
Denis Allard

To cite this version:


HAL Id: hal-01236545
https://hal.science/hal-01236545
Submitted on 1 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
J.-P. Chilès, P. Delfiner: Geostatistics: Modeling Spatial Uncertainty

Denis Allard

© International Association for Mathematical Geosciences 2012

Geostatistics: Modeling Spatial Uncertainty by J.-P. Chilès and P. Delfiner published in 1999 has been one of the most cited reference book in Geostatistics and spatial statistics in the last decade. There are very good reasons for this: it is, in my opinion, the most comprehensive book on geostatistics, with an in-depth coverage of all aspects, from random field theory and spectral representation to virtually all practical issues arising when modeling spatial data, including variogram fitting, (co) kriging, estimating nonlinear quantities in presence of a change of support, or when performing conditional simulations. In a unique way, it reconciles theory and practice. It provides very precise theoretical results and excellent advice on the practice of modeling spatial data. It is profusely illustrated with numerous application examples, mostly in earth science.

Time has come for a new edition, which would include the many developments that have been made since 1999. The authors had a constraint: The second edition should be no longer than the first one. They decided to remove most of the material on stochastic hydrogeology, which “was too detailed for the casual reader and too incomplete for the specialist.” Material of interest for geostatisticians was distributed throughout the relevant chapters, such as the section on stochastic differential equations now part of the section on Intrinsic Random Functions of order $k$ (IRF-$k$). In other places, some material was removed or condensed. Some figures were reduced in size.

This left the authors with about 40 to 50 pages for new material. They thoroughly revised many aspects of the book while keeping the same general architecture. First of all, they included two brand new sections, on space–time models (Sect. 5.8, 16 pages)
and on models for maxima and extreme-values (Sect. 6.11, 8 pages), the latter written by C. Lantuéjoul.

Space–time statistics has seen a surge of theoretical and methodological advances in the last decade. This topic deserved a whole chapter, but the authors preferred “to outline the main approaches and refer to recent review papers and book” rather than adding “pages to an already thick book.” Space–time models and methods are very diverse, depending on the type of data at hand (time series at multiple locations, time repetitions of spatial measurements, measurements at different time and location, etc.) and whether or not there exists a physical model allowing the use of assimilation methods. In a very nice introduction, the authors propose a clear typology, which will be used to organize the rest of the section. The main results on nonseparable covariance functions are presented, along with the references to the main papers. Three examples of specific space–time models are then given that practitioners with similar or close to similar situations will find useful for developing their own model. The rest of the section is then devoted to sequential data assimilation, the Kalman filter, and its extensions. Very often, but not always, the size of the dataset is one of the problems the practitioner has to address when analyzing and modeling space–time data. This problem is not explicitly mentioned in this section, probably due to space constraints. Interested readers will have to find more details for implementing their methods in some other book, e.g., the recent Statistics for Spatio-Temporal Data by Cressie and Wikle (2011).

Predicting the spatial extent of an extreme quantile of a variable, e.g., precipitation or pollutant concentration, is recognized as one of the most challenging problems in spatial statistics. It is one of the hot research topics of the moment. The authors are to be congratulated for having included such a section in this new edition. Written by C. Lantuéjoul, coauthor of several papers on the subject, it contains the most important results obtained in the last years. To my knowledge, the material in this section is presented in a geostatistics book for the very first time. My only mild criticism regards its title and location. Presenting the extreme-value theory as a particular case of the change of support problem is very elegant from a theoretical point of view. It is, however, not intuitive for the noninformed reader to guess the content from the reading of the table of contents. A more usual title such as “Models for Maxima” would have made the material more visible.

There are many other additions distributed all along the text. It is beyond the scope of this review to make a comprehensive list. Instead, I will highlight the ones I found the most important. The chapter on the structural analysis has been enriched with covariance models on spheres and on river networks. The link between the Matérn class and Gaussian Markov random field (GMRF) is presented and a reference to the very recent work by Lindgren et al. (2011) is made. Universal Kriging and Kriging with External Drift are now presented in a unified framework highlighting that UK is nothing but a KED where the external functions are equal to the basis functions. Two approaches are presented for reducing the size of the kriging neighborhood in Sect. 3.6: covariance tapering and the GMRF approximation. Fixed-rank kriging (Cressie and

1 Possibly due to the characteristic modesty of the author, two papers written by C. Lantuéjoul on this topic are not cited in the book. References are given below.
Johannesson 2008), which was proven to be very efficient to deal with very large datasets is unfortunately not presented. I consider that tackling very large data set is another big challenge in spatial and spatiotemporal modeling (in particular for climatic data). A specific section on this topic would certainly have been welcome. Dual kriging is presented in relationship with the radial basis function approach. There is also a useful section on the use of kriging for the design and analysis of computer experiments; this question has recently received great attention in the experimental design community. The literature is rapidly growing on the subject. This (too) short section is a tribute to this relatively new field of application for geostatistics.

Chapter 5, on multivariate methods, was largely rewritten and augmented. Quoting the authors “the chapter now includes a thorough review of different forms of collocated cokriging, with a clear picture of which underlying models support the approach.” Comparison with KED is systematic and very well discussed. The chapter also includes new material on Min/Max autocorrelation factors and on compositional data. As already written above, the last section considered as a “threat for the size of the book” by the authors is about space–time models.

Chapter 6, on nonlinear geostatistics, has been considerably reorganized. The presentation of the critical step of transforming a variable to a Gaussian transform has been expanded and updated. Prediction of local point distributions by conditional expectation, indicator (co) kriging and disjunctive kriging are now presented first. Change of support models are then discussed in depth, including new variants that appeared in the 2000s. The chapter ends with the section on extreme-value theory and models for maxima already discussed above.

Conditional simulation (Chap. 7) has been a very active field of research in the last decade and the numerous advances had to be incorporated in the second edition of the book. Simulations of fractional Brownian motion and IRF-k, now possible exactly, are presented as extensions of each method (dilution methods, spectral methods), instead of being the subject of a separate section. The text gains in continuity and clarity. Presentation of MCMC methods has also been expanded. A yet unpublished new Gibbs propagation algorithm proposed by C. Lantuéjoul, at the same time very elegant and very efficient, is presented for the first time. A new section, entitled Beyond standard conditioning includes the very recent developments in multipoint simulation and the gradual deformation method.

The authors made also some subtle vocabulary changes. Now that spatial statistics (and thus geostatistics as a major subdomain of spatial statistics) is definitely recognized as part of “main stream statistics” (most statistic programs include a spatial statistics course; most statistical conferences include spatial statistics sessions), time has come to adopt some of the “main stream vocabulary.” K-Bessel covariance functions are now Matérn covariance functions; “Sensitivity of Kriging to Misspecification of the variogram” (Sect. 3.4.3) now is “Kriging with Estimated Parameters” (Sect. 3.4.4), etc. In the process of adopting mainstream vocabulary, the book has a less typical “old geostatistics” flavor. The flavor is not completely gone though. Variogram models are still “fitted” in this book, whereas its parameters are “estimated” for statisticians. This semantic distinction emphasizes the importance of integrating prior scientific knowledge in the process of modeling variograms (regularity at the origin, anisotropy, hole effect, etc.).
To summarize, this book is an excellent update of the first edition. The authors kept the same level of mathematical rigor, while avoiding technicalities that are not deemed essential for geostatisticians. Two important sections were added. Most chapters underwent a thorough revision. Many sections have been reorganized and rewritten at the light of the new advances made in the last decade. Relationships to new methods, such as the Gaussian–Markov random field, fractional Brownian motion, and extreme-value theory are very clearly presented.

Among the things I like most in this book is the very good balance between the mathematical rigor and the modeling point of view. This book is an excellent reference book for anyone working with spatial data, from the academic (geo)statistician to the advanced practitioner. If, in addition, he/she works in earth science, this book is definitely a “must have” book. It has not been specifically designed to be a textbook, but the most advanced students will find a clear presentation of all major methodological issues. There are a couple of things I regret. Not about what is in the book, but about what is not. I still believe that space–time modeling deserved more than 16 pages. I also believe that the “big data” challenge, i.e., the set of methods designed for coping with $10^6$ to $10^9$ data also deserved a section on its own. Last, but not least, there is barely any mention of hierarchical spatial models and the model-based geostatistics approach (Diggle et al. 1998; Diggle and Ribeiro 2007), which allows the modeling of nonlinear, complex, random relationships between the underlying spatial process of interest and the measurement process. The cost would have been, of course, an extra 50 to 100 pages to an already thick book.

Chiles and Delfiner first edition was my favorite book on geostatistics. Today, it has a rather charming tatty look, reminding me how often I used it, and how fast time flies. The first edition will enjoy a well deserved retirement and second edition will take its place on the shelf and in my short list of favorite books.

References