Statistical Science
2015, Vol. 30, No. 2, 164–166
DOI: 10.1214/15-STS515
Main article DOI: 10.1214/14-STS487

Beyond the Valley of the Covariance Function

Daniel Simpson, Finn Lindgren and Håvard Rue

1. INTRODUCTION

Multivariate models are under-represented in the literature on spatial statistics. There is a basic reason for this: univariate models are sufficiently complicated to keep us busy. Genton and Kleiber have done a fabulous job compiling and investigating the available models, with a focus on the important class of models that they, with collaborators, introduced. This paper gives a solid state of the art and points out just how many holes there are in the theory and practice associated with these fields. This gives us licence to point out some other holes and to suggest some important directions for the future.

2. THERE IS POWER IN A SPECTRUM

If we were to quibble about one thing in Genton and Kleiber's paper, it would be that we disagree over the extent to which the class of multivariate GRFs has been categorized. Note that this is different from *explicitly* constructing valid cross-covariance functions! To wit, if a multivariate GRF has a spectral representation, the spectral representation given in Section 1.2 completely characterizes the class of stationary multivariate random fields that admit an absolutely continuous spectral measure. This represents a large chunk of interesting GRFs. We note that the paper, by restricting the cross-spectral densities to be real, implicitly assumes that $C_{ij}(\mathbf{h}) = C_{ij}(-\mathbf{h})$, when the minimal necessary requirement is only that $C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$, which allows for phase differences between the model components. The representation can then be employed constructively as follows. Let $\boldsymbol{\omega} \to \mathbf{S}(\boldsymbol{\omega})$ be a mapping from $\mathbb{R}^d$ to the set of Hermitian nonnegative definite matrices, the elements of which the cross-spectral densities, denoted f_{ij} in the paper, are here subject to $f_{ij}(\boldsymbol{\omega}) = \overline{f_{ji}(\boldsymbol{\omega})}$. Then, for any complex, matrix-valued function $\mathbf{L}(\boldsymbol{\omega})$ such that $L_{ij}(\boldsymbol{\omega}) = \overline{L_{ij}(-\boldsymbol{\omega})}$ and $\mathbf{S}(\cdot) = \mathbf{L}(\cdot)\overline{\mathbf{L}(\cdot)}$,

$$\begin{aligned}\mathbf{x}(\mathbf{s}) &= \int_{\mathbb{R}^d} \mathbf{L}(\boldsymbol{\omega}) \mathrm{e}^{i\mathbf{s}\cdot\boldsymbol{\omega}}\, d\widetilde{\mathbf{W}}(\boldsymbol{\omega}) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{L}(\boldsymbol{\omega}) \mathrm{e}^{i(\mathbf{s}-\mathbf{s}')\cdot\boldsymbol{\omega}}\, d\boldsymbol{\omega}\, d\mathbf{W}(\mathbf{s}'),\end{aligned} \tag{1}$$

where $d\widetilde{\mathbf{W}}(\cdot) \in \mathbb{C}^p$ and $d\mathbf{W}(\cdot) \in \mathbb{R}^p$ are Gaussian white noise processes on $\mathbb{R}^d$ understood as random measures with $d\widetilde{W}_i(\boldsymbol{\omega}) = \overline{d\widetilde{W}_i(-\boldsymbol{\omega})}$, $\mathbb{E}[d\widetilde{\mathbf{W}}(\boldsymbol{\omega}) \cdot \overline{d\widetilde{\mathbf{W}}(\boldsymbol{\omega}')}] = \delta(\boldsymbol{\omega} - \boldsymbol{\omega}')\mathbf{I}\, d\boldsymbol{\omega}$, and $\mathbb{E}[d\mathbf{W}(\mathbf{s})\overline{d\mathbf{W}(\mathbf{s}')}] = \delta(\mathbf{s}-\mathbf{s}')\mathbf{I}\, d\mathbf{s}$ (Adler and Taylor, 2007; Lindgren, 2012). This representation only covers multivariate GRFs with absolutely continuous spectral measures; however, the same procedure applies to fields with an atomic spectral representation. The abstract feature that is hiding in all of this specificity is that we are explicitly constructing a square root of the multivariate covariance operator and using this square root to filter the multivariate white noise. On a compact domain, the covariance operator is a compact, trace class operator, and so this square root is well defined using the usual functional calculus.

Another reason to further emphasize this spectral representation is that it is not only constructive in its own right, but also useful when transformed back to the nonspectral domain. Kernel convolution methods (Higdon, 1998) have a storied history in univariate spatial statistics and their generalization to the multivariate case is straightforward (Simpson, Lindgren and Rue, 2012; Bolin and Lindgren, 2013). Their advantage is that it is never necessary to identify the spectrum of the process or, in fact, the cross-covariance structure. Rather, for any L^2 matrix-valued function $\mathbf{K}(\cdot,\cdot)$, $\mathbf{x}(s) = \int \mathbf{K}(\mathbf{s},\mathbf{s}')\, d\mathbf{W}(\mathbf{s}')$ is a valid mutlivariate GRF, which can be approximated by (carefully) approximating the corresponding integral with a sum. In

Daniel Simpson is CRiSM Fellow, Department of Statistics, University of Warwick, CV4 7AL, Coventry, United Kingdom (e-mail: D.P.Simpson@warwick.ac.uk).
Finn Lindgren is Reader in Statistics, Department of Mathematical Sciences, University of Bath, BA2 7AY, Bath, United Kingdom. Håvard Rue is Professor of Statistics, Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491, Trondheim, Norway (e-mail: hrue@math.ntnu.no).

fact, there is nothing special about white noise in this situation, $\mathbf{W}(\cdot)$ can be any independently scattered $\mathbb{R}^d$-valued random measure. This leads to a natural way to construct non-Gaussian random fields (Åberg and Podgórski, 2011).

