p}(h) = \varepsilon\left(\left[-c\frac{\partial}{\partial t} + \left((-\Delta) + a^2\right)^p\right]^{-1} h\right), \quad \forall h \in L^2(T\times\mathbf{R}^n).$$

Note that the mean-square fractional regularity order of $X_{-1,-p}$ is $(1+2pn)/(n+1)$, with 1 representing the regularity order in time and $2p$ the regularity order in space.

Model (8.108) is included in a more general family

$$\left(\frac{\partial}{\partial t}+L_x^{\nu}\right)\mathscr{X}_{-1,-\nu} \underset{\text{m.s.}}{=} \varepsilon, \tag{8.109}$$

where L_x^{ν} is a spatial linear self-adjoint differential operator of order ν on $L^2(\mathbf{R}^n)$. In the case where $L_x^{\nu} = F_{\nu}(-\Delta)$, i.e. L_x^{ν} a function of $-\Delta$, with Δ denoting, as before, the Laplacian operator on $\mathbf{R}^n$, particularly, for F_{ν} given in terms of fractional powers of $-\Delta$, a fractional version of the heat equation emerges with m.s. solution given by

$$\mathscr{X}_{-1,-\nu}(t,x) \underset{\text{m.s.}}{=} \int_{\mathbf{R}_+\times\mathbf{R}^n} l(t,x;s,y)d\varepsilon(s,y),$$

for $(1+\nu n)/(n+1) > (n+1)/2$, where

$$l(t,x;s,y) = (2\pi)^{-n-1}\int_{\Lambda_t\times\Lambda_{\mathbf{x}}} \exp\left(i\,\langle (t-s,x-y),(\mu,\lambda)\rangle\right)\frac{1}{i\mu+F_{\nu}(|\lambda|^2)}d\lambda\,d\mu.$$

Its spectral density is then defined as

$$f_{1,\nu}(\mu,\lambda) = \frac{1}{\mu^2+\left(F_{\nu}(|\lambda|^2)\right)^2}.$$

In particular, model (8.108) corresponds to the case where $\nu = 2p$, and $F_{\nu}(-\Delta) = \left((-\Delta)+a^2\right)^p$. If $F_{\nu}(-\Delta) = (I-\Delta)^{\alpha/2}(-\Delta)^{\gamma/2}$ (i.e. $\nu = \alpha+\gamma$), one obtains the fractional heat equation considered in [3], where the local mean quadratic variation properties of the solution are studied.

8.7.1 Mean Quadratic Local Variation Properties

The semigroup-based formulation of the solution to equation (8.109), as well as its mean quadratic local variation properties are derived in the following result:

Proposition 8.6. *The solution to equation (8.109) admits the following representation:*

$$c\,(t,x) = \int_{\mathbf{R}^n} e^{i\,\langle x,\lambda\rangle}\int_0^t \exp\left\{-(t-s)\,\mathscr{P}_{\nu}(\lambda)\right\}\hat{\varepsilon}_s\,(d\lambda)\,ds, \quad (t,x)\in\mathbf{R}_+\times\mathbf{R}^n,$$

where $\mathscr{P}_{\nu}$ denotes the characteristic polynomial of operator L_x^{ν} (continuous function of the spectrum of $-i\frac{\partial}{\partial x}$), and ν its asymptotic order. Also, the following assertions hold:

(i) The limit behavior for the variance

$$\sigma_x^2(\tau) = E[c(t+\tau, x) - c(t, x)]^2.$$

of the time increments is given by

$$\lim_{|\tau|\to 0} \frac{\sigma_x(\tau)}{|\tau|^{\frac{1}{2}\left(1-\frac{n}{\nu}\right)}} = \textit{constant}.$$

Also, for $\nu > n$, the spatial increment process displays the following local asymptotic behavior:

$$\lim_{|z|\to 0} \frac{\left\{E\left[c(t, x+z) - c(t, x)\right]^2\right\}^{1/2}}{|z|^{\frac{\nu-n}{2}}} = \textit{constant}.$$

8.7.2 Geostatistical Fractional Pseudodifferential Models

Now, consider $\{Y(t, x),\ t > 0,\ x \in \mathbf{R}^n\}$, a spatio-temporal random field defined as the output of the following fractional differential filter applied to a spatio-temporal random field $\mathscr{X}_{(-\beta,-\nu)}$ with $\nu + 1 = \sum_{i=1}^n \nu_i$

$$\frac{\partial^{\nu+\beta+2}}{\partial x_1^{\nu_1} \partial x_2^{\nu_2} \ldots \partial x_n^{\nu_n} \partial t^{\beta+1}} \mathscr{X}_{(-\beta,-\nu)} = Y(t, x), \quad \beta > 0,\ \nu_i > 0,\ i = 1, \ldots, n. \tag{8.110}$$

Here, Y is assumed to be a zero-mean spatially homogeneous/temporally stationary field. Thus, Y belongs to the class of spatio-temporal random fields (S/TRF$_{\nu/\beta}$ models) considered in [33], and can be interpreted as a generalized spatio-temporal random field defined from $\mathscr{X}_{(-\beta,-\nu)}$, in terms of the test function family (see [33], p. 255)

$$q_{t,x}(s, y) = (2\pi)^{-(n+1)} \int_{\Lambda_t \times \Lambda_{\mathbf{x}}} (i\lambda_1)^{\nu_1} \ldots (i\lambda_n)^{\nu_n} (i\mu)^{(\beta+1)} \times \exp\left(i \left\langle (t-s,\ x-y), (\mu, \lambda) \right\rangle\right) d\lambda\, d\mu, \quad (t, x) \in \mathbf{R}_+ \times \mathbf{R}^n, \tag{8.111}$$

where $\Lambda_t \times \Lambda_{\mathbf{x}}$ is the continuous spectrum of operator

$$(\mathscr{L}_{x_1}, \ldots, \mathscr{L}_{x_n}, \mathscr{L}_t) = (-i\,\partial/\partial x_1, \ldots, -i\,\partial/\partial x_n, -i\,\partial/\partial t), \tag{8.112}$$

and $(2\pi)^{-(n+1)} \exp\left(i \left\langle (t-s,\ x-y), (\mu, \lambda) \right\rangle\right)$ is its spectral kernel. Note that operator

$$\delta^{\nu+\beta+2} = \frac{\partial^{\nu+\beta+2}}{\partial x_1^{\nu_1} \partial x_2^{\nu_2} \dots \partial x_n^{\nu_n} \partial t^{\beta+1}}$$

can be defined via the operator $(\mathscr{L}_{x_1}, \dots, \mathscr{L}_{x_n}, \mathscr{L}_t)$ in equation (8.112) as

$$\delta^{\nu+\beta+2} = (\mathscr{L}_{x_1})^{\nu_1} \dots (\mathscr{L}_{x_n})^{\nu_n} (\mathscr{L}_t)^{\beta+1},$$

with $\mathscr{D}(\delta^{\nu+\beta+2}) = H^{\nu+\beta+2}(\mathbf{R}_+ \times \mathbf{R}^n)$ dense in $L^2(\mathbf{R}_+ \times \mathbf{R}^n)$. The parametric family of (ν/β)-random field models defined above provides a useful tool in heterogeneity analysis in modern spatio-temporal geostatistics. This parametric family allows, by properly selecting the test functions q, to represent the degree of departure from homogeneity (parameter ν) and from stationarity (parameter β) (see [33]).

In the case where random field Y is a generalized white noise on $L^2(\mathbf{R}_+ \times \mathbf{R}^n)$, it can be proved similarly to Theorem 8.11 that $\mathscr{X}_{(-\beta,-\nu)}$ is well-defined in the mean-square sense as

$$\begin{aligned} &\int_{\mathbf{R}_+ \times \mathbf{R}^d} \bar{g}(t,x) \mathscr{X}_{(-\beta,-\nu)}(t,x) dx dt \\ &\underset{\text{m.s.}}{=} \int_{\mathbf{R}_+ \times \mathbf{R}^d} \bar{g}(t,x) \int_{\mathbf{R}_+ \times \mathbf{R}^d} l(t,x;s,y) d\varepsilon(s,y) dx\, dt, \end{aligned} \tag{8.113}$$

for $g \in L^2(\mathbf{R}_+ \times \mathbf{R}^n)$, where

$$\begin{aligned} l(t,x;s,y) = (2\pi)^{-(n+1)} \int_{\Lambda_t \times \Lambda_{\mathbf{x}}} &(i\lambda_1)^{-\nu_1} \dots (i\lambda_n)^{-\nu_n} (i\mu)^{-(\beta+1)} \\ &\exp\left(i \left\langle (t-s,\ x-y), (\mu,\lambda) \right\rangle\right) d\lambda\, d\mu. \end{aligned} \tag{8.114}$$

8.8 Multifractional Pseudodifferential Equations

The results previously described can be extended to the multifractional context considering the theory of fractional Besov, and, in particular, fractional Sobolev spaces of variable order as given in [78], [79], [80], [99], [100] provide, in the framework of generalized random fields, the covariance factorization and multifractional pseudodifferential representation of zero-mean second order random fields with RKHS isomorphic to a fractional Sobolev space of variable order, including the case of compact $d(\cdot)-$sets with variable local fractal dimension (see also [98]). In [94], the associated functional filtering and prediction problems are addressed in the functional regression context. Here, we study the heterogeneous mean-quadratic local variation properties of such random fields.

Let us first consider some preliminary definitions and results (see [99]):

Let δ and ρ be real numbers with $0 \leq \delta < \rho \leq 1$, σ be a real-valued function in $\mathscr{B}^{\infty}(\mathbf{R}^n)$, the set of all C^{∞} functions whose derivatives of each order are bounded. We say that a function $p(\mathbf{x}, \boldsymbol{\xi}) \in \mathscr{B}^{\infty}\left(\mathbf{R}^n_{\mathbf{x}} \times \mathbf{R}^n_{\boldsymbol{\xi}}\right)$ belongs to $\mathscr{S}^{\sigma}_{\rho,\delta}$ if and only if for any multi-indices α and β there exists some positive constant $C_{\alpha,\beta}$ such that

$$|D^{\alpha}_{\boldsymbol{\xi}} D^{\beta}_{\mathbf{x}} p(\mathbf{x}, \boldsymbol{\xi})| \leq C_{\alpha,\beta} \langle \boldsymbol{\xi} \rangle^{\sigma(\mathbf{x}) - \rho|\alpha| + \delta|\beta|}, \tag{8.115}$$

where $D^{\alpha}_{\boldsymbol{\xi}}$ and $D^{\beta}_{\mathbf{x}}$ refer to derivatives with respect to $\boldsymbol{\xi}$ and $\mathbf{x}$, respectively, and $\langle \boldsymbol{\xi} \rangle = (1 + |\boldsymbol{\xi}|^2)^{1/2}$. The following semi-norm is considered for the elements of $\mathscr{S}^{\sigma}_{\rho,\delta}$:

$$|p|^{(\sigma)}_{l} = \max_{|\alpha+\beta| \leq l} \sup_{(\mathbf{x},\boldsymbol{\xi}) \in \mathbf{R}^n \times \mathbf{R}^n} \left\{ |D^{\alpha}_{\boldsymbol{\xi}} D^{\beta}_{\mathbf{x}} p(\mathbf{x}, \boldsymbol{\xi})| \langle \boldsymbol{\xi} \rangle^{-\sigma(\mathbf{x}) - \rho|\alpha| + \delta|\beta|} \right\}.$$

Definition 8.2. ([61], [62]) For $u \in \mathscr{S}(\mathbf{R}^n)$ (the set of rapidly decreasing functions of the Schwartz space) and $p \in \mathscr{S}^{\sigma}_{\rho,\delta}$, let $P : \mathscr{S}(\mathbf{R}^n) \longrightarrow \mathscr{S}(\mathbf{R}^n)$ be defined as

$$Pu(\mathbf{x}) = \int_{\mathbf{R}^n} e^{i\mathbf{x}\boldsymbol{\xi}} p(\mathbf{x}, \boldsymbol{\xi}) \hat{u}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{8.116}$$

where $\hat{u}(\boldsymbol{\xi}) = \int_{\mathbf{R}^n} e^{-i\mathbf{x}\boldsymbol{\xi}} u(\mathbf{x}) d\mathbf{x}$ (the Fourier transform of u) and $d\boldsymbol{\xi} = (2\pi)^{-n} d\boldsymbol{\xi}$. We refer to $P = p(\mathbf{x}, D_{\mathbf{x}})$ as a pseudodifferential operator of variable order with symbol $p \in \mathscr{S}^{\sigma}_{\rho,\delta}$. The set of all pseudodifferential operators with symbol p of class $\mathscr{S}^{\sigma}_{\rho,\delta}$ is denoted by $\mathscr{S}^{\sigma}_{\rho,\delta}$.

A pseudodifferential operator $P \in \mathscr{S}^{\sigma}_{\rho,\delta}$ is elliptic if there exists $c > 0$ and $M > 0$ such that

$$|p(\mathbf{x}, \boldsymbol{\xi})| \geq c \langle \boldsymbol{\xi} \rangle^{\sigma(\mathbf{x})}, \quad (|\boldsymbol{\xi}| \geq M). \tag{8.117}$$

Furthermore, $Q \in \mathscr{S}^{\infty}_{\rho,\delta} = \bigcup_{m \in \mathbf{R}} \mathscr{S}^{m}_{\rho,\delta}$ is said to be a left (resp. right) parametrix of P if there exists $R_L \in \mathscr{S}^{-\infty}_{\rho,\delta} = \bigcap_{m \in \mathbf{R}} \mathscr{S}^{m}_{\rho,\delta}$ (resp. $R_R \in \mathscr{S}^{-\infty}_{\rho,\delta} = \bigcap_{m \in \mathbf{R}} \mathscr{S}^{m}_{\rho,\delta}$) such that

$$QP = I + R_L \quad (\text{resp.} \quad PQ = I + R_R),$$

where I denotes the identity operator. Pseudodifferential operator Q is a parametrix for P if Q is simultaneously a left and right parametrix of P.

The following properties hold for the class of pseudodifferential operators with variable order $\mathscr{S}^{\sigma}_{\rho,\delta}$.

Proposition 8.7. *(i)* $\mathscr{S}^{\sigma_1}_{\rho,\delta} \subset \mathscr{S}^{\sigma_2}_{\rho,\delta}$ *for* $\sigma_1(\mathbf{x}) \leq \sigma_2(\mathbf{x})$. *In particular,* $\mathscr{S}^{\underline{\sigma}}_{\rho,\delta} \subset \mathscr{S}^{\sigma}_{\rho,\delta} \subset \mathscr{S}^{\overline{\sigma}}_{\rho,\delta}$, *for* $\underline{\sigma} = \inf_{\mathbf{x} \in \mathbf{R}^n} \sigma(\mathbf{x})$ *and* $\overline{\sigma} = \sup_{\mathbf{x} \in \mathbf{R}^n} \sigma(\mathbf{x})$.

(ii) For $P_1 \in \mathscr{S}^{\sigma_1}_{\rho,\delta}$ *and* $P_2 \in \mathscr{S}^{\sigma_2}_{\rho,\delta}$, $P = P_1 \cdot P_2 \in \mathscr{S}^{\sigma_1+\sigma_2}_{\rho,\delta}$.

(iii) *For* $P \in \mathscr{S}^{\sigma}_{\rho,\delta}$, *the formally adjoint operator* P^* *defined by* $(Pu, v) = (u, P^*v)$, *for* $u, v \in \mathscr{S}(\mathbf{R}^n)$, *belongs to* $\mathscr{S}^{\sigma}_{\rho,\delta}$.

(iv) *If* $P \in \mathscr{S}^{\sigma}_{\rho,\delta}$ *is elliptic, then there exists a parametrix of* P *in* $\mathscr{S}^{-\sigma}_{\rho,\delta}$.

We now consider the definition of fractional Sobolev spaces of variable order.

Definition 8.3. Let σ be a real-valued function in $\mathscr{B}^{\infty}(\mathbf{R}^n)$. The Sobolev space of variable order σ on $\mathbf{R}^n$ is defined as

$$H^{\sigma(\cdot)}(\mathbf{R}^n) = \left\{ u \in H^{-\infty} = \bigcup_{s\in\mathbf{R}} H^s(\mathbf{R}^n) : \langle D_{\mathbf{x}}\rangle^{\sigma(\mathbf{x})} u \in L^2(\mathbf{R}^n) \right\},$$

where

$$\langle D_{\mathbf{x}}\rangle^{\sigma(\mathbf{x})} u = \int_{\mathbf{R}^n} (2\pi)^{-n} \exp(i\mathbf{x}\boldsymbol{\xi}) \langle \boldsymbol{\xi}\rangle^{\sigma(\mathbf{x})} \hat{u}(\boldsymbol{\xi}) d\boldsymbol{\xi},$$

and

$$H^s(\mathbf{R}^n) = \left\{ u \in \mathscr{S}'(\mathbf{R}^n) : \langle D_{\mathbf{x}}\rangle^s u \in L^2(\mathbf{R}^n) \right\}.$$

Proposition 8.8. *(see [62]) The above introduced fractional Sobolev spaces of variable order satisfy the following properties:*

(i) *If* $u \in H^{\sigma(\cdot)}(\mathbf{R}^n)$, *then, for* $P \in \mathscr{S}^{\sigma}_{\rho,\delta}$, $Pu \in L^2(\mathbf{R}^n)$.

(ii) *Let* σ_1 *and* σ_2 *be functions in* $\mathscr{B}^{\infty}(\mathbf{R}^n)$ *with* $\sigma_1(\mathbf{x}) \geq \sigma_2(\mathbf{x})$, *for each* $\mathbf{x} \in \mathbf{R}^n$. *Then,* $H^{\sigma_1(\cdot)}(\mathbf{R}^n) \subset H^{\sigma_2(\cdot)}(\mathbf{R}^n)$. *In particular,* $H^{\sigma(\cdot)}(\mathbf{R}^n) \subset H^{\underline{\sigma}(\cdot)}(\mathbf{R}^n)$.

(iii) $H^{\sigma(\cdot)}(\mathbf{R}^n)$ *is a Hilbert space with the inner product*

$$\langle u, v\rangle_{H^{\sigma(\cdot)}(\mathbf{R}^n)} = \int_{\mathbf{R}^n} \left(\langle D_{\mathbf{x}}\rangle^{\sigma(\mathbf{x})} u\right)(\mathbf{x}) \overline{\left(\langle D_{\mathbf{x}}\rangle^{\sigma(\mathbf{x})} v\right)(\mathbf{x})} d\mathbf{x} + \int_{\mathbf{R}^n} \left(\langle D_{\mathbf{x}}\rangle^{\underline{\sigma}} u\right)(\mathbf{x}) \overline{\left(\langle D_{\mathbf{x}}\rangle^{\underline{\sigma}} v\right)(\mathbf{x})} d\mathbf{x}. \tag{8.118}$$

Moreover, $\mathscr{S}(\mathbf{R}^n)$ *is dense in* $H^{\sigma(\cdot)}(\mathbf{R}^n)$.

(iv) *Let* σ *and* τ *be functions in* $\mathscr{B}^{\infty}(\mathbf{R}^n)$. *Suppose that* $P \in \mathscr{S}^{\sigma}_{\rho,\delta}$. *Then, there exists some constant* $C > 0$ *independent of* P *and some positive integer* l *depending only on* $\sigma, \tau, \rho, \delta$, *and* n *such that*

$$\|Pu\|_{H^{\tau(\cdot)}(\mathbf{R}^n)} \leq C\, |p|^{(\sigma)}_l \|u\|_{H^{\sigma(\cdot)+\tau(\cdot)}(\mathbf{R}^n)},$$

for $u \in H^{\sigma(\cdot)+\tau(\cdot)}(\mathbf{R}^n)$, *which provides the continuity of* P *from* $H^{\sigma(\cdot)+\tau(\cdot)}(\mathbf{R}^n)$ *into* $H^{\tau(\cdot)}(\mathbf{R}^n)$.

Remark 8.8. Equivalent norms can be defined on $H^{\sigma(\cdot)}(\mathbf{R}^n)$, considering in Equation (8.118) the inner product $\langle u, v\rangle_{H^s(\mathbf{R}^n)}$ instead of the inner product $\langle u, v\rangle_{H^{\underline{\sigma}}(\mathbf{R}^n)}$, for $s \leq \underline{\sigma}$.

The following result is fundamental in the characterization of the class of generalized random fields of variable order we consider in the next section.

Theorem 8.12. *(see [62]) Let $P \in \mathscr{S}^{\sigma}_{\rho,\delta}$, be elliptic. Then,*

$$H^{\sigma(\cdot)}(\mathbf{R}^n) = \left\{u \in H^{-\infty}(\mathbf{R}^n) : \ Pu \in L^2(\mathbf{R}^n)\right\} \tag{8.119}$$

as a set. Moreover, the norm $\|u\|_{H^{\sigma(\cdot)}}(\mathbf{R}^n)$ is equivalent to the norm

$$\|u\|_{H^{\sigma(\cdot),P}}(\mathbf{R}^n) := \left(\|Pu\|^2_{L^2(\mathbf{R}^n)} + \|u\|^2_{H^{\underline{\sigma}}(\mathbf{R}^n)}\right)^{1/2}. \tag{8.120}$$

The following results on embeddings and lifting properties for fractional Sobolev spaces of variable order on $L^p(\mathbf{R}^n)$ also hold (see [57]).

Theorem 8.13. *Let $1 < p < \infty$ and $j \in \mathbf{N}$, and let $\sigma(\mathbf{x}) = s + \psi(\mathbf{x})$, with $\psi \in \mathscr{S}(\mathbf{R}^n)$, satisfying $0 < m' \leq \sigma(\mathbf{x}) \leq m \leq 2$, for all $\mathbf{x} \in \mathbf{R}^n$. Then, the following asserions hold:*

(i) The space $H_p^{j,\sigma(\cdot)}(\mathbf{R}^n)$ is a Banach space and $C_0^{\infty}(\mathbf{R}^n)$ is dense in this space, where

$$H_p^{j,\sigma(\cdot)}(\mathbf{R}^n) = \left\{f \in \mathscr{S}'(\mathbf{R}^n) : \langle D_{\mathbf{x}}\rangle^{j\sigma(\mathbf{x})} f \in L^2(\mathbf{R}^n)\right\}.$$

(ii) For $m' j > n/p$, the embedding of $H_p^{j,\sigma(\cdot)}(\mathbf{R}^n)$ into $C^{\infty}(\mathbf{R}^n)$ is continuous.

Theorem 8.14. *Assume that the conditions of Theorem 8.13 hold, and let $b(\mathbf{x}, \langle D_{\mathbf{x}}\rangle^{\sigma(\cdot)})$ be a parametrix for $\langle D_{\mathbf{x}}\rangle^{\sigma(\cdot)}$ in $\mathscr{S}^{-\sigma}_{\rho,\delta}$. Then, the following assertions hold:*

(i) The operator $\langle D_{\mathbf{x}}\rangle^{\sigma(\cdot)}$ maps $H_p^{j+1,\sigma(\cdot)}(\mathbf{R}^n)$ continuosly into $H_p^{j,\sigma(\cdot)}(\mathbf{R}^n)$, and $b(\mathbf{x}, \langle D_{\mathbf{x}}\rangle^{\sigma(\cdot)})$ maps $H_p^{j-1,\sigma(\cdot)}(\mathbf{R}^n)$ continuosly into $H_p^{j,\sigma(\cdot)}(\mathbf{R}^n)$.

(ii) If $\langle D_{\mathbf{x}}\rangle^{\sigma(\cdot)} u = f$ holds for $f \in H_p^{j,\sigma(\cdot)}(\mathbf{R}^n)$ and $u \in L^p(\mathbf{R}^n)$, then we have $u \in H_p^{j+1,\sigma(\cdot)}(\mathbf{R}^n)$.
There exists $\lambda_0 \in \mathbf{R}$ such that, for all $\lambda \geq \lambda_0$, the operator $\langle D_{\mathbf{x}}\rangle^{\sigma(\cdot)} + \lambda I$ maps $H_p^{j+1,\sigma(\cdot)}(\mathbf{R}^n)$ onto $H_p^{j,\sigma(\cdot)}(\mathbf{R}^n)$.

8.8.1 Mean-Quadratic Local Variation Properties

Let us define X as the solution to the equation

$$\mathscr{L}_{s(\cdot)} X(x) = \varepsilon(x), \quad x \in \mathbf{R}^n, \tag{8.121}$$

where $\mathscr{L}_{s(\cdot)}$ is a multifractional elliptic pseudodifferential operator of functional order $s(\cdot)$, whose inverse $\mathscr{L}^{-1}_{s(\cdot)}$ has symbol l such that the following identity holds:

$$\mathscr{L}^{-1}_{s(x)}\phi(x) = \int_{\mathbf{R}^n} \exp\left(i\,\langle x, \lambda\rangle\right) l_{s(x)}(x,\lambda)\widehat{\phi}(\lambda) d\lambda, \quad \forall \phi \in \mathscr{D}(\mathscr{L}^{-1}_{s(\cdot)}).$$

Here, ε denotes a zero-mean, second-order random innovation process with RKHS isomorphic to a fractional Sobolev space $H^{\beta(\cdot)}(\mathbf{R}^n)$, of variable order $\beta(\cdot)$, i.e., it admits the following weak-sense covariance factorization (see [99]):

$$R_\varepsilon = \mathscr{T}_\varepsilon \mathscr{T}_\varepsilon^*,$$

where $\mathscr{T}_\varepsilon : L^2(\mathbf{R}^n) \longrightarrow H^{\beta(\cdot)}(\mathbf{R}^n)$ defines an isomorphism.

Remark 8.9. The case where the model is defined on D, a compact domain with C^∞ boundary, or a fractal compact $d-$set, can be treated in a similar way to Section 8.6.2 by definition of the corresponding trace spaces and trace pseudodifferential operators of variable order.

The following result provides the covariance factorization and mean-quadratic local variation properties of the solution to equation (8.121).

Proposition 8.9. *Let X be the solution to equation (8.121). If $\inf_{x\in\mathbf{R}^n} \beta(x) + s(x) > n/2$, the following assertions hold:*

(i) The covariance function $B_X(x,y) = E[X(x)X(y)]$, $x, y \in \mathbf{R}^n$, admits the following factorization:

$$\int_{\mathbf{R}^n} B_X(x,y)\phi(y)dy = \int_{\mathbf{R}^n} \exp\left(i\,\langle x,\lambda\rangle\right) l_{s(x)}(x,\lambda) t_\varepsilon(x,\lambda)\bar{t}_\varepsilon(x,\lambda)\bar{l}_{s(x)}(x,\lambda) \times \widehat{\phi}(\lambda)d\lambda, \quad \forall \phi \in \mathscr{D}(R_X),$$

where $\mathscr{D}(R_X)$ denotes, as before, the domain of the covariance operator R_X of X, and t_ε denotes the symbol of the pseudodifferential operator $\mathscr{T}_\varepsilon$.

(ii) The increments of X satisfy that there exists a positive constant C, such that

$$E[X(x+h) - X(x)] \leq C\,|h|^{2(\beta(x)+s(x))-n}, \quad h \longrightarrow 0, \quad \forall x \in \mathbf{R}^n.$$

Proof. (i) The proof follows from the assumption on the isomorphic relationship of the RKHS of the innovation process and the multifractional Sobolev space $H^{\beta(\cdot)}(\mathbf{R}^n)$, as well as from the fact that $\mathscr{L}_{s(\cdot)}$ is an elliptic pseudodifferential operator of variable order.
(ii) From (i), and equations (8.115) and (8.117), the asymptotic order of the symbol of R_X, the covariance operator of X having kernel B_X, is $\beta(\cdot) + s(\cdot)$. Specifically, the symbol

$$r_X(x,\lambda) = l_{s(x)}(x,\lambda) t_\varepsilon(x,\lambda)\bar{t}_\varepsilon(x,\lambda)\bar{l}_{s(x)}(x,\lambda), \quad x, \lambda \in \mathbf{R}^n,$$

of R_X displays an equivalent behavior to the symbol

$$\langle\lambda\rangle^{-2(\beta(x)+s(x))} = (1+|\lambda|^2)^{-(\beta(x)+s(x))}, \text{ for all } x, \lambda \in \mathbf{R}^n,$$

i.e.,

$$\langle\lambda\rangle^{-2(\beta(x)+s(x))} \simeq r_X(x,\lambda), \quad x, \lambda \in \mathbf{R}^n, \quad |\lambda| \longrightarrow \infty.$$

The rest of the proof follows from the fact that the functions of the RKHS of X then display a local behavior characterized in terms of the functional local Hölder exponent $\beta(\cdot) + s(\cdot) - n/2$.

Acknowledgements N.N. Leonenko and M.D. Ruiz-Medina partially supported by grant of the European commition PIRSES-GA-2008-230804 (Maric Curie), projects MTM2009-13393 of the DGI, and P09-FQM-5052 of the Andalousian CICE, Spain, and the Australian Research Council grant DP 0345577

References

1. Adler, R.J.: The Geometry of Random Fields. Wiley, Chichester (1981)
2. Albeverio, S., Molchanov, S.A. and Surgailis, D.: Stratified structure of the Universe and Burgers' equation: A probabilistic approach. Prob. Theor. Rel. Fields **100**, 457–484 (1994)
3. Angulo, J.M., Ruiz-Medina, M.D., Anh, V.V., Grecksch, W.: Fractional diffusion and fractional heat equation. Adv. in Appl. Prob. **32**, 1077–1099 (2000)
4. Anh V.V., Angulo, J.M., Ruiz-Medina, M.D.: Possible long-range dependence in fractional random fields. J. Stat. Plan. Infer. **80**, 95–110 (1999)
5. Anh, V.V., Leonenko, N.N.: Non-Gaussian scenarios for the heat equation with singular initial conditions. Stoch. Proc. Appl. **84**, 91–114 (1999)
6. Anh, V. V., Leonenko, N. N.: Scaling laws for fractional diffusion-wave equations with singular data. Statist. Probab. Lett. **48**, 239–252 (2000)
7. Anh, V. V., Leonenko, N. N.: Spectral analysis of fractional kinetic equations with random data. J. Statist. Phys. **104**, 1349–1387 (2001)
8. Anh, V.V., Leonenko, N.N.: Renormalization and homogenization of fractional diffusion equations with random data. Prob. Theor. Rel. Fields **124**, 381–408 (2002)
9. Anh, V.V., Leonenko, N. N.: Spectral theory of renormalized fractional random fields. Theory Probab. Math. Statist. **66**, 1–13 (2003)
10. Anh, V.V., Leonenko, N.N.: Harmonic analysis of random fractional diffusion-wave equations. Appl. Math. Comput. **141**, 77–85 (2003)
11. Anh, V.V., Leonenko, N.N., Melnikova, O.O.: Scaling laws for fractional Volterra equations with chi-square random data. Math. Commun. **7**, 159–175 (2002)
12. Anh, V. V., Leonenko, N.N., Moldavskaya, E.M., Sakhno, L.M.: Estimation of spectral densities with multiplicative parameter. Acta Appl. Math. **79**, 115–128 (2003)
13. Anh, V.V., Leonenko, N.N., Sakhno, L.M.: Higher-order spectral densities of fractional random fields. J. Statist. Phys. **111**, 89–814 (2003)
14. Anh, V., Leonenko, N.N., Sakhno, L.M.: On a class of minimum contrast estimators for fractional stochastic processes and fields. J. Stat. Plan. Inf. **123**, 161–185 (2004)
15. Anh, V.V., Leonenko, N.N., Sakhno, L.M.: Spectral properties of Burgers and KPZ turbulence. J. Statist. Physics **122**, 949–974 (2006)
16. Anh, V.V., Leonenko, N.N., Shieh, N.R. Multifractal scaling of products of birthdeath processes. Bernoulli **15**, 508–531 (2009)
17. Bakhtin, Yu.: A functional central limit theorem for transformed solutions of the multidimensional Burgers equation with random initial data. Theory Probab. Appl. **46**, 387–405 (2001)
18. Bakhtin, Y.: Burgers equation with random boundary conditions. Proc. Amer. Math. Soc. **135**, 2257–2262 (2007)
19. Barabasi, A.L., Stanley, H.E.: Fractal Concepts in Surface Growth. Cambridge University Press, Cambridge (1995)
20. Barndorff-Nielsen, O.E., Leonenko, N.N.: Burgers turbulence problem with linear or quadratic external potential. J. Appl. Probab. **42**, 550–565 (2005)

21. Bec, J., Khanin, K.: Burgers turbulence. Phys. Rep. **447**, 1–66 (2007)
22. Beghin, L., Knopova, V. P., Leonenko, N. N., Orsingher, E.: Gaussian limiting behavior of the rescaled solution to the linear Korteweg-de Vries equation with random initial conditions. J. Statist. Phys. **99**, 769–781 (2000)
23. Benassi, A., Jaffard, S., Roux, D.: Elliptic Gaussian random processes. Rev. Mat. Iberoam. **13**, 19–90 (1997)
24. Berline, N., Getzler, E., Vergne, M.: Heat Kernels and Dirac Operators. Springer, Berlin (1992)
25. Bertini, L., Cancrini, N.: The stochastic heat equation: Feynman-Kac formula and intermittence. J. Stat. Physics **78**, 1377–1401 (1995)
26. Bertoin, J.: The inviscid Burgers equation with Brownian initial velocity. Commun. Math. Phys. **193**, 397–406 (1998)
27. Breuer, P., Major, P.: Central limit theorem for non-linear functionals of Gaussian fields. J. Multiv. Anal. **13**, 425-441 (1983)
28. Bulinski, A.V.: Central limit theorem for the solution of the multidimensional Burgers equation with random data. Ann. Acad. Sci. Fenn. Ser. A I Math. **17**, 11–22 (1992)
29. Bulinski, A.V., Molchanov, S.A.: Asymptotic Gausianness of solutions of the Burgers equation with random initial conditions. Theor. Prob. Appl. **36**, 217–235 (1991)
30. Burgers, J.: The Nonlinear Diffusion Equation. Kluwer, Dordrecht (1974)
31. Chigirinskaya, Y., Marsan, D.: Multifractal cascade dynamics and turbulent intermittency. Fractals **5**, 427–471 (1997)
32. Chorin, A.J.: Lecture Notes in Turbulence Theor. Publish or Perish. Berkeley, CA (1975)
33. Christakos, G.: Modern Spatiotemporal Geostatistics. University Press, Oxford (2000)
34. Cycon, H.L., Froese, R.G., Kirsch.,W., Simon, B.: Schrödinger Operators. Springer, Berlin (1987)
35. Demuth, M., Van Casteren, J.A.: Stochastic Spectral Theory for Selfadjoint Operators. Birkhauser-Verlag, Basel (2000)
36. Deriev, I.,Leonenko, N.: Limit Gaussian behavior of the solutions of the multidimensional Burgers' equation with weak-dependent initial conditions. Acta Applicandae Math. **47**, 1–18 (1997)
37. Dermone, A., Hamadéne, S., Ouknine, Y.: Limit theorem for the statistical solution of Burgers equation. Stoch. Proc. Appl. **81**, 217–230 (1999)
38. Dobrushin, R.L., Major, P.: Non-central limit theorems for nonlinear fuctionals of Gaussian fields. Z. Wahrsch. verw. Gebiete. **50**, 1–28 (1979)
39. Doukhan, P., Oppenheim, G., Taqqu, M.S.: Theory and Applications of Long-Range Dependence. Birkhauser, Boston (2003)
40. Falconer, K.: Techniques in Fractal Geometry. Wiley, New York (1997)
41. Frisch, U.: Turbulence. Cambridge University Press, Cambridge (1995)
42. Funaki, T., Surgailis, D., Woyczynski, W.A.: Gibbs-Cox random fields and Burgers turbulence. Ann. Appl. Prob. **5**, 461–492 (1995)
43. Gay, R., Heyde, C.C.: On a class of random field model which allows long dependence. Biometrika **77**, 401–403 (1990)
44. Giraitis, L., Molchanov, S. A., Surgailis, D.: Long Memory Shot Noises and Limit Theorems with Application to Burgers' Equation. New Directions in Time Series Analysis, Part II, 153–176, IMA Vol. Math. Appl. **46**. Springer, New York (1993)
45. Griniv, O.O.: A central limit theorem for the Burgers equation. Theoret. and Math. Phys. **88**, 678–682 (1992)
46. Gupta, V.K., Waymire, E.: A statistical analysis of mesoscale rainfall as a random cascade. J. Appl. Meteorol. **32**, 251–267 (1993)
47. Gurbatov, S.: Universality classes for self-similarity of noiseless multidimensional Burgers turbulence and interface growth. Phys. Rev. E **61**, 2595-2604 (2000)
48. Gurbatov, S., Malakhov, A., Saichev, A.: Non-linear Waves and Turbulence in Nondispersive Media: Waves, Rays and Particles. Manchester University Press, Manchester (1991)
49. Gurbatov, S.N., Simdyankin, S.I. Aurell, E., Frisch, U., Tó th, G.: On the decay of Burgers turbulence. J. Fluid Mech. **344**, 339–374 (1997)

50. Hilfer, R. (Ed.): Applications of Fractional Calculus in Physics. World Scientific, Singapore (2000)
51. Hodges, S.D., Carverhill, A.P.: Quasi mean reversion in an efficient stock market: The characterisation of economic equilibria which support Black-Scholes option pricing. Economic Journal **103**, 395–405 (1993)
52. Hopf, E.: The partial differential equation $u_x + uu_x = \mu u_{xx}$. Commun. Pure Appl. Math. **3**, 201–230 (1950)
53. Hosking, J.R.M.: Fractional differencing. Biometrika **68**, 165–176 (1981)
54. Iribarren, I., León, J.R.: Central limit theorem for solutions of random initialized differential equations: A simple proof. J. Appl. Math. Stoch. Anal. doi:10.1155/JAMSA/2006/35206 (2006)
55. Ishiyama, K.: Methods for evaluating density functions of exponential functionals represented as integrals of geometric Brownian motion. Methodology and Computing in Applied Probability **7**, 271-283 (2005)
56. Ivanov, A.V., Leonenko, N.N.: Statistical Analysis of Random Fields. Kluwer, Dordrecht (1989)
57. Jacob, N., Leopold, H.G.: Pseudo differential operators with variable order of differentiation generating Feller semigroups. Integr. Equat. Oper. Th. **17**, 544–553 (1993)
58. Jaffard, S.: The multifractal nature of Lévy processes. Probab. Theory Rel. **114**, 207–227 (1999)
59. Kampe de Ferier: Random solutions of the partial differential equations, Proc. 3rd Berkeley Symp. Math.Stat. Probab. Vol. III, 199–208. University of California Press, Berkley, California (1955)
60. Kelbert, M., Leonenko, N., Ruiz-Medina, M.D.: Fractional random fields associated with stochastic fractional heat equations. Adv. Appl. Probab. **37**, 108–133 (2005)
61. Kikuchi, K., Negoro, A.: Pseudo differential operators with variable order of differentation. Rep. Fac. Liberal Arts, Shizuoka University, Sciences **31**, 19–27 (1995)
62. Kikuchi, K., Negoro, A.: On Markov processes generated by pseudodifferentail operator of variable order. Osaka J. Math. **34**, 319–335 (1997)
63. Kochmanski, S.: On the evolution operators for some equations of mathematical physics with variable coefficients. Ukr. Math. J. **46**, 938–952 (1994)
64. Leonenko, N.: Limit Theorems for Random Fields with Singular Spectrum. Kluwer, Dordrecht (1999)
65. Leonenko, N.N, Li Z.B., Rybasov, K.V.: Non-Gaussian limit distributions of solutions of the multidimensional Burgers equation with random data. Ukrain. Math. J. **47**, 330–336 (1995)
66. Leonenko, N.N., Melnikova, O.A.: Renormalization and homogenization of solutions of heat equation with linear potential and related Burgers equation with random data. Theor. Prob. Math. Stat. **62**, 72–82 (2000)
67. Leonenko, N., Orsingher, E.: Limit theorems for solutions of Burgers equation with Gaussian and non-Gaussian initial data. Theor. Prob. Appl. **40**, 387–403 (1995)
68. Leonenko, N. Orsingher, E., Parkhomenko, N.: On the rate of convergence to the normal law for solutions of the Burgers equation with singular initial data. J. Stat. Phys. **82**, 915–930 (1995)
69. Leonenko, N.N, Orsingher, E., Rybasov, K.V.: Limit distributions of solutions of the multidimensional Burgers equation with random intial data I,II. Ukrain. Math. J. **46**, 870–877 (1994)
70. Leonenko, N.N., Ruiz-Medina, M.D.: Scaling laws for the multidimensional Burgers equation with quadratic external potential. J. Statist. Physics **124** 191–205 (2006a).
71. Leonenko, N.N., Ruiz-Medina, M.D.: Strongly dependent Gaussian scenarios for the Burgers turbulence problem with quadratic external potential. Random Operators and Stoch. Equ. **14**, 259–274 (2006b)
72. Leonenko, N.N., Ruiz-Medina, M.D.: Gaussian scenario for the heat and Burgers equations with quadratic external potential and weakly dependent data with applications. Methodology and Computing in Applied Probability **10**, 595–620 (2008)

73. Leonenko N.N., Ruiz-Medina, M.D.: Spatial scaling for randomly initialized heat and Burgers equation with quadratic potential. Stoch. Anal. Appl. **28**, 1–19 (2010)
74. Leonenko, N.N., Zhanbing, L., Rybasov, K.V.: Non-Gaussian limit distributions of solutions of the Burgers equation with strongly dependent random initial conditions. Random Oper. Stoch. Equations **2**, 95–102 (1994).
75. Leonenko, N.N., Woyczynski, W.A.: Exact parabolic asymptotics for singular $n - \mathscr{D}$ Burgers random fields: Gaussian approximation. Stoch. Proc. Appl. **76**, 141–165 (1998)
76. Leonenko, N.N., Woyczynski, W.A.: Scaling limits of solutions of the heat equation for singular non-Gaussian data. J. Stat. Phys. **91**, 423–438 (1999)
77. Leonenko, N. N., Woyczynski, W. A.: Parameter identification for stochastic Burgers' flows via parabolic rescaling. Probab. Math. Statist. **21**, 1–55 (2001)
78. Leopold, H.G.: On Besov spaces of variable order o differentiation. Zeitschrift für Analysis und ihre Anwendungen **8**, 69–82 (1989)
79. Leopold, H.G.: On function spaces of variable order of differentiation. Forum Mathematicum **3**, 1–21 (1991)
80. Leopold, H.G.: Embedding of function spaces of variable order of differentation in function spaces of variable order of integration. Czechoslovak Mathematical Journal **49**, 633–644 (1999).
81. Liu, G.R., Shieh, N.R.: Scaling limit for some P.D.E. systems with random initial conditions. Stochastic Analysis and Applications **28**, 505–522 (2010a)
82. Liu, G.R., Shieh, N.R.: Scaling limit for some P.D.E. systems with random initial conditions. Stochastic and Dynamics **10**, 1–35 (2010b).
83. Liu, G.R., Shieh, N.R.: Homogenization of fractional kinetic systems with random initial data, submitted (2010c)
84. Mandelbrot, B.B.: Fractals, Form, Chance and Dimension. Freeman, San Francisco (1977)
85. Mandelbrot, B.B., Van Ness, J.W.: Fractional Brownian motion, fractional noises and applications. SIAM Review **10**, 422–437 (1968)
86. Mandelbrot, B. B., Fisher, A., Calvet L.: A Multifractal Model of Asset Returns. Cowles Foundation Discussion Paper **1164**. Yale University, New Haven (1997).
87. Molchanov, S.A., Surgailis, D., Woyczynski, W.A.: Hyperbolic asymptotics in Burgers turbulence. Commun. Math. Phys. **168**, 209–226 (1995).
88. Molchanov, S.A., Surgailis, D., Woyczynski, W.A.: The large-scale structure of the Universe and quasi-Voronoi tessellation of shock fronts in forced Burgers turbulence in R^d. Ann. Appl. Prob. **7**, 220–223 (1997)
89. Nualart, D., Peccati, G. Central limit theorems for sequences of multiple stochastic integrals. Convergence in law of multiple stochastic integrals. Ann. Prob. **33**, 177–193 (2005)
90. Ramm, A.G.: Random Fields Estimation Theory, Logman Scientific & Technical (1990)
91. Riedi, R.H.: Multifractal processes. In Theory and Applications of Long-Range Dependence P. Doukhan, G. Oppenheim, and M.S. Taqqu (Eds.), 625–716, Birkhäuser (2003)
92. Rosenblatt, M.: Some remark on the Burgers equation. J. Math. Phys. **9**, 1129–1136 (1968)
93. Rosenblatt, M.: Scale renormalization and random solutions of Burgers equation. J. Appl. Prob. **24**, 328–338 (1987)
94. Ruiz-Medina, M.D.: Functional denoising and reconstruction of fractal image sequences. Operators and Stochastic Equations **17**, 275–293 (2009)
95. Ruiz-Medina, M.D., Angulo, J.M., Anh, V.V.: Scaling limit solution of a fractional Burgers equation. Stoch. Proc. Appl. **93**, 285–300 (2001)
96. Ruiz-Medina, M.D., Angulo, J.M., Anh, V.V.: Fractional generalized random fields on bounded domains. Stoch. Anal. Appl **21** 465–492 (2003)
97. Ruiz-Medina, M.D., Angulo, J.M., Anh, V.V.: Karhunen-Loève-Type representations on fractal domains. Stoch. Anal. Appl. **24**, 195–219 (2006)
98. Ruiz-Medina, M.D., Angulo, J.M., Anh, V.V.: Multifractality in Space-Time Statistical Models. Stoch. Env. Res. Risk A. **22**, 81–86 (2008)
99. Ruiz-Medina, M.D., Anh, V.V., Angulo, J.M.: Fractional generalized random fields of variable order. Stoch. Anal. Appl. **22**, 775–800 (2004)

100. Ruiz-Medina, M.D., Anh, V.V., Angulo, J.M.: Multifractional Markov processes in heterogeneous domains. Stoch. Anal. Appl. **29**, 15–47 (2010)
101. Ryan, R.: The statistics of Burgers turbulence initiated with fractional Brownian-noise data. Commun. Math. Phys. **191**, 1008–1038 (1998)
102. Saichev A.I., Woyczynski, W.A.: Evolution of Burgers' turbulence in the presence of external forces. J. Fluid Mech. **331**, 313–343 (1997)
103. Shandarin, S.F., Zeldovich, Ya.B.: Turbulence, intermittency, structures in a left-gravitating medium: The large scale structure of the Universe. Rev. Modern Phys. **61**, 185–220 (1989)
104. Sinai, Ya. G.: Statistics of shocks in solutions of inviscid Burgers equation. Commun. Math. Phys. **148**, 601–621 (1992)
105. Stein, M. L.: Interpolation of Spatial Data. Springer, Berlin (1999)
106. Surgailis, D.: Asymptotics of solutions of Burgers' equation with random piecewise constant data. Stochastic Models in Geosystems, 427–441, IMA Vol. Math. Appl., 85. Springer, New York (1997).
107. Surgailis, D., Woyczynski, W.A.: Burgers' equation with non-local shot noise data. J. Appl. Probab. **31**, 351–362 (1994a)
108. Surgailis, D., Woyczynski, W.A.: Scaling limits of solutions of the Burgers equation with singular Gaussian initial data. In: HoudrLe, C., PLerez-Abreu, V. (Eds.). Multiple Wiener-Ito Integrals and Their Applications. CRC Press, Boca Raton, (1994b)
109. Surgailis, D., Woyczyński, W.A.: Limit theorems for the Burgers equation initialized by data with long-range dependence. Theory and applications of long-range dependence, 507–523. Birkhäuser Boston, Boston, MA (2003)
110. Taqqu, M.S.: Weak convergence to fractional Brownian motion and to the Rosenblatt process. Z. Wahrsch. verw. Gebiete **31**, 287–302 (1975)
111. Taqqu, M.S.: Convergence of integrated processes of arbitrary Hermite rank. Z. Wahrsch. verw. Gebiete **40**, 203–238 (1979)
112. Triebel, H.: Interpolation Theory, Function Spaces, Differential Operators. North-Holland Publishing Co. Amsterdam (1978)
113. Triebel, H.: Fractals and Spectra. Birkhauser (1997)
114. Whittle, P.: Stochastic processes in several dimensions. Bull. Ins. Internat. Stattist. **40**, 974–994 (1963)
115. Winkel, M.: Burgers turbulence initialized by a regenerative impulse. Stochastic Process. Appl. **93**, 241–268 (2001)
116. Winkel, M.: Limit clusters in the inviscid Burgers turbulence with certain random initial velocities. J. Statist. Phys. **107**, 893–917 (2002)
117. Witham, G.B.: Linear and Nonlinear Waves. Wiley, New York (1974)
118. Woyczynski, W.A.: Burgers-KPZ Turbulence. Göttingen Lectures. Lecture Notes in Mathematics **1706**. Springer, Berlin (1998)
119. Yadrenko, M.I.: Spectral Theory of Random Fields. Optimization Software. New York (1983)

Chapter 9
On Some Local, Global and Regularity Behaviour of Some Classes of Covariance Functions

Emilio Porcu and Michael L. Stein

Abstract Two critical properties of stationary random fields are their degree of smoothness and the rate of decay of correlation at long lags. These properties are in turn closely connected to the behaviour of the random fields' spectral densities at infinity and at the origin, respectively. Many standard models have flexibility at one but not both of these scales. Recent works have proposed a number of models with at least some flexibility at both scales. This chapter summarizes some of these proposed models and analyzes their local and global behaviour, both in terms of their covariance functions and the associated spectra. Some ways to obtain models allowing greater flexibility are described.

9.1 Introduction

Recent literature emphasizes a growing interest in the study of Gaussian random fields (RFs) characterized by parametric families of covariance functions; see, *e.g.*, [14, 16] and [28] for the generalized Cauchy covariance function or [29] for the generalized Matérn covariance function. Two critical aspects of random fields are their behavior over short scales (or equivalently, at high frequencies) and over long scales (or low frequencies). Some standard models have flexibility at one but not both of these scales. Recent years have seen a number of proposed models that have at least some flexibility at both scales. For example, the generalized Cauchy

E. Porcu (✉)
Universität Göttingen, Institut for Mathematics Stochastics, Universidad de Castilla-La Mancha, Facultad de Derecho y Ciencias Sociales, Área de Estadística, Avenida Camilo Jose Cela 14, 13005 Ciudad Real, España,
e-mail: emilio.porcu@uclm.es

M.L. Stein
Department of Statistics, The University of Chicago, Chicago, USA
e-mail: stein@galton.uchicago.edu

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_9,

family allows different values for the fractal dimension, a descriptor of short-scale behavior, and for the Hurst parameter, a descriptor of large-scale behavior [16]. Houdre and Villa [21] generalized fractional Brownian motion parametrized by a single Hurst index to the bifractional Brownian motion characterized by two indices. Other notable works include [3, 51], and [2] for a more abstract representation of generalized RFs and covariance factorizations in the Riesz sense. Some properties of fractional Riesz-Bessel field are offered in [30]. Properties of the generalized Linnik probability density are studied in [11, 38] and [26], whilst those of the generalized Ornstein-Uhlenbeck process are offered in [31]. For a statistical view of the subject, we cite [9] and [59], as well as [46, 47] and [45]. The subject has been extensively treated in the field of termomechanics, for which we cite [35–37]. Finally, excellent textbook references are [1, 12, 27] and [53].

Properties of a Gaussian RF are intimately connected with that of its associated covariance function. In the weakly stationary case, (in chronological order), Matheron [32], Christakos [7], Yaglom [61], Stein [55] and Christakos [8] show that we can establish several properties of RFs through the study of their correlation functions or, equivalently, spectral densities.

This chapter offers a view of the most important covariance functions stemming from the statistical literature and for which critical properties of the corresponding Gaussian RFs can be obtained. In order to do this, we appeal to literature coming from fields as diverse as statistics, engineering, mathematical physics and, last but not least, numerical analysis. The last is especially important for the extensive study of the properties of covariances, called kernels, in terms of reproducing kernel Hilbert spaces (RKHS) and Sobolev spaces associated with them that allow one to identify the regularity properties of RFs with a given kernel. Sobolev spaces can be viewed as special cases of RKHS and the celebrated imbedding theorems allow the study of continuity properties of Gaussian fields. Also, RFs evaluated on Hilbert spaces constitute a fundamental tool for the study of the equivalence of Gaussian measures and its relationship to spatial interpolation; see Ibragimov and Rozanov [22] Yadrenko [60] Stein [54];[55] and Zhang [62]. In particular, under infill asymptotics, Stein [54] shows that the effect of covariance misspecification on optimal linear prediction (known as kriging in the geostatistical literature) is asymptotically negligible if the Gaussian measures corresponding to the correct and misspecified covariance functions are equivalent. As a consequence, it turns out that for spatial interpolation, some parameters of standard models do not have to be estimated well in order to get nearly optimal predictions and accurate assessments of mean squared error.

The plan of the chapter is the following: in Sect. 2 we offer some basic tools for ordinary and generalized RFs, long range dependence and fractal dimension, and RKHS and Sobolev spaces. In Sect. 3 we resume discussion of the properties of the main classes of (ordinary or generalized) covariance functions through the study of their spectra. In particular, we describe ways to generate families of covariance functions that have the same flexibility as the Matérn family at high frequencies and the same flexibility at low frequencies as the generalized Cauchy family.

9.2 Preliminaries

In this section we report some basic facts about weakly stationary and isotropic Gaussian RFs, the characterization of the associated family of covariance functions, and the identification of their local regularity and asymptotic orders in the fractal and long-range dependence case. Basic facts concerning the RKHS theory are also provided.

Throughout this chapter we shall refer to real-valued zero mean weakly stationary Gaussian RFs $\{X(x),\ x \in \mathbf{R}^n\}$ having covariance structure

$$C(\xi) := \mathbf{E}\,(X(x+\xi)X(x))\,.$$

Covariance functions are positive definite, *i.e.* for any finite system of points $\{x_i\}_{i=1}^N$ in $\mathbf{R}^n$ and collection of real constants $\{c_i\}_{i=1}^N$,

$$\sum_{i,j=1}^{N} a_i C(x_i - x_j) a_j \geq 0.$$

Bochner's theorem establishes the equivalence between the class of continuous positive definite functions on $\mathbf{R}^n$ and that of Fourier transforms of positive and bounded measures defined on $\mathbf{R}^n$; that is,

$$C(\xi) = \int \exp(i\tau\xi)\mathrm{d}\mu(\tau).$$

Here, we are only interested in real-valued random fields, in which case, we only need to consider measures μ that are symmetric about the origin. Additionally, if $\mu(\cdot)$ is absolutely continuous with respect to Lebesgue measure (when $C \in L_1(\mathbf{R}^n)$), then the above integral can be written with respect to the spectral density function $f(\cdot) = (\mathrm{d}/\mathrm{d}\cdot)\,\mu(\cdot)$.

We shall work under the additional assumption of isotropy or radial symmetry, that is $C(\xi) := \widetilde{C}(\|\xi\|)$, for $\|\cdot\|$ denoting the Euclidean norm and $\widetilde{C} : \mathbf{R} \to \mathbf{R}$ a continuous function such that the composition $\widetilde{C}(\|\cdot\|)$ is positive definite on $\mathbf{R}^n$. Thus, a covariance function is isotropic if $C(\xi_1) = C(\xi_2)$ whenever $\|\xi_1\| = \|\xi_2\|$.

A function $g : [0,\infty[\to \mathbf{R}$ is completely monotone if

$$(-1)^k g^{(k)}(t) \geq 0, \qquad 0 < t, \quad \forall k \in \mathbf{N}.$$

By Schoenberg's theorem [52], a function $g : [0,\infty[\to \mathbf{R}$, $g(\|\xi\|^2)$ is continuous and positive definite on $\mathbf{R}^n$ for all $n \geq 1$ if and only if g is completely monotone on $[0,\infty[$. Thus, the search for radial functions that are positive definite on n-dimensional Euclidean spaces for all n coincides with that of completely monotone functions on the positive real line. A correlation function is just a covariance function that equals 1 at the origin. We see that correlation functions

of isotropic RFs are also characteristic functions of random vectors whose measures are invariant to rotations about the origin. Thus, we may use equivalently these definitions whenever no confusion may arise. As a last remark, positive definiteness conditions are called *permissibility* conditions in Christakos [7].

9.2.1 Fractals and Long Memory

The fractal dimension of a surface in $\mathbf{R}^n$ provides its roughness measure, with range $[n, n+1)$. The long memory in time series or spatial data is associated with power law correlations, and often referred to as Hurst effect (H effect). Long memory dependence is then characterized by the H parameter.

Local regularity properties of the sample-paths of a Gaussian RF have an intimate connection with its second-order regularity properties (see [1]). In particular, in the weakly stationary case, if, for some $\alpha \in (0, 2]$,

$$\lim_{\|\xi\| \to 0} \quad (1 - C(\xi)) \, \|\xi\|^{-\alpha} = C_0, \quad 0 < C_0 < \infty, \tag{9.1}$$

then, with probability one, the random field X satisfies

$$d = \dim(\mathrm{Gr}\, X) = \min\left(\frac{n}{\alpha/2}, n + 1 - \alpha/2\right),$$

where, as before, C denotes the covariance function of the random field X. Here, Gr X denotes the graph$(X) = \{(t, X(t)),\ t \in [-1, 1]^n\} \subset \mathbf{R}^{n+1}$, thus the estimate of α provides an estimate of the fractal dimension d. Equation (9.1) refers to the issue of scaling laws, which describe the way in which rather elementary measurements vary with the size of measurement unit, and we refer to Hall and Wood [20] for a detailed analysis of the relation between the fractal index α and the fractal dimension d, as well as to the previous work in [1] on Gaussian index-β random fields, with $\beta = \alpha/2$ in this case.

Now consider the behavior of C at long lags. If, for some $\beta \in (0, n)$ and $0 < C_0 < \infty$,

$$\lim_{\|\xi\| \to \infty} \quad C(\xi) \|\xi\|^{-n+\beta} = C_0, \tag{9.2}$$

then the process is said to have long memory, with Hurst coefficient $H = \beta/2$. Note that (9.2) implies that C is not integrable on $\mathbf{R}^n$. If $H \in (n/2, n)$ or $H \in (0, n/2)$ the correlation is said to be respectively persistent or anti-persistent. Tauberian and Abelian theorems (see, e.g., [5]) relate the behavior of a function at the origin with that of its Fourier transform at infinity. Thus, the parameter α, which in (9.1) controls the behavior at the origin of C, is related to the rate of decay of the spectral density at high frequencies, while the parameter β in (9.2) is associated with the behavior of the spectral density at low frequencies.

In relation to the generalized Gaussian RF framework, we remark that, for k a positive integer, we say that a function $K : \mathbf{R}^n \to \mathbf{R}$ is conditionally positive definite of order k if, for any collection $x_1, \ldots, x_N \in \mathbf{R}^n$ and reals $\alpha_1, \ldots, \alpha_N$, such that $\sum_{j=1}^N \alpha_j\, p(x_j) = 0$, for all polynomials $p(\cdot)$ of degree at most k, the inequality $\sum_{i,j=1}^N a_i\, K(x_i - x_j) a_j \geq 0$ holds. A counterpart of Bochner's characterization is available and we refer the reader to Wendland [57] for more precise characterization theorems in this context. For the weak-sense definition of covariance functions, from the theory of distributions, we refer the reader to [19] and, in the fractional pseudodifferential case, to [44]. The Moak class defined in equation (9.15) thus represents the correlation function of some Gaussian generalized RF defined on a suitable class of test functions.

9.2.2 Reproducing-Kernel Hilbert Spaces

In this section, we shall introduce the notion of a reproducing-kernel Hilbert space, which is the base to describe the classes of functions that are dealt with in statistics. Requiring a covariance function to be continuous is not always necessary, but allowing for discontinuous covariances (called *kernels* in the literature stemming from numerical analysis) would complicate many of our considerations, so we shall generally stick to continuity as one of our working assumptions in this and all subsequent chapters. For a symmetric positive definite function C on some domain $\mathscr{D}$, define

$$H_C := \left\{ \sum_{i=1}^m a_i\, C(x_i, \cdot) \;:\; a_i \in \mathbf{R},\; x_i \in \mathscr{D},\; m \in \mathscr{N} \right\}, \tag{9.3}$$

with inner product

$$\left(\sum_{i=1}^m a_i\, C(x_i, \cdot), \sum_{j=1}^{m'} b_j\, C(x_j, \cdot) \right)_{\mathscr{H}_C} := \sum_{i=1}^m \sum_{j=1}^{m'} a_i\, b_j\, C(x_i, x_j). \tag{9.4}$$

By the positive definiteness of C we have $(f, f)_{\mathscr{H}_C} \geq 0$ for all $f \in H_C$, and $(f, f)_{\mathscr{H}_C} = 0$ if and only if $f \equiv 0$, so the inner product (9.4) defines a norm $\|f\|_{\mathscr{H}_C} = (f, f)_{\mathscr{H}_C}^{1/2}$ on H_C. Furthermore, for any $f \in H_C$, we have

$$\big(f, C(x, \cdot)\big)_{\mathscr{H}_C} = \left(\sum_{i=1}^m a_i\, C(x_i, \cdot), C(x, \cdot) \right)_{\mathscr{H}_C} = \sum_{i=1}^m a_i\, C(x_i, x) = f(x), \tag{9.5}$$

which is called the *reproducing kernel* property.

The closure of H_C under $\|\cdot\|_{\mathscr{H}_C}$ is a space of real-valued functions, denoted by $\mathscr{H}_C$, and called the *reproducing kernel Hilbert space* of C. By the continuity of the inner product, the reproducing equation (9.5) carries over to $\mathscr{H}_C$.

9.2.3 Sobolev Spaces

Following [50] and [57] we introduce an important class of RKHSs, the *Sobolev spaces*. These spaces guarantee a certain smoothness of the functions they contain. They are the natural function spaces for the sample paths of second order random fields. For an (arbitrary) domain $\mathscr{D} \subseteq \mathbf{R}^n$ we denote by $\mathscr{C}(\mathscr{D})$ the space of continuous (real-valued) functions, by $\mathscr{C}^k(\mathscr{D})$ the space of k times continuously differentiable functions and by $\mathscr{C}^\infty(\mathscr{D})$ the space of infinitely differentiable functions $f : \mathscr{D} \to \mathbf{R}$. Finally, for $f \in \mathscr{C}^k(\mathscr{D})$ and a multi-index $\alpha \in \mathscr{N}_0^n$ of order $|\alpha| \leq k$, $|\alpha| := \sum_{i=1}^n \alpha_i$, let $D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial {e_1}^{\alpha_1} \cdots \partial {e_n}^{\alpha_n}}$, where e_i is the unit vector in $\mathbf{R}^n$ in the direction of the i^{th} coordinate axis. $D^\alpha f$ is the (weak) partial derivative in the direction $\left({e_1}^{\alpha_1}, \ldots, {e_n}^{\alpha_n}\right)'$. Let $\mathscr{D}$ be a domain in $\mathbf{R}^n$ and $1 \leq p \leq \infty$. The *Sobolev space* $W^{k,p}(\mathscr{D})$ consists of all locally integrable (i.e., integrable over all compact sets) functions $f : \mathscr{D} \to \mathbf{R}$ such that for each multi-index α with $|\alpha| \leq k$, $D^\alpha f$ exists in the weak sense and belongs to $L^p(\mathscr{D})$. $W^{k,p}(\mathscr{D})$ is the Sobolev space of (integer) order k over $\mathscr{D}$. If it only belongs to $L^p_{\text{loc}}(\mathscr{D})$, we obtain the *local Sobolev space* $W^{k,p}_{\text{loc}}(\mathscr{D})$. For $f \in W^{k,p}(\mathscr{D})$ we define its norm to be

$$\|f\|_{W^{k,p}(\mathscr{D})} := \begin{cases} \left(\sum_{|\alpha| \leq k} \|D^\alpha f\|^p_{L^p(\mathscr{D})} \right)^{1/p}, & 1 \leq p < \infty \\ \sum_{|\alpha| \leq k} \operatorname{ess\,sup}_{\mathscr{D}} |D^\alpha f|, & p = \infty \end{cases}.$$

Sobolev spaces have a nice mathematical structure: for each $k \in \mathscr{N}$ and $1 \leq p \leq \infty$, the Sobolev space $W^{k,p}(\mathscr{D})$ is a Banach space. The special case $W^{k,2}(\mathscr{D})$ is a Hilbert space. The notion of Sobolev spaces can be extended to non-integer orders [50], which, all together, yield a class of function spaces with continuously parametrized degree of smoothness. The scale of fractional Sobolev spaces provides a suitable framework for the local characterization of fractal functions, since the regularity properties of functions in the spaces of such a scale interpolate the local regularity properties displayed by the functions in Sobolev spaces of integer order. Specifically, in the case $\mu \in (0, n)$, we have continuity for $s > n/2$, but not differentiability [48]. The embedding of fractional Sobolev spaces of order $s > n/2$ in the Hölder-Zygmund space of continuous functions of order $s - n/2$ allows the introduction of fractal functions in the framework of fractional Sobolev spaces. The equivalence between the norm generated by the covariance function and the norm of a fractional Sobolev space of order $s \in (n/2, n)$ is fundamental in the definition of fractal random fields [42–44]. Specifically, such an equivalence leads to the Hölder

continuity of order $s - n/2$, in the mean-square sense, as well as to the Hölder continuity of fractional order (fractality) of the sample paths, in the Gaussian case (see [44]).

An alternative characterization of $W^{\mu,2}(\mathbf{R}^n)$ as (cf. [57, p. 141]) will be useful throughout the chapter:

$$W^{\mu,2}(\mathbf{R}^n) = \left\{ f \in L^2(\mathbf{R}^n) : \hat{f}(\cdot)(1 + \|\cdot\|^2)^{\mu/2} \in L_2(\mathbf{R}^n) \right\}. \tag{9.6}$$

We shall also make use of the following useful result, for which we refer to Scheuerer [50] and the references therein.

Theorem 9.1. *Suppose that $C \in L_1(\boldsymbol{R}^n) \cap \mathscr{C}(\boldsymbol{R}^n)$ has a Fourier transform f that satisfies*

$$c_1 \left(1 + \|\omega\|^2\right)^{-\tau} \;\le\; f(\omega) \;\le\; c_2 \left(1 + \|\omega\|^2\right)^{-\tau}, \qquad \omega \in \boldsymbol{R}^n$$

with $\tau > \frac{n}{2}$ and two positive constants $c_1 \le c_2$. Then the RKHS $\mathscr{H}_C$ coincides with the Sobolev space $W^{\tau,2}(\boldsymbol{R}^n)$, and the norms $\|\cdot\|_{\mathscr{H}_C}$ and $\|\cdot\|_{W^{\tau,2}(\boldsymbol{R}^n)}$ are equivalent.

The equivalence of the norms discussed above should not be confused with that of Gaussian measures, this last concept being the basis for the theory of optimal asymptotic prediction as discussed in [55, 60] and to which the reader is referred.

The main imbedding results can be resumed as in the following [50].

Theorem 9.2. *Let $\mathscr{D}$ be a bounded $\mathscr{C}^\infty$ domain in R^n. Then, for $\mu > k + \frac{n}{2}$ we have the implication*

$$f \in W^{\mu,2}(\mathscr{D}) \quad \Longrightarrow \quad \exists\, \tilde{f} \in \mathscr{C}^k(\overline{\mathscr{D}}) \;\; \textit{so that} \;\; \tilde{f} = f \quad \textit{a.e. on } \mathscr{D}.$$

A more general result can be obtained by considering Hölder spaces $\mathscr{C}^{k,\beta}(\overline{\mathscr{D}})$ equipped with the norm

$$\|f\|_{\mathscr{C}^{k,\beta}(\overline{\mathscr{D}})} := \sum_{|\alpha| \le k} \|D^\alpha f\|_{\mathscr{C}(\overline{\mathscr{D}})} + \sum_{|\alpha| = k} |D^\alpha f|_{\mathscr{C}^{0,\beta}(\overline{\mathscr{D}})},$$

where for $0 < \beta \le 1$ we define the β^{th} Hölder seminorm of $f : \mathscr{D} \to \mathbf{R}$ by

$$|f|_{\mathscr{C}^{0,\beta}(\overline{\mathscr{D}})} = \sup_{\substack{s,t \in \mathscr{D} \\ s \ne t}} \frac{|f(t) - f(s)|}{\|t - s\|^\beta}.$$

Theorem 9.3. *Let $\mathscr{D}$ be a bounded $\mathscr{C}^\infty$ domain in R^n, further let $k \in \mathscr{N}_0$ and $0 < \beta < 1$. Then, for $\mu > k + \beta + \frac{n}{2}$ we have the implication*

$$f \in W^{\mu,2}(\mathscr{D}) \quad \Longrightarrow \quad \exists\, \tilde{f} \in \mathscr{C}^{k,\beta}(\overline{\mathscr{D}}) \;\; \textit{so that} \;\; \tilde{f} = f \quad \textit{a.e. on } \mathscr{D}.$$

9.3 Results on Some Families of Covariance Functions

This section describes six (labeled I-VI) families of covariance functions.

(I) The generalized Cauchy function (e.g., [14], [16]) *i.e.*

$$C_{\alpha,\gamma}(\xi) := (1 + \|\xi\|^{\alpha})^{-\gamma}, \tag{9.7}$$

which is positive definite on $\mathbf{R}^n$ for all $n \in \mathbf{N}$, for $0 < \alpha \leq 2$ and $\gamma > 0$. Special cases of this class will be also of interest. In particular, $C_{2,\gamma}$ is the characteristic function of the symmetric Bessel distribution, $C_{\alpha,\alpha}$ is the characteristic function of the Linnik distribution, and $C_{1,\gamma}$ is the symmetric generalized Linnik characteristic function. For these special cases, according to [48] we shall use the restriction $\gamma \in (0, n)$, for n the dimension of the Euclidean space associated to the argument ξ, for technical reasons that are thoroughly explained therein. Note that for this and most of the models we consider here, it is possible to add a variance parameter θ and a range parameter ϕ to the model by considering $C_{\alpha,\gamma,\theta,\phi} = \theta(1 + \|\xi/\phi\|^{\alpha})^{-\gamma}$ for θ and ϕ positive and it is this form of the model that would generally be used in practice.

The generalized Cauchy class represents a breaking point with respect to earlier literature based on the dogmatic assumption of self similarity, since it decouples the fractal dimension and the Hurst effect [16]. Gneiting and Schlather [16] argue convincingly that the generalized Cauchy RF allows independent treatment of the parameters identifying the fractal dimension and the Hurst effect. Then, in [28], the same class of covariance functions is considered to study the local and global properties of the associated Gaussian RF and Gaussian sheet. Also, the properties of the associated self-similar RF, obtained through Lamperti transformations, are studied therein.

In a more recent contribution, Lim and Teo [28] consider the generalized Cauchy class and show that, according to Kent and Wood [25], the Gaussian RF with this covariance is locally self similar of order $\alpha/2$ and that its tangent field is the Lévy fractal Brownian field of order $\alpha/2$. They then give a formal justification of the decoupling effect by showing that the model in equation (9.7) is long range dependent if and only if $0 < \alpha\gamma < n$. To show this, they appeal to the integral identity [18]

$$\int_0^{\infty} x^{\mu-1}(1 + x^{\rho})^{-\nu} \mathrm{d}x = \frac{1}{\rho} B\left(\frac{\mu}{\rho}, \nu - \frac{\mu}{\rho}\right),$$

for B the Beta function, and to integration in polar coordinates. Thus, the reparametrization

$$C_{\alpha,\gamma}(\xi) := (1 + \|\xi\|^{\alpha})^{-\gamma/\alpha} \tag{9.8}$$

gives a model that decouples the fractal dimension and the Hurst effect.

Lim and Teo also study the local and long tail behavior of the associated spectra, which we denote hereafter as $f_{\alpha,\gamma}$. In particular, as referenced by the same authors, a detailed study of these spectral densities has been carried out by Kotz et al. [26] for $0 < \alpha < 2$, $\gamma = 1$ and $n = 1$; by Ostrovskii [38] for $0 < \alpha < 2$, $\gamma = 1$ and $n \in \mathbf{N}$; and by Erdogan and Ostrovskii [11] for $0 < \alpha < 2$, $\gamma > 0$ and $n = 1$. Lim and Teo derive the asymptotic properties of the spectral densities $f_{\alpha,\gamma}$ for $0 < \alpha < 2$, $\gamma > 0$ and $n \in \mathbf{N}$, which can be regarded as an extension to the results on generalized multivariate Linnik distributions. The proofs of their results are very beautiful and the reader is referred to their paper for them. Here, we report and rephrase the results for the sake of completeness. The authors show that $f_{2,\gamma}(\omega) = c(\gamma)\|\omega\|^{\gamma-n}(1-\theta(\omega))$ where $\theta(\omega) \to 0$ as $\|\omega\| \to 0$, and $c(\gamma) = \frac{1}{2\Gamma(\gamma)\cos(\gamma\pi/2)}$. Also, for $\|\omega\| \to \infty$, we have $f_{2,\gamma}(\omega) = \frac{1}{2^{\gamma/2}\Gamma(\gamma/2)}\|\omega\|^{\gamma-\frac{n+1}{2}} \exp(-\|\omega\|)\left(1 + \frac{\gamma(\gamma-2)}{8\|\omega\|} + \dots\right)$. Moreover, for the spectral densities $f_{\alpha,\alpha}$ and $f_{1,\gamma}$ the following local and asymptotic identities hold: when $\|\omega\| \to 0$, $f_{\alpha,\alpha}(\omega) = c(\alpha)\|\omega\|^{\alpha-n}(1-\theta(\omega))$ and $f_{1,\gamma}(\omega) = c(\gamma)\|\omega\|^{\gamma-n}(1-\theta(\omega))$, where $\theta(\omega) \to 0$ as $\|\omega\| \to 0$. Finally, for $\|\omega\| \to \infty$, $f_{\alpha,\alpha}(\omega) = c_\alpha\|\omega\|^{-n-\alpha}(1+o(1))$, and $f_{1,\gamma}(\omega) = c_\gamma\|\omega\|^{-2}(1+o(1))$. The main result is eventually contained in the following

Proposition 9.1 (Lim and Teo, 2008 [28]).

$$f_{\alpha,\gamma}(\omega) = c(\gamma)\|\omega\|^{\gamma-n}(1-\theta(\omega)), \quad \text{where} \quad \theta(\omega) \to 0 \quad \text{as} \quad \|\omega\| \to 0 \tag{9.10}$$

$$f_{\alpha,\gamma}(\omega) = c_1\|\omega\|^{-n-\alpha}(1+o(1)), \quad \|\omega\| \to \infty, \tag{9.11}$$

Lim and Teo then show that, for $0 < \alpha < 2$ and for any positive β,

$$f_{\alpha,\beta}(\omega) = -\frac{\|\omega\|^{\frac{2-n}{2}}}{2^{\frac{2-n}{2}}\pi^{\frac{2+n}{2}}}\mathrm{Im}\int_0^\infty \frac{\mathscr{K}_{(n-2)/2}(\|\omega\|u)}{\left(1+\mathrm{e}^{i\pi\alpha/2}u^\alpha\right)^\beta}u^{n/2}du,$$

where $\mathscr{K}_\nu$ is the modified Bessel function of the third kind of order ν. Gneiting and Schlather [16] demonstrate decoupling of the generalized Cauchy model through a simulation study. They show that with sufficiently large sample sizes on a regular grid spread over a large enough domain, it is possible to estimate both α and γ in (9.8) well using spectral methods. Thus, it is not always necessary to assume Gaussian RFs are self-similar in order to obtain good estimates of their behavior at both small and large scales.

Ruiz Medina *et al.* [48] give further properties and characterizations of a Gaussian RF with a Cauchy covariance. In particular, they obtain a local pseudodifferential representation in terms of the local regularity properties of the functions in its RKHS. Namely, they show that the local behavior of the functions in the RKHS of the Cauchy Gaussian RF coincides with the local behavior of the functions in the fractional Sobolev space $W^{\frac{n+\alpha}{2},2}(\mathbf{R}^n)$. Thus,

the Gaussian RF with a Cauchy covariance as in equation (9.8) admits the following local pseudodifferential representation, in the mean-square sense:

$$(-\Delta)^{\frac{n+\alpha}{4}} X = \varepsilon, \tag{9.12}$$

where Δ denotes the Laplacian operator, and, here and subsequently, ε denotes Gaussian white noise.

As a corollary of this result, the fractional mean quadratic variation order, as well as the modulus of continuity of the sample paths, is derived by the same authors, who show that the Gaussian RF with a Cauchy covariance has mean quadratic variation order given from the identity

$$E[X(\xi+\xi_0)-X(\xi)]^2 = \mathscr{O}\left(\|\xi_0\|^{\alpha}\right), \quad \|\xi_0\| \to 0, \quad \forall \xi \in \mathbf{R}^n. \tag{9.13}$$

Also, with probability one, for any $\epsilon > 0$, the following inequality holds for its sample paths

$$\operatorname*{Sup}_{|\xi_0|<\delta} |X(\xi+\xi_0)-X(\xi)| \leq Y\delta^{\frac{\alpha-\epsilon}{2}}, \quad \delta \to 0, \quad \forall \xi \in \mathbf{R}^n, \tag{9.14}$$

where Y is an almost surely finite random variable.

(II) The Moak class [33] of completely monotonic functions has the form

$$\frac{1}{\|\xi\|^{\alpha}(1+\|\xi\|^2)}, \quad \alpha \geq 0 \tag{9.15}$$

for which several nice contributions can be found about its complete monotonicity (cf. Berg *et al.*, [4], and references therein). For $\alpha > 0$, this function has a singularity at the origin. Thus, it cannot be the covariance of a Gaussian RF defined in the ordinary sense. This class will be reconsidered subsequently in point (IV) below.

(III) For $\nu > 0$ the *Matérn* covariance function is given by

$$C_{\phi,\nu}(\xi) := \left(\frac{\|\xi\|}{\phi}\right)^{\nu} \mathscr{K}_{\nu}\left(\frac{\|\xi\|}{\phi}\right). \tag{9.16}$$

A huge literature for this covariance function can be found in spatial statistics and we refer the reader to Matérn [34] for an early treatment, to Stein [55] for a discussion of its importance as a flexible model for the local behavior of a Gaussian RF, and to the excellent survey in Gneiting and Guttorp [17] for an historical view of the use of this kernel in several branches of science. The Matérn kernel has corresponding spectral density with respect to Lebesgue measure (up to a multiplicative constant)

$$f_\nu(\omega) = \left(\phi^{-2} + \|\omega\|^2\right)^{-\nu - n/2}. \tag{9.17}$$

In the numerical analysis literature this kernel is also called the *Sobolev kernel* [6]. Its value in both numerical analysis and spatial statistics is that the parameter ν quantifies the smoothness of the associated RKHS and the associated random field.

Note that the kernel $C_{\phi,\nu}$ itself (and hence any finite linear combinations of kernel translates) is contained in the Sobolev space $W^{\mu,2}(\mathbf{R}^n)$ if and only if $\mu < 2\,\nu - \frac{n}{2}$, which directly follows from

$$\left(1 + \|x\|^2\right)^{-s} \in L^1(\mathbf{R}^n) \quad \Longleftrightarrow \quad s > \tfrac{n}{2},$$

and from the alternative characterization of Sobolev spaces given in equation (9.6). A complementary and nice result is given in Scheuerer [50]: for the Matérn kernel $C_{\phi,\nu}$ it holds that

$$C_{\phi,\nu} \in \mathscr{C}^{k,\beta}_{\mathrm{loc}}(\mathbf{R}^n) \quad \Longrightarrow \quad 2\nu \geq k + \beta + n, \qquad k = 0, 1, \quad 0 < \beta \leq 1.$$

For $k = 1$ and $\beta = 1$ we even have the strict inequality

$$C_{\phi,\nu} \in \mathscr{C}^{1,1}_{\mathrm{loc}}(\mathbf{R}^n) \quad \Longrightarrow \quad 2\nu > 2 + n.$$

Whittle [58] shows that the covariance in equation (9.16) is generated by the stochastic differential equation

$$\left(-\Delta + \lambda^2\right)^{\nu/2} Y(x) = \varepsilon(x). \tag{9.18}$$

Lim and Teo [29] consider a stochastic differential equation with two fractional orders

$$\left((-\Delta)^\alpha + \lambda^2\right)^{\nu/2} Y(x) = \varepsilon(x),$$

generating the spectra $f_{\alpha,\nu}(\omega) = (2\pi)^{-n}\left(\|\omega\|^{2\alpha} + \lambda^2\right)^{-\nu}$, which is square integrable on the real line if and only if $\alpha\nu > n/2$, so that the associated field is well defined in the ordinary sense for this range of parameters. Otherwise, the associated RF can be regarded as a generalized RF in the Schwartz space of test functions. Lim and Teo analyze the asymptotic behaviour of the covariance function $C_{\alpha,\nu}$ given by the Fourier transform of $f_{\alpha,\nu}$. [29] show the Fourier representation of $C_{\alpha,\nu}$. A very interesting result contained therein asserts that, for $\|\xi\| \to \infty$, the leading term of $C_{\alpha,\nu}$ is proportional to $\|\xi\|^{-2\alpha-n}$ when α is not an integer. To study the local behavior of the RF, one needs to distinguish several subcases of the parameter space.

(IV) Let us recall the equation of the Dagum function family

$$\xi \mapsto K_{\beta,\gamma}(\xi) := 1 - \left(\frac{|\xi|^\beta}{1 + \|\xi\|^\beta}\right)^\gamma, \tag{9.19}$$

for which sufficient conditions of complete monotonicity are $\beta\gamma \leq 1$ and $\beta < 1$. The characterization is not complete yet, as explained in [4], although the authors conjecture that this condition is also necessary. Arguments in [40] show that the Dagum function is positive definite on any n-dimensional Euclidean space for $\beta \in (0, 2]$ and $\gamma \in (0, 1]$. Larger ranges of permissibility of these parameters are given in Berg *et al.* [4] where the authors give an almost complete picture for the permissibility of this class. A simulation-based study in the spirit of [16] highlights, after a suitable reparametrization, an explicit decoupling of the local behavior and long-range dependence in the Dagum covariance function.

We begin by illustrating the relationship between the Dagum function and what has been called in [48] *auxiliary* family of generalized covariance functions, hereafter denoted as $C_{\alpha,\beta,\gamma}$ and whose equation is

$$\widetilde{C}_{\alpha,\beta,\gamma}(\xi) := \left(\frac{1}{\|\xi\|^{\alpha}(1 + \|\xi\|^{\beta})}\right)^{\gamma/\beta}. \tag{9.20}$$

In particular, in the results derived below, we work under the choice $\alpha = (n - \beta\gamma)/(1 + \gamma)$ and $\gamma/\beta = 1 + \gamma$, so that $C_{\alpha,\beta,\gamma}$ can be easily seen as the first derivative of the Dagum family $K_{\beta,\gamma}$ with respect to $|\xi|$. This is not a mere algebraic fact, since it constitutes a key point for studying the properties of a Dagum Gaussian RF in relation to those Gaussian RFs whose covariance function is a special case of the auxiliary family above. It is worth mentioning that the generalized Cauchy class and its related particular cases are obtained as special cases of the auxiliary family, since $C_{\alpha,\gamma} = \widetilde{C}_{0,\alpha,\gamma}$. Also, the Moak class in equation (9.15) corresponds to the choice $\widetilde{C}_{\alpha,2,2}$.

For this generalized covariance, [48] show that for $\gamma\beta \in (0, n)$, the following local and asymptotic behaviors hold:

$$\widetilde{f}_{\alpha,\beta,\gamma}(\omega) = \mathscr{O}\left(\frac{1}{1 + \|\omega\|^{\beta}}\right), \quad \|\omega\| \to 0 \qquad \text{and}$$

$$\widetilde{f}_{\alpha,\beta,\gamma}(\omega) = \mathscr{O}\left(\|\omega\|^{-\gamma\beta}\right), \quad \|\omega\| \to \infty.$$

where $\widetilde{f}_{\alpha,\beta,\gamma}$ is the generalized Fourier transform of $\widetilde{C}_{\alpha,\beta,\gamma}$.

A relevant remark concerns the fact that the Cauchy, Dagum and auxiliary families do not admit a closed form for the associated (ordinary or generalized) spectra. [48] show that, for $0 < \alpha\gamma < n$,

$$g_{\beta,\gamma}(\omega) = \mathscr{O}\left(\|\omega\|^{-n+\beta\gamma}\right), \quad \|\omega\| \to 0 \qquad \text{and}$$

$$g_{\beta,\gamma}(\omega) = \mathscr{O}\left(\|\omega\|^{-\gamma\beta-n}\right), \quad \|\omega\| \to \infty, \tag{9.21}$$

for $g_{\beta,\gamma} := \mathscr{F}[K_{\beta,\gamma}]$

The local pseudodifferential representation and local RKHS characterization of Dagum Gaussian and auxiliary generalized RFs is provided in the following result, also due to [48], where it is shown that the local behavior of the functions in the RKHS of the Dagum Gaussian RF can be identified with the local behavior of the functions in the fractional Sobolev space $W^{(n+\gamma\beta)/2,2}(\mathbf{R}^n)$, continuously embedded into the Hölder-Zygmund space of fractional order $\gamma\beta/2$. This means that the following local identity holds in the mean-square sense:

$$(-\Delta)^{(n+\gamma\beta)/4} X = \varepsilon.$$

Finally, [48] show that the mean quadratic variation order of the Dagum Gaussian RF X is given by

$$E[X(\xi+\xi_0) - X(\xi)]^2 = \mathcal{O}\left(\|\xi_0\|^{\gamma\beta}\right), \quad \|\xi_0\| \to 0, \quad \forall \xi \in \mathbf{R}^n. \tag{9.22}$$

That is, X is an index-$(\gamma\beta)/2$ RF. Therefore, the Dagum Gaussian RF satisfies a stochastic Hölder condition of order α for every $\alpha < \gamma\beta$. Equivalently, with probability one, for any $\epsilon > 0$, the following inequality holds for its sample paths

$$\operatorname*{Sup}_{|\xi_0|<\delta} |X(\xi+\xi_0) - X(\xi)| \leq Z\delta^{\frac{\gamma\beta-\epsilon}{2}}, \quad \delta \to 0, \quad \forall \xi \in \mathbf{R}^n, \tag{9.23}$$

where Z is an almost surely finite random variable.

(V) Compactly supported correlation functions.

The Wendland class of correlation functions has been repeatedly used in applications involving the so-called tapered likelihood [10]. We recall here Wendland's [57] construction (it is rephrased in Gneiting [15]). Let

$$\psi_{\nu,0,\beta}(x) := \left(1 - \frac{\|\xi\|}{\beta}\right)_+^{\nu}, \qquad \xi \in \mathbf{R}^n, \quad \nu \in \mathbf{R}_+ \quad \text{and} \quad \beta > 0, \tag{9.24}$$

be the truncated power function, also known as the Askey function when ν is an integer; it is compactly supported over a ball in $\mathbf{R}^n$ with radius β, and $\psi_{\nu,0,\beta}$ is positive definite on $\mathbf{R}^n$ for $\nu \geq \lfloor \frac{n}{2} \rfloor + 1$, where $\lfloor a \rfloor$ denotes the greatest integer less than or equal to a. Many applications refer to the exponent $\nu \in \mathbf{N}$, while the real part of the exponent is a basic characteristic of the associated Sobolev space that determines the regularity properties of a Gaussian RF with such covariance structure. For pertinent results on this topic see Wendland [57] and, for the cases not treated there, Schaback's [49] complementary contribution. Note that we have defined ψ over $x \in \mathbf{R}^n$, but in fact it could as well be defined over $y = \|x\| \in \mathbf{R}_+$, and the notation often exploits this varied use while maintaining consistency of meaning for ψ.

For any $g \in \Phi(\mathbf{R}^n)$ for which $\int_{\mathbf{R}_+} u\, g(u)\, \mathrm{u} < \infty$, Matheron's montée operator I is defined by

$$I_g(t) = \frac{\int_t^\infty u\, g(u)\, du}{\int_0^\infty u\, g(u)\, du} \qquad (t \in \mathbf{R}_+).$$

Wendland defines $\psi_{\nu,k,\beta}$ via k-fold iterated application of the montée operator on the Askey function $\psi_{\nu,0,\beta}(\xi)$ defined in (9.24). Wendland [57] proves that $\psi_{\nu,0,\beta} \in \Phi(\mathbf{R}^n)$ for $\psi_{\nu,k,\beta} \in \Phi(\mathbf{R}^{n-2k})$. The implications in terms of differentiability of $\psi_{\nu,k,\beta}$ are well summarized by Gneiting [15], and Wendland [57] shows that the degree of the piecewise polynomials is minimal for the given smoothness and dimension for which the radial basis function should be positive definite.

The following beautiful result allows for a characterization of the Fourier transform related to Wendland functions. Its proof can be found in ([57], Thm 10.35) but the assertion and notation are rephrased here for consistency with the rest of the chapter. For what follows, let us denote by $\widehat{\psi}_{\nu,k,\beta}$ the Fourier transform of $\psi_{\nu,k,\beta}$.

Theorem 9.4. *Let $n \geq 3$ if $k = 0$. Then there exists constants c_1, c_2 depending exclusively on n and k such that*

$$c_1(1 + \|\omega\|)^{-n-2k-1} \leq \widehat{\psi}_{\lfloor d/2 \rfloor + k + 1, k, 1}(\omega) \leq c_2(1 + \|\omega\|)^{-n-2k-1},$$

i.e. Wendland functions are defined on the classical Sobolev space $W^{d/2+k+1/2,2}(\boldsymbol{R}^n)$.

The complementary cases are completed in the recent contribution by Schaback [49].

(VI) Some new covariance functions.

None of the covariance functions we have described enjoy both the flexibility of the Matérn model at short scales together with that of the generalized Cauchy or Dagum families at long scales. Specifically, by adjusting the parameter ν in the Matérn model, we can get any degree of mean-squared differentiability in the corresponding random field. However, no Matérn model exhibits long-range dependence; indeed, every Matérn covariance function decays exponentially as its argument increases. In contrast, the generalized Cauchy and Dagum families can both attain any degree of long-range dependence, but correspond either to processes with no mean-squared derivatives or to processes that are infinitely mean-squared differentiable (for $\alpha = 2$ in Cauchy case and $\beta = 2$ in Dagum case).

There are a number of ways to obtain families of covariance functions that allow the local flexibility of the Matérn class and freedom in the degree of long-range dependence. A simple solution is just to add a Matérn covariance and a Cauchy model with $\alpha = 2$:

$$C(\xi) = \theta_1 \|\xi/\phi_1\|^{\nu} \mathscr{K}_{\nu}(\|\xi\|/\phi_1) + \frac{\theta_2}{(1 + \|\xi/\phi_2\|^2)^{\gamma}}, \tag{9.25}$$

where θ_1, θ_2 nonnegative and $\phi_1, \phi_2, \nu, \gamma$ positive are sufficient to make C positive definite in any number of dimensions n. Note that because we are taking a sum of positive definite functions and there is no prior basis for fixing the relative contribution of the two terms, we have included variance parameters θ_1 and θ_2 in the model. We also include a range parameter for each component of the model (ϕ_1 and ϕ_2), although one might be able to make a case for setting $\phi_1 = \phi_2$ if one wanted to simplify the model. Because all Matérn models decay exponentially as $\|\xi\|$ increases, the first term on the right side of (9.25) has no impact on the long-range dependence of the model (9.25). Similarly, because the Cauchy models with $\alpha = 2$ all correspond to infinitely differentiable processes, as long as $\theta_1 > 0$, the second term on the right side of (9.25) has no impact on the local behavior of the model (9.25).

Even if we set $\phi_1 = \phi_2$, the model in (9.25) has five parameters, which is arguably one more than necessary. Specifically, if we allow one variance parameter and one range parameter, we should be able to attain any degree of local behavior and any degree of long-range dependence with two additional parameters for a total of four. In fact, we can specify such a four-parameter model in the spectral domain by

$$f(\omega) = \frac{\theta \|\omega\|^{\alpha - n}}{(1 + \|\omega/\phi\|^2)^{\kappa}}, \tag{9.26}$$

which, for θ and ϕ positive, is integrable over $\mathbf{R}^n$ if and only if $0 < \alpha < 2\kappa$. For any n, the corresponding covariance function can be written in terms of the generalized hypergeometric function ${}_1F_2$ ([18], 6.565.8). The parameter α controls the long-range behavior of the process (for $\alpha \in (0, n)$, the resulting covariance function has Hurst coefficient $\frac{1}{2}(n - \alpha)$) and $\kappa - \frac{1}{2}\alpha$ controls its local behavior, playing essentially the role of ν in the Matérn model.

References

1. Adler, R.J.: The Geometry of Random Fields, Chichester: Wiley (1981)
2. Anh, V.V., Ruiz-Medina, M.D. and Angulo, J.M.: Covariance factorization and abstract representation of generalized random fields. Bull. Autral. Math. Soc. **62**, 319–334 (2000)
3. Ayache, A., Xiao, Y.: Asymptotic properties and Hausdorff dimensions of fractional Brownian sheets, J. Fourier Anal. Appl. **11**, 407–439 (2005)
4. Berg, C., Mateu, J., Porcu, E.: The Dagum family of isotropic correlation functions. Bernoulli **14**, (4), 1134–1149 (2008)
5. Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular Variation. Cambridge University Press, New York (1987)
6. Buhmann, M. A new class of radial basis functions with compact support. Math. Comp. **70**, 307–318 (2001)

7. Christakos, G.: On the problem of permissible covariance and variogram models, Water Resour. Res. **20**, 2, 251–265 (1984)
8. Christakos G, Hristopoulos D.T.: Spatiotemporal environmental health modeling: a tractatus stochasticus, Kluwer, Boston.
9. Davies,S. and Hall, P. (1999), Fractal analysis of surface roughness by using spatial data, J. R. Statist. Soc. B **61**, 3–37
10. Du, J., Zhang, H. and Mandrekar, V. (2009). Infill asymptotic properties of tapered maximum likelihood estimators. Ann. Statist. **37** (6a) 3330–3361
11. Erdogan, M.B. and Ostrovskii, I.V. (1998), Analytic and asymptotic properties of generalized Linniks probability density, J. Math. Anal. Appl. **217**, 555–578
12. Falconer, K.J.: Fractal geometry, John Wiley & Sons Inc., Hoboken, NJ (2003)
13. Fernández-Pascual, R., Ruiz-Medina, M.D., Angulo, J.M.: Estimation of intrinsic processes affected by additive fractal noise. J. Multivariate Anal. **97** 1361–1381 (2006)
14. Gneiting, T.: Power-law correlations, related models for long-range dependence and their simulation. Journal of Appl. Probab. **37**, 1104–1109 (2000)
15. Gneiting, T.: Compactly supported correlation functions. J. Multivariate Anal. **83**, 493–508 (2002)
16. Gneiting, T., Schlather, M.: Stochastic models that separate fractal dimension and the Hurst effect. SIAM Rev. **46**, 269–282 (2004)
17. Guttorp, P., Gneiting, T.: Studies in the history of probability and statistics XLIX: On the Matérn correlation family. Biometrika **93**, 989–995 (2006)
18. Gradshteyn, I.S., Ryzhik, I.M.: Tables of integrals, series and products, sixth edition, Academic Press, San Diego (2000)
19. Guelfand, M., Vilenkin, N.Y.: Les Distributions **4**, Dunod, Paris (1967)
20. Hall, P., Wood, A. On the performance of box-counting estimators of fractal dimension. Biometrika **80**, 246–252 (1993)
21. Houdre, C., Villa, J.: An example of infinite dimensional quasi-helix, Contemporary Math. **336**, 195–201 (2003)
22. Ibragimov, I. A., Rozanov, Y. A. Gaussian Random Processes, trans. A. B. Aries. Springer, New York (1978)
23. Kato, T.: Perturbation Theory of Linear Operators. Springer, Berlin (1995)
24. Kang, SC., Koh, H.M., Choo, J.F.: An efficient response surface method using moving least squares approximation for structural reliability analysis. Probabilistic Engineering Mechanics, in press. (1995)
25. Kent, J.T., Wood, A.T.A.: Estimating fractal dimension of a locally self-similar Gaussian process by using increments. Statistics Research Report SRR 034–95, Centre for Mathematics and Its Applications, Austalia National University, Canberra (1995); J. R. Statist. Soc. B **59** 679–699 (1995)
26. S. Kotz, S. Ostrovskii, I.V., Hayfavi, A.: Analytic and asymptotic properties of Linniks probability density I, II. J. Math. Anal. Appl. **193**, 353–371, 497–521 (1995)
27. Leonenko, N.: Limit Theorems for Random Fields with Singular Spectrum, Kluwer Academic Publishers, Boston (1999)
28. Lim, S.C., Teo, L.P.: Gaussian fields and Gaussian sheets with generalized Cauchy covariance structure. Submitted to Arxiv, 2008, Arxiv number arXiv:0807.0022
29. Lim, S.C., Teo, L.P.: Generalized Whittle-Matern random field as a model of correlated fluctuations. Submitted to Arxiv, 2008, Arxiv number arXiv:0901.3581
30. Lim S.C., Li, M.: Generalized Cauchy process and its application to relaxation phenomena. J. Phys. A: Math. Gen. **39**, 2935–2951 (2006)
31. S.C. Lim, Muniandy, S.M.: (2003), Generalized Ornstein–Uhlenbeck processes and associated selfsimilar processes. J. Phys. A: Math. Gen. **36**, 3961–3982
32. Matheron, G. The intrinsic random functions and their applications. Adv. Appl. Probab. **5**(3), 439–468 (1973)
33. Moak, D.: Completely monotonic functions of the form $(1 + |x|)^{\beta}|x|^{\alpha}$. Rocky Mt. J. Math. **17**, 719–725 (1987)

34. Matérn, B.: Spatial Variation, 2nd edition. Berlin: Springer (1986)
35. Ostoja-Starzewski, M.: Microstructural randomness versus representative volume element in thermomechanics. ASME J. Appl. Mech. **69**, 25–35 (2002)
36. Ostoja-Starzewski, M.: Towards thermomechanics of fractal media. Journal of Applied Mathematics and Physics. **58**, 1085–1096 (2007)
37. Ostoja-Starzewski, M.: Microstructural Randomness and Scaling in Mechanics of Materials. Chapman & Hall/CRC Press (2008)
38. Ostrovskii, I.V.: Analytic and asymptotic properties of multivariate Linniks distribution. Math. Phys. Anal. Geom. **2**, 436–455 (1995)
39. Porcu, E., Mateu, J., Zini, A., Pini, R.: Modelling spatio-temporal data: A new variogram and covariance structure roposal. Stat. Probabil. Lett. **77**(1), 83–89 (2007)
40. Porcu, E., Mateu, J. and Nicolis, O.: A note on decoupling of local and global behaviours for the Dagum Random Field. Probabilist. Eng. Mech. **22**(4), 320–329 (2007)
41. Rytov, S.M., Kravtsov, Y.A., Tatarskii, V.I.: Principles of Statistical Radiophysics. **4**, Springer, Berlin (1989)
42. Ruiz-Medina, M.D., Angulo, J.M., Anh, V.V.: Fractional-order regularization and wavelet approximation to the inverse estimation problem for random fields. J. Multivariate Anal. **85**, 192–216 (2003a)
43. Ruiz-Medina, M.D., Angulo, J.M., Anh, V.V.: Stochastic Fractional-order differential models on fractals. Theor. Probab. Appl. **67**, 130–146 (2002)
44. Ruiz-Medina, M.D., Angulo, J.M., Anh, V.V.: Fractional generalized random fields on bounded domains. Stoch. Anal. Appl. **21**, 465–492 (2003b)
45. Ruiz-Medina, M.D., Angulo, J.M., Anh, V.V.: Karhunen-Loéve-Type representations on fractal domains. Stoch. Anal. Appl. **24**, 195–219 (2006)
46. Ruiz-Medina M.D., Anh V.V., Angulo, J.M. Fractal random fields on domains with fractal boundary. Infin. Dimen. Anal. Q.U. **7**, 395–417 (2004)
47. Ruiz-Medina, M.D., Angulo, J.M., Fernández-Pascual, R.: Wavelet-vaguelette decomposition of spatiotemporal random fields. Stoch. Env. Res. Risk. A. **21**, 273–281 (2007)
48. Ruiz Medina, M.D., Porcu, E., Fernandez-Pascual, R.: The Dagum and auxiliary covariance families: towards reconciling two-parameter models that separate fractal dimension and Hurst effect. Probabilistic Engineering Mechanics, **26**, 259–268 (2011)
49. Schaback, R.: The missing Wendland functions. Adv. Comput. Math., (to appear). DOI 10.1007/s10444-009-9142-7 (2009)
50. Scheuerer, M. A comparison of models and methods for spatial interpolation in statistics and numerical analysis. Doctoral thesis, University of Goettingen (2009)
51. Shkarofsky, I.P. Generalized turbulence space-correlation and wave-number spectrumfunction pairs. Can. J. Phys. **46**, 2133–53 (1968)
52. Schoenberg, I.J.: Metric Spaces and Completely Monotone Functions, Ann. Math. **39**(4) 811–841 (1938)
53. G. Samorodnitsky, Taqqu, M.: Stable non-Gaussian Random Processes, Chapman & Hall, London (1994)
54. Stein, M.L.: Asymptotically efficient prediction of a random field with a misspecified covariance function. Ann. Stat. **16**, 55–63 (1988a)
55. Stein, M.L.: Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York (1999)
56. Stein, M. L.: Equivalence of Gaussian measures for some nonstationary random fields. J. Stat. Plan. Infer. **123**, 1–11 (2004)
57. Wendland, H.: Scattered data Approximation, Cambridge University press (1994)
58. Whittle, P.: Stochastic processes in several dimensions. Bull. Inst. Int. Statist. **40**, 974–994 (1963)
59. Wood, A.T.A., Chan, G.: Increment-based estimators of fractal dimension for twodimensional surface data. Stat. Sinica **10**, 343–376 (1994)
60. Yadrenko, M.: Spectral Theory of Random Fields. New York, NY : Optimization Software (1983)

61. Yaglom, A.M.: An Introduction to the Theory of Stationary Random Functions, Dover Phoenix Editions (1987)
62. Zhang, H. (2004), Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics, J. Am. Stat. Assoc. **99**, 250–261.

Chapter 10
Asymptotics and Computation for Spatial Statistics

Hao Zhang

Abstract Asymptotics describe large sample properties of statistical inferences. Useful asymptotics are those that can help with the statistical inferences. For example, asymptotics can help identify statistically and computationally efficient estimators. The study of asymptotics in spatial statistics is complicated by the fact there are more than one asymptotic frameworks in spatial statistics and the asymptotic results are very different under the different asymptotic frameworks. This chapter reviews some results under these asymptotic frameworks and shows how the asymptotic results can help alleviate the computational challenges in the analysis of massive spatial data.

10.1 Introduction

In this chapter, we assume a spatial process $Y(\mathbf{s}), \mathbf{s} \in R^2$ is observed at n locations $\mathbf{s}_1, \cdots, \mathbf{s}_n$ in a bounded region $D \subset R^2$. This kind of geostatistical data arise in many disciplines in environmental, agricultural and atmospheric sciences. We are concerned of two particular problems in this section. One is the prediction of $Y(\mathbf{s})$ at a spatial location that is not observed. The best linear unbiased prediction is commonly referred to as kriging, which is computed by using either the covariogram or variogram. Recall that the variogram for a process is

$$\gamma(\mathbf{s}, \mathbf{x}) = \frac{1}{2} E(Y(\mathbf{s}) - Y(\mathbf{x}))^2 \tag{10.1}$$

and the covariogram is defined as

$$C(\mathbf{s}, \mathbf{x}) = Cov(Y(\mathbf{s}), Y(\mathbf{x})).$$

H. Zhang (✉)
Department of Statistics, Purdue University, USA
e-mail: zhanghao@purdue.edu

E. Porcu et al. (eds.), *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics 207, DOI 10.1007/978-3-642-17086-7_10,

Under the intrinsic stationarity, $\gamma(\mathbf{s}, \mathbf{x}) = \gamma(\mathbf{s} - \mathbf{x})$ and under the second order stationarity, $C(\mathbf{s}, \mathbf{x}) = C(\mathbf{s} - \mathbf{x})$ where we slightly abused the notation. We will consider only stationary processes and use $C(\cdot)$ and $\gamma(\cdot)$ to denote the stationary covariogram and variogram, respectively. Then we have $C(\mathbf{x}) = \sigma^2 - \gamma(\mathbf{x})$ where σ^2 is the variance. Quite often, we assume the covariogram belongs to a parametric family such as the Matérn model. We will estimate the model parameters by the maximum likelihood method or some other methods so that the plug-in kriging can be carried out. This leads to the second problem in this chapter, i.e., the parameter estimation. More specifically, we will study which parameters can be estimated well and what parameters are virtually impossible to be estimated well. We will also study what parameters are most important to kriging and what parameters are less important to kriging. The precise meanings of these last two sentences will become clear later as we use rigorous statistical terms.

The following is frequently encountered in the analysis of spatial data, which deserves some theoretical investigation. Two quite different covariograms could yield very similar prediction results; Estimators of some parameters in the covariogram tend to have large variances even when the sample size is sufficiently large, regardless of estimation method. These can be explained quite well by asymptotic results. [16] established rigorous mathematical results on kriging which we will review in Sects. 10.2.1 and 10.2.2. He employed the fixed-domain or infill asymptotic framework by assuming that the sampling domain D is bounded, i.e., $\sup_{x \in D} \|x\| < \infty$.

For parameter estimation, however, different asymptotic frameworks have been studied. When the sampling domain is bounded and more data are observed from this bounded region, this is the fixed domain asymptotic framework. No all parameters can be estimated consistently under the infill asymptotic framework [9, 22]. When the distance between any two sampling locations is bounded away from 0, this is the increasing domain asymptotic framework because the sampling domain have to increase in order for us to observe an arbitrarily large number of data. Under this asymptotic framework, all parameters can be estimated consistently under regularity conditions and the maximum likelihood estimators are asymptotically normal [13]. [9] considered a mixed asymptotic framework where the spatial domain is increasing but the minimum distance among the sampling distance also tends to 0. Since estimators have different asymptotic properties under the two asymptotic frameworks, it is important to choose an appropriate asymptotic framework for the practical problem at hand. [26] have done some theoretical investigation and numerical study to address this issue.

Recent studies on asymptotic properties of parameter estimation have brought about more understanding of the behavior of estimators of spatial model parameters and lead to estimation methods that can reduce computation in the estimation and prediction in spatial data analysis. These methods are particularly useful in the analysis of massive spatial data. The objective of this chapter is to review the asymptotic results that can help with the spatial data analysis.

10.2 Infill Asymptotic Framework

In this section, we assume that the spatial process $Y(\mathbf{s}), \mathbf{s} \in D \subset R^d$ is a stationary Gaussian process with mean 0 and a covariance function $C(\mathbf{h}, \theta)$ where θ is the parameter. We first review a probabilistic concept that has proved to be quite useful in the study of infill asymptotics.

10.2.1 Equivalent Probability Measures

Two probability measures P_0 and P_1 are equivalent on a measurable space $\{\Omega, \mathscr{F}\}$ if $\mathrm{P}_1(A) = 1$ for any $A \in \mathscr{F}$ implies $\mathrm{P}_0(A) = 1$ and vice versa. It says that if an event occurs with probability one under one measure then it occurs with probability one under the other measure. On the other hand, two measures are orthogonal if there exists an event A such that $P_1(A) = 1$ but $P_0(A) = 0$. We usually restrict the event A to the σ-algebra generated by $\{Y(\mathbf{s}), \mathbf{s} \in D\}$ where D is a subset of the space $\boldsymbol{R}^d$. We emphasize this restriction by saying the two measures are equivalent on the paths of $\{Y(\mathbf{s}), \mathbf{s} \in D\}$. If D is bounded, two Gaussian measures are either equivalent or orthogonal on the paths $\{Y(\mathbf{s}), \mathbf{s} \in D\}$ if $Y(\mathbf{s})$ has a continuous covariogram (This is not a necessary condition though).

Example 10.1. Let Y_k, $k = 1, 2, \ldots$, be i.i.d. under probability measure P_0 and P_1 and $E_i Y_k = 0$ and $E_i Y_k^2 = \sigma_i^2$ for any k, where the expectation E_i corresponds to P_i for $i = 0, 1$. If $\sigma_0^2 \neq \sigma_1^2$, the two measures are orthogonal because by the law of large numbers, $P_i(A) = i, i = 0, 1$ where A is the event that $\lim_{n\to\infty}(1/n)\sum_{k=1}^n Y_k^2 = \sigma_1^2$.

Example 10.2. Let $Y_i = Y(\mathbf{s}_i)$ where $\mathbf{s}_i \in D$ and $D \subset R^2$ is bounded, and the process $Y(\mathbf{s})$ be stationary Gaussian with mean 0 and an isotropic exponential covariogram $\sigma^2 \exp(-\alpha h)$ where σ^2 and α are the two parameters. Denote by P_i the Gaussian measure corresponding to parameters σ_i and α_i. Then even if $(\sigma_1^2, \alpha_1) \neq (\sigma_0^2, \alpha_0)$, the two measures could still be equivalent. Indeed, P_1 and P_0 are equivalent on the paths of $\{Y(\mathbf{s}_i), i = 1, 2, \cdots\}$ if and only if $\sigma_1^2\alpha_1 = \sigma_0^2\alpha_0$. Hence $P_1\{(1/n)\sum_{i=1}^n Y_i^2 \to \sigma_1^2\} = 1$ cannot be true.

In Example 2, if we reparameterize by letting $\vartheta = \sigma^2\alpha$ and write the model parameters as $\theta = (\vartheta, \alpha)$. If P_θ denotes the Gaussian probability measure corresponding to parameter θ and $\theta_i = (\vartheta_i, \alpha_i)$, $i = 0, 1$, then P_{θ_0} and P_{θ_1} are equivalent if and only if $\vartheta_1 = \vartheta_2$. Such a parameter ϑ is called a microergodic parameter and α is called a non-microergodic parameter. Non-microergodic parameters can not be estimated consistently. Only microergodic parameters can be estimated consistently.

Now consider the Gaussian process $Y(\mathbf{s})$ under two equivalent probability measures P_1 and P_0. Let $Y_i = Y(\mathbf{s}_i)$ for some locations $\mathbf{s}_i$, $i \geq 1$. Write

$$e_{1,i} = Y(\mathbf{s}_1),\ e_{k,i} = Y(\mathbf{s}_k) - \mathrm{E}_i[Y(\mathbf{s}_k)|Y(\mathbf{s}_l), 1 \leq l < k],\ k = 2, \cdots, n,\ i = 0, 1. \tag{10.2}$$

It is known (e.g., [15], p 501) that P_0 and P_1 are equivalent on $\mathscr{F} = \sigma(Y_i, i = 1, 2, \cdots,)$ if and only if

$$\sum_{n=2}^{\infty} \frac{\mathrm{E}_0[(e_{n,1} - e_{n,0})^2]}{\mathrm{E}_1 e_{n,1}^2} < \infty, \text{ and } \sum_{n=2}^{\infty} \left(\frac{\mathrm{E}_0 e_{n,0}^2}{\mathrm{E}_1 e_{n,1}^2} - 1 \right)^2 < \infty. \tag{10.3}$$

It follows (10.3) that

$$\lim_{n\to\infty} \frac{\mathrm{E}_1[(e_{n,1} - e_{n,0})^2]}{\mathrm{E}_0 e_{n,0}^2} = 0 \tag{10.4}$$

$$\lim_{n\to\infty} \frac{\mathrm{E}_1 e_{n,1}^2}{\mathrm{E}_0 e_{n,0}^2} = 1. \tag{10.5}$$

The last two equations have a nice interpretation. Since $e_{n,i}$ is the prediction error for predicting Y_n given $Y_k, k < n$, these two equations imply that prediction given under P_1 are asymptotically equally as good as those given under P_0.

In practice, we are interested in predicting $Y(\mathbf{s})$ at any location $\mathbf{s}$ not just at $\mathbf{s}_n$. Let $e_i(\mathbf{s}, n) = Y(\mathbf{s}) - E_i(Y(\mathbf{s})|Y(\mathbf{s}_k), k = 1, \cdots, n-1)$. A sensible measure for how good predictions given P_1 are when P_0 is the true model is

$$\sup_s \frac{\mathrm{E}_0(e_1(\mathbf{s}, n) - e_0(\mathbf{s}, n))^2}{\mathrm{E}_0 e_0(\mathbf{s}, n)^2},$$

i.e., how large the mean squared difference of predictions is relative to correct mean squared error, where the supremum is taken over $\mathbf{s}$ such that $\mathrm{E}_0 e_0(\mathbf{s}, n)^2 > 0$. Because the mean squared error is often calculated in practice, it is also of interest to compare the two MSEs by evaluating the ratio $\mathrm{E}_1 e_1(\mathbf{s}, n)^2 / \mathrm{E}_0 e_0(\mathbf{s}, n)^2$. The following theorem can be derived from (10.3) and is a special version of Stein ([16], 1999, p. 35).

Theorem 10.1. *Let P_0 and P_1 be two equivalent Gaussian probability measures on the paths of $Y(\mathbf{s}), \mathbf{s} \in D$, and the set of sampling sites $\{\mathbf{s}_k, k = 1, 2, \cdots\}$ is dense in D, where $D \subset \boldsymbol{R}^d$ is bounded, then uniformly in $\mathbf{s} \in D$ such that $\mathrm{E}_0 e_0(\mathbf{s}, n)^2 > 0$,*

$$\lim_{n\to\infty} \frac{\mathrm{E}_0(e_1(\mathbf{s}, n) - e_0(\mathbf{s}, n))^2}{\mathrm{E}_0 e_0(\mathbf{s}, n)^2} = 0 \tag{10.6}$$

$$\lim_{n\to\infty} \mathrm{E}_1 e_1(\mathbf{s}, n)^2 / \mathrm{E}_0 e_0(\mathbf{s}, n)^2 = 1. \tag{10.7}$$

So far we have assumed that the process is Gaussian. [2] established a result that shows that the predictive distribution under two equivalent measures will eventually agree as more data are collected. We now rephrase the theorem to make it directly applicable here. Let $Y_i,\ i \geq 1$ be random variables on a measurable space $(\Omega, \mathscr{F})$ and $P_i, i = 0, 1$ be two probability measures (not necessarily Gaussian) on $\mathscr{F}$ such that P_0 is absolutely continuous with respect to P_1 constrained on $\sigma(Y_i, i \geq 1)$, the σ-algebra generated by $Y_i, i \geq 1$. Then with P_0-probability 1,

$$\sup |P_0(A|Y_1, \cdots, Y_n) - P_1(A|Y_1, \cdots, Y_n)| \to 0, \text{ as } n \to \infty$$

where the supremum is taken over all $A \in \sigma(Y_i, i > n)$. In particular,

$$\sup_{i>n, B} |P_0(Y_i \in B|Y_1, \cdots, Y_n) - P_1(Y_i \in B|Y_1, \cdots, Y_n)| \to 0, \text{ as } n \to \infty. \tag{10.8}$$

The two measures are not necessarily Gaussian and the predictors are not necessarily linear. Hence (10.8) applies to non-linear prediction and implies the predictive distributions under the two measures are asymptotically equal.

Interestingly, Theorem 1 can be established from (10.8) . When the measures are Gaussian, the predictive distribution, i.e., the conditional distribution of Y_i given $Y_1, \cdots, Y_n$, is determined by the conditional mean (the predictor) and the conditional variance (the prediction variance). Then (10.8) can be given in terms of prediction error and the prediction variance, as given by (10.6) and (10.7).

Finally in this section, we review a sufficient condition for equivalence of Gaussian probability measures in terms of spectral density. Let the process $Y(\mathbf{s}), \mathbf{s} \in R^d$ be Gaussian stationary with mean 0 and have a spectral density $f_i(\boldsymbol{\lambda})$ under probability measure $P_i,\ i = 0, 1$. If, for some $a > 0$, $f_1^*(\boldsymbol{\lambda})|\boldsymbol{\lambda}|^a$ is bounded away from 0 and ∞ as $|\boldsymbol{\lambda}| \to \infty$, and for some finite c,

$$\int_{|\boldsymbol{\lambda}|>c} \left\{ \frac{f_1(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda})}{f_0(\boldsymbol{\lambda})} \right\}^2 d\boldsymbol{\lambda} < \infty, \tag{10.9}$$

then P_1 and P_0 are equivalent on the paths of $Y(\mathbf{s}), \mathbf{s} \in D$ for any bounded subset $D \subset R^d$.

When the spectral densities can be expressed in closed-form, it is possible to verify condition (10.9).

10.2.2 Spectral Densities and Kriging

In this section, we focus exclusively on linear prediction. Linear prediction only requires the first two moments and does not require the distributional properties. We

assume that the process is second order stationary with mean 0 and will carry out kriging using two covariograms, $C_i, i = 0, 1$. Assume that C_i has a spectral density $f_i(\lambda), i = 0, 1$. The following theorem of Stein (1999, p.135) provides a simpler condition in terms of spectral density for (10.6) and (10.7) to hold.

Theorem 10.2. *Let the underlying process $Y(\mathbf{s})$ be Gaussian under probability P_i with mean 0 and spectral density f_i, $i = 0, 1$. If for some $\varphi > 1$, $f_0(\boldsymbol{\lambda}) \|\boldsymbol{\lambda}\|^{\varphi}$ is bounded away from 0 and ∞ and*

$$\frac{f_1(\boldsymbol{\lambda})}{f_0(\boldsymbol{\lambda})} \to 1 \text{ as } \|\boldsymbol{\lambda}\| \to \infty,$$

then (10.6) and (10.7) hold.

10.2.3 Infill Asymptotics for the Matérn Model

The Matérn model for covariogram has been widely used primarily because it is capable of modeling the smoothness of the underlying process. It has a spectral density in a simple form

$$f(\boldsymbol{\lambda}) = \frac{\Gamma(\nu + d/2)}{\pi^{d/2}\Gamma(\nu)} \frac{\sigma^2\alpha^{2\nu}}{(\alpha^2 + \|\boldsymbol{\lambda}\|^2)^{\nu+d/2}}, \boldsymbol{\lambda} \in R^d, \tag{10.10}$$

where σ^2 is the variance, α is the scale parameter controlling how fast the correlation function decays to 0, and ν is the smoothness parameter. When $\nu > 1$, the process is mean square differentiable.

Consider two Matérn covariograms with parameters $(\sigma_i^2, \alpha_i, \nu), i = 0, 1$. It can be shown [22] that the two spectral densities satisfy condition (10.9). [22] showed the following result.

Theorem 10.3. *Let $P_i, i = 0, 1$ be two probability measures such that under P_i, the process $Y(\mathbf{s}), \mathbf{s} \in R^d$ is stationary Gaussian with mean 0 and an isotropic Matérn covariogram in R^d with a variance σ_i^2, a scale parameter α_i, $i = 0, 1$ and the same smoothness parameter ν, where $d = 1, 2$ or 3. For any bounded infinite set $D \subset R^d$, $P_1 \equiv P_2$ on the paths of $Y(\mathbf{s}), \mathbf{s} \in D$ if and only if $\sigma_1^2\alpha_1^{2\nu} = \sigma_2^2\alpha_2^{2\nu}$.*

There are two immediate corollaries. First, the parameter $\sigma^2\alpha^{2\nu}$ is microergodic and the individual parameter σ^2 and α are both non-microergodic. Hence both σ^2 and α are not consistently estimable if the sampling domain is bounded in R^d for $d \leq 3$. The inconsistency of the estimator for a non-microergodic parameter implies that the variance likely does not vanish as the sample increases. It certainly does not if the estimator is asymptotically unbiased. This may also explain why the profile likelihoods for these parameters tend to be flat as observed in many simulation studies and data analysis [22, 25].

Second, two Matérn covariograms with parameters $(\sigma_i^2, \alpha_i, \nu)$, $i = 0, 1$ yield asymptotically equal predictive distribution and asymptotically equally optimal. More precisely, (10.6) and (10.7) hold. This is not intuitive at all, but it explains why two different covariograms can yield very similar prediction results.

Therefore, from the point of view of prediction, we only need to estimate the microergodic parameters well. The non-microergodic parameters cannot be estimated well anyway.

In the higher dimensions, $d > 4$, Theorem 10.3 no longer holds. Indeed, both σ^2 and α are consistently estimable [1]. The case for $d = 4$ remains open.

10.3 Increasing Domain Asymptotic Framework

In the increasing domain asymptotic framework, the distance between any two sampling locations is bounded away from 0. Properties like (10.6) and (10.7) in general fail to hold. Therefore, properties of kriging are not studied under the increasing domain asymptotic framework. However, asymptotic properties of estimators can be given. Indeed, it is generally easier to study the asymptotic properties of estimators under the increasing domain asymptotic framework. One reason is that various mixing conditions can be given with the increasing domain asymptotic framework. Limit theorems have been established under the mixing conditions [7].

[13] gave some conditions under which the maximum likelihood estimators for the parameters in the probability distribution of a stationary spatial process are asymptotically normal. The asymptotic variance is given by the inverse of the Fisher information matrix.

[10] studied the asymptotic distribution of least squares estimators of variogram parameters under the increasing domain asymptotic framework.

10.4 Which Asymptotic Framework to Use?

In any real application, spatial data are observed at a finite number of points with no intention or possibility of taking more observations, and it is not clear which asymptotic framework to appeal to. Stein (1999) gives a cogent argument for using infill asymptotics if interpolation of the spatial process is the ultimate goal, as we discussed previously. However, for parameter estimation, it is not obvious which asymptotic framework should be used.

I take a pragmatic approach. I strongly believe that the reason to establish any asymptotic results is to ultimately apply them to a finite sample for statistical inferences. We hope that the asymptotic distribution of an estimator provides a good approximation to the finite sample distribution in the particular problem at hand. We

therefore choose a framework on the basis of how well the asymptotic distributions of estimators of parameters of interest approximate the finite-sample distributions of those estimators.

The finite sample distribution of an estimator can be obtained through simulations. [22] and [26] considered extensive simulation studies on exponential models. Their results show that the finite sample distribution of the MLE for the variance σ^2 or the scale parameter α appears to be skewed for moderately large sample size while the finite sample distribution of the MLE for the microergodic parameter $\sigma^2\alpha$ appears to be more symmetric and normal. This can be easily explained by the infill asymptotics though it is not obvious under the increasing domain asymptotics.

[26] considered a Gaussian process $Y(t), t \in R$ that has an isotropic exponential covariogram $\theta_1 \exp(-\theta_2 h), h \geq 0$. Asymptotic results for this model have been established under both asymptotic framework. [21] has shown that the following infill asymptotic distribution for the microergodic parameter $\theta_1\theta_2$. If the process is observed at n locations in $[0, 1]$,

$$\sqrt{n}(\hat{\theta}_1\hat{\theta}_2 - \theta_1\theta_2) \to N(0, 2(\theta_1\theta_2)^2).$$

If the process is observed at equally spaced locations $\delta i, i = 1, \cdots, n$ where $\delta > 0$ is a fixed constant, this is the increasing domain framework. The MLE has the following asymptotic framework

$$\sqrt{n}(\hat{\theta}_1\hat{\theta}_2 - \theta_1\theta_2) \to N(0, \sigma^2), \tag{10.11}$$

where

$$\sigma^2 = 2(\theta_1\theta_2)^2 + \{(1-\rho^2)(\delta\rho)^{-1} - 2\theta_2\rho\}^2\theta_1^2(1-\rho^2)^{-1} \tag{10.12}$$

and $\rho = \exp(-\theta_2\delta)$.

The two asymptotic distributions differ in the asymptotic variance. However, when both are applied to a finite sample, little difference will result. The reason is the following. Suppose we observed at n equally spaced locations in an interval $[0, a]$. The distance is usually rescaled by a linear transformation so that the maximum distance in the sampling domain is 1 [3]. Then the n locations can be denoted by $t_i = \delta i$ for $\delta = 1/n$. For this δ, the asymptotic variance σ^2 in (10.12) under the increasing domain asymptotic framework is approximately $2(\theta_1\theta_2)^2$. Indeed,

$$\sigma^2 \to 2(\theta_1\theta_2)^2, \text{ as } n \to \infty.$$

Therefore, there will be no or little difference which asymptotic result we choose to apply, either the infill or the increasing domain asymptotics.

However, for a non-microergodic parameter, there are big differences in the two asymptotic frameworks. We already know that the non-microergodic parameter is not consistently estimable under the infill asymptotic framework but is consistently

estimable in general under the increasing domain asymptotic framework. The asymptotic distribution of a non-microergodic parameter under the infill framework usually converges to a non-degenerate distribution though this asymptotic distribution is usually unknown except for the exponential model we discuss here. It is plausible that the MLE of α a sample at n locations in $[0, 1]$ has an infill asymptotic distribution, which is the distribution of

$$\hat{\phi}_{2,\infty} = -\frac{\int_0^1 Y(t)dY(t)}{\int_0^1 Y(t)^2 dt}. \tag{10.13}$$

$\hat{\phi}_{2,\infty}$ is the MLE of θ_2 when $Y(t)$ is observed everywhere in $[0, 1]$ (*e.g.* [12], Theorem 7.7, p. 248).

10.5 Asymptotics for Computation

Asymptotic results are particularly helpful for the statistical inferences if they lead to statistically and computationally efficient estimation. In particular, spatial data analysis usually involves more computation due to the spatial correlation and the fact that the spatial sample size can be huge. We now discuss some applications of infill asymptotics that reduce the amount of computation in the spatial data analysis.

10.5.1 Hybrid Estimation

[22] constructed a consistent estimator for the microergodic parameter $\sigma^2\alpha^{2\nu}$ in the Matérn model (10.10) for the dimension $d = 1, 2$ or 3. The idea is to fixed the non-microergodic parameter α at any chosen value α_1. The variance is estimated by maximizing the likelihood function in which the parameter α is fixed at α_1. The resulting estimate of σ^2 is given in closed-form

$$\hat{\sigma}^2 = \frac{1}{n} Y_n' R(\alpha_1)^{-1} Y_n$$

where $R(\alpha_1)$ is the correlation function that depends only on α_1 (the smoothness parameter is assumed to be known here). Zhang (2004) showed that $\hat{\sigma}^2\alpha_1^{2\nu}$ is an consistent estimator of $\sigma^2\alpha^{2\nu}$. This idea of fixing the non-microergodic parameter at a known value is used in the subsequent work [4, 8].

[4] established the infill asymptotic distribution of this estimator when $d = 1$:

$$\sqrt{n}(\hat{\sigma}^2\alpha_1^{2\nu} - \sigma_0^2\alpha_0^{2\nu}) \to N(0, 2(\sigma_0^2\alpha_0^{2\nu})^2). \tag{10.14}$$

The right hand of side is the asymptotic distribution of full MLE as established by [21]. Therefore there is no loss of statistical efficiency when we fix the non-microergodic parameter α.

In practice when we have a finite sample, the choice of α may affect the statistical efficiency. [27] suggested fix α at a rough estimator that is computationally easy to obtain. They used the weighted least squares estimator for α and estimate the microergodic parameter by maximizing the likelihood function when all the non-microergodic parameters are fixed at the rough estimates. They call it the hybrid estimation. The hybrid estimation requires a little more computation than the least squares method and has superior predictive performance. It requires much less computation than the MLE but has comparable predictive performance as the MLE.

10.5.2 Covariance Tapering

Covariance tapering is a computational technique to handle a large covariance matrix. Covariance tapering results in a sparse covariance matrix that takes less memory to store and to operate. Despite the computational advantage, the covariance tapering may not reduce the statistical efficiency of the estimators and the predictive performance.

The tapered covariogram is of the form

$$C(\mathbf{h}, (t)) = \rho(\mathbf{h}) K(\mathbf{h}, (t)) \tag{10.15}$$

where $K(\mathbf{h}, (t))$ is the covariance function of the underlying process that depends on a vector of parameter (t) and $\rho(\mathbf{h})$ is the taper, a known correlation function with a compact support. Hence $\rho(\mathbf{h})$ is zero when $\|\mathbf{h}\| > h_0$ for some $h_0 > 0$. Some examples of taper can be found in [6, 14, 18–20].

When the covariogram K is a Matérn model, explicit infill asymptotic results have been established that assure that an appropriate taper yields no loss in predictive performance or statistical efficiency. First note that if the K and ρ each have a spectral density, denoted by $f_0(\boldsymbol{\lambda})$ and $g(\boldsymbol{\lambda})$, then C has a spectral density as a convolution

$$f_1(\boldsymbol{\lambda}) = \int_{\mathbf{R}^d} f_0(\boldsymbol{\lambda} - \mathbf{x}) g(\mathbf{x}) d\mathbf{x}. \tag{10.16}$$

If

$$f_0(\boldsymbol{\lambda}) \propto \sigma^2 \alpha^{2\nu} / (\alpha^2 + \|\boldsymbol{\lambda}\|^2)^{\nu + d/2} \text{ for } \boldsymbol{\lambda} \in \mathbf{R}^d \tag{10.17}$$

and the spectral density $g(\boldsymbol{\lambda})$ of the taper satisfies the following taper condition:

$$0 < g(\boldsymbol{\lambda}) < M(1 + \|\boldsymbol{\lambda}\|^2)^{-\nu - d/2 - \epsilon} \tag{10.18}$$

for some $\epsilon \geq 0$ and $M > 0$. Then $f_1(\boldsymbol{\lambda}) / f_0(\boldsymbol{\lambda})$ has a finite limit as $\|\boldsymbol{\lambda}\| \to \infty$ and this limit equals 1 if $\epsilon > 0$, where f_1 is given by (10.16)[5, 23] Therefore,

predictions under both models will be nearly the same when a large sample is obtained from a bounded.

[4] established the first infill asymptotic distribution for tapered MLE of the microergodic parameter $\sigma^2\alpha^{2\nu}$ in the Matérn model when the Gaussian process is observed in an bounded interval on the real line. Let $\tilde{V}_n$ denote the tapered covariance function whose (i, j)th element is $C(\|\mathbf{s}_i - \mathbf{s}_j\|)$ where $\mathbf{s}_i, i = 1, \cdots, n$ are n sampling locations. The tapered likelihood function is defined as

$$l_{n,tap}(\alpha, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log[\det \tilde{V}_n] - \frac{1}{2}\mathbf{X}_n'\tilde{V}_n^{-1}\mathbf{X}_n. \tag{10.19}$$

Fixing α at an arbitrary value $\alpha_1 > 0$, let

$$\hat{\sigma}_n^2 = \hat{\sigma}_{n,tap}^2 = \text{ArgMax}\, l_{n,tap}(\alpha_1, \sigma^2).$$

[4] established the following infill asymptotic result

$$\sqrt{n}(\hat{\sigma}_n^2\alpha_1^{2\nu} - \sigma_0^2\alpha_0^{2\nu}) \xrightarrow{d} N(0, 2(\sigma_0^2\alpha_0^{2\nu})^2). \tag{10.20}$$

Comparing with (10.14), we see that the tapered MLE is as efficient as the true MLE.

I am aware of some new results that are not yet published and extend (10.20) to higher dimensions $d = 2$ and 3.

10.5.3 Approximation of Likelihood

When the spatial sample size is large, the inverse of the covariance matrix can be a computational challenge. Hence, the Gaussian likelihood function needs to be approximated to facilitate computation. [17] proposed an approximation to the likelihood. This approximation is based on the following simple idea. Given $Y_n = (Y_1, \cdots, Y_n)$ where $Y_j = Y(\mathbf{s}_j)$ is the observation at location $\mathbf{s}_j$, the log likelihood can be expressed exactly as

$$L_n((t)) = \log p(Y_1) + \sum_{j=2}^{n} \log p(Y_j|Y_{j-1}, \cdots, Y_1; (t)).$$

[17] suggested to approximate $p(Y_j|Y_{j-1}, \cdots, Y_1)$ by the conditional probability density of Y_j on a subset $\boldsymbol{S}_{(j-1)} \subset \{Y_{j-1}, \cdots, Y_1\}$ and recommended choosing $S_{(j-1)}$ to be made up of a few values that are nearest to the side j. This results in a quasi-likelihood

$$\tilde{L}_n((t)) = \log p(Y_1) + \sum_{j=2}^{n} \log p(Y_j|S_{(j-1)}; (t)).$$

Assume the process is Gaussian with mean 0 and a Matérn covariogram of known smoothness parameter. Write $\theta = (\sigma^2, \alpha)$. Assume σ_0 and α_0 are the true value of the parameters and $\alpha_1 > 0$ is an arbitrarily chosen value. Define the estimator

$$\hat{\sigma}^2 = \text{AugMax}\tilde{L}_n(\sigma^2, \alpha_1)$$

We can give sufficient conditions such that the following is true

$$\sqrt{n}(\hat{\sigma}^2\alpha_1^{2\nu} - \sigma_0^2\alpha_0^{2\nu}) \xrightarrow{d} N(0, 2(\sigma_0^2\alpha_0^{2\nu})^2). \tag{10.21}$$

Let $\sigma_1^2 = \sigma_0^2\alpha_0^{2\nu}/\alpha_1^{2\nu}$ and let E_i denote the expectation corresponding to the parameter value (σ_i^2, α_i), $i = 0, 1$. Define

$$e_{1,0} = Y(\mathbf{s}_1),\ e_{k,0} = Y(\mathbf{s}_k) - \text{E}_0[Y(\mathbf{s}_k)|Y(\mathbf{s}_l), 1 \le l < k],\ k = 2, \cdots, n$$
$$e_{1,1} = Y(\mathbf{s}_1),\ e_{k,1} = Y(\mathbf{s}_k) - \text{E}_1[Y(\mathbf{s}_k)|S_{(k-1)}],\ k = 2, \cdots, n.$$

Theorem 10.4. *If*

$$\lim_{n\to\infty} \frac{\text{E}_0(e_{n,1} - e_{n,0})^2}{\text{E}_0 e_{n,0}^2} = 0. \tag{10.22}$$

and

$$\frac{1}{n}\sum_{k=1}^{n}\left(\frac{\text{E}_0(e_{k,1} - e_{k,0})^2}{\text{E}_0 e_{k,0}^2}\right)^{1/2} = o(n^{-1/2}),\ \textit{and}\ \frac{1}{n}\sum_{k=1}^{n}\left|\frac{\text{E}_0 e_{k,0}^2}{\text{E}_1 e_{k,1}^2} - 1\right| = o(n^{-1/2}). \tag{10.23}$$

then (10.21) holds.

Proof of this follows the unpublished result and [24].

10.6 Spatio-Temporal Processes

Quite often a spatial location is observed at different time points. This leads to spatio-temporal data. It is often sensible to employ the increasing domain asymptotic in the time domain as in the time series literature. In the spatial domain, we still have the choice whether to apply the fixed domain or increasing domain asymptotic framework. Currently, most asymptotics for spatio-temporal data assume the increasing domain asymptotics in both the spatial domain and temporal domain [11]. It would be interesting to consider that the time domain is increasing while the spatial domain is fixed. There are no asymptotic results that I am aware of for this kind of spatio-temporal asymptotic framework. One conjuncture is that all model parameters are consistently estimable in this latter case

because the time domain is increasing. However, the asymptotic distribution might depend on whether the spatial domain is fixed or increasing.

Acknowledgements This research is supported by the US National Science Foundation grants DMS-0833323 and IIS-1028291.

References

1. Anderes, E.: On the consistent separation of scale and variance for gaussian random fields. Ann. Stat. **38**(2), 870–893 (2010)
2. Blackwell, D. and Dubins, L.: Merging of opinions with increasing information. Ann. Math. Stat. **33**, 882–886 (1962)
3. Diggle, P., Tawn, J.A., Moyeed, R.A.: Model-based geostatistics (with discussion). J. Roy. Stat. Soc. C-App, **47**, 299–350 (1998)
4. Du, J., Zhang, H., Mandrekar, V.: Fixed-domain asymptotic properties of tapered maximum likelihood estimators. Ann. Stat. **37**, 3330–3361 (2009)
5. Furrer, R., Genton, M.G., Nychka, D.: Covariance tapering for interpolation of large spatial datasets. J. Comput. Graph. Stat. **15**(3), 502–523 (2006)
6. Gneiting, T.: Compactly supported correlation functions. J. Multivariate Anal. **83**(2), 493–508 (2002)
7. Guyon, X.: Random Fields on a Network: Modelling, Statistics and Applications. Springer, New York (1995)
8. Kaufman, C., Schervish, M., Nychka, D.: Covariance tapering for likelihood-based estimation in large spatial datasets. J. Am. Stat. Assoc.**103**, 1545–1555 (2009)
9. Lahiri, S.N.: On inconsistency of estimators based on spatial data under infill asymptotics. Sankhya. Ser. A., **58**, 403–417 (1996)
10. Lahiri, S.N., Lee, Y., Cressie, N.: On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. J. Stat. Plan. Infer., **103**(1-2), 65–85 (2002)
11. Li, B., Genton, M.G., Sherman, M.: Testing the covariance structure of multivariate random fields. Biometrika, **95**(4), 813–829 (2008)
12. Liptser, R.S., Shiryayev, A.N.: Statistics of Random Processes. Springer, New York (1977)
13. Mardia, K.V., Marshall, R.J.: Maximum likelihood estimation of models for residual covariance in spatial statistics. Biometrika. **71**, 135–146 (1984)
14. Mitra, S., Gneiting, T., Sasvári, Z.: Polynomial covariance functions on intervals. Bernoulli. **9**(2), 229–241 (2003)
15. Shiryayev, A.N.: Probability. Springer, New York (1984)
16. Stein, M.L.: Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York (1999)
17. Vecchia, A.V.: Estimation and model identification for continuous spatial processes. J. Roy. Stat. Soc. B. **50**, 297–312 (1998)
18. Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Adv. Comput. Math. **4**, 289–396 (1995)
19. Wendland, H.: Error estimates for interpolation by compactly supported radial basis functions of minimal degree. Adv. Comput. Math. **93**, 258–272 (1998)
20. Wu, Z.M.: Compactly supported positive definite radial functions. Adv. Comput. Math. **4**, 283–292 (1995)
21. Ying, Z.: Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. J. Multivariate Anal. **36**, 280–296 (1991)
22. Zhang, H.: Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. J. Am. Stat. Assoc. **99**, 250–261 (2004)

23. Zhang, H., Du, J.: Covariance tapering in spatial statistics. In J. Mateu and E. Porcu, editors, Positive definite functions: From Schoenberg to space-time challenges. 181–196 (2008a)
24. Zhang, H., Du, J.: A new approach to deriving infill asymptotics in spatial statistics. Unpublished manuscript (2008b)
25. Zhang, H., Wang, Y.: Kriging and cross-validation for massive spatial data environmetrics. Environmetrics. 290–304 (2010)
26. Zhang, H., Zimmerman, D.L.: Towards reconciling two asymptotic frameworks in spatial statistics. Biometrika. **92**, 921–936 (2005)
27. Zhang, H., Zimmerman, D.L.: Hybrid estimation of semivariogram parameters. Math. Geol., **39**, 247–260 (2007)