Different choices of $\mathbf{K}(\cdot,\cdot)$ generating the same covariance function will in the non-Gaussian case affect the dependence structure. Similarly, the choice of square root in the spectral representation becomes relevant, and the two integrals in (1) are no longer guaranteed to give the same process model. We also note that $\mathbf{K}(\mathbf{s},\mathbf{s}')$ does not have to be a function of $\mathbf{s}-\mathbf{s}'$ and, hence, this resulting field does not need to be stationary.

An important advantage to the multivariate spectral and convolution kernel constructions is that the spectral operator $\mathbf{L}(\boldsymbol{\omega})$ and the kernel matrix $\mathbf{K}(\cdot,\cdot)$ can reflect the modeler's knowledge about the physical process under consideration. This can lead to useful, informative covariance structures that are tailored to the specific application. This idea falls under the auspices of the physics-constrained cross-covariance specifications mentioned in Section 7.3, except for one key difference: *while a cross-covariance structure is present, identifying it is not necessary for inference, either conceptually or computationally*!

3. THINK LOCAL, ACT LOCAL

In fact, there are many parts of the above construction that we can live without. The SPDE approach introduced by Lindgren, Rue and Lindström (2011) as a computationally efficient reformulation of GRFs is an example of the same procedure where we never construct the kernel matrix. Instead, multivariate models can be constructed using systems of equations, such that the kernel corresponds to (matrix-valued) Green's function of some linear partial differential operator (Hu et al., 2013a, 2013b, 2013c). Using the partial differential operator construction has several advantages over the direct kernel specification. First, when the representation is Markovian, it allows us to localize the process. This is especially convenient when moving to non-Euclidean spaces; the great tragedy of spatial statistics is that the Earth turned out not to be flat. The second major advantage is that this localization allows us to construct local approximations to the resulting random fields. For an n-dimensional approximation to a field observed at N points, this reduces the computational cost of fitting the field from $\mathcal{O}((pN)^3)$ to $\mathcal{O}(pN + p^3n^{3/2})$, for the case $d = 2$. This also compares well with non-Markovian methods, such as the aforementioned n-dimensional convolution kernel methods that have computational cost of $\mathcal{O}(pn^2N + p^3n^3)$. Given that we are now living in the age of "big data," this is a serious advantage to the SPDE specification. The third major advantage is that it is straightforward to construct nonstationary models by locally varying the partial differential operator in the model. This corresponds to the "physics" view, where dependency structures are specified locally and extended to a global covariance structure through a conditioning argument. In our experience, this is an extremely powerful tool for specifying useful covariance and cross-covariance structures.

4. WITH LOW POWER COMES GREAT RESPONSIBILITY

One of the principal challenges that we have encountered when applying likelihood methods to multivariate GRFs is that their likelihood surfaces tend to be flat. This is perhaps not a surprise. If fitting p univariate models requires the estimation of $\mathcal{O}(p)$ parameters, then a p-component multivariate model will require the same data to be informative about $\mathcal{O}(p^2)$ parameters. This parameter inflation becomes noticeable already for bivariate models, but for large p it is a serious issue, that can sometimes be partially alleviated by using a low rank model, ideally motivated by problem-specific knowledge. Another option is to impose a sparse structure on the linear filter matrix operator or to impose constraints between the parameters.

While flat likelihoods resulting from an exploding parameter space are annoying, there is a far more pathological problem for univariate GRFs. Fundamentally, the range and variance parameters are not identifiable under the usual infill regime (Zhang, 2004). This leads to a ridge in the parameter space that can only be resolved using careful prior modeling (Simpson et al., 2014). Ridges will also seriously challenge numerical optimizers and MCMC schemes unless they use enough second-order information to resolve it. It is currently unclear to what extent these problems extend to multivariate models; however, we suspect that they do, due to the aforementioned parameter inflation.

REFERENCES

ÅBERG, S. and PODGÓRSKI, K. (2011). A class of non-Gaussian second order random fields. *Extremes* **14** 187–222.

ADLER, R. J. and TAYLOR, J. (2007). *Random Fields and Geometry*. Springer, Berlin.

BOLIN, D. and LINDGREN, F. (2013). A comparison between Markov approximations and other methods for large spatial data sets. *Comput. Statist. Data Anal.* **61** 7–32.

HIGDON, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environ. Ecol. Stat.* **5** 173–190.

HU, X., LINDGREN, F., SIMPSON, D. and RUE, H. (2013a). Multivariate Gaussian random fields with oscillating covariance functions using systems of stochastic partial differential equations. Dept. Mathematical Sciences, NTNU, Norway. Available at arXiv:1307.1384.

HU, X., SIMPSON, D., LINDGREN, F. and RUE, H. (2013b). Multivariate Gaussian random fields using systems of stochastic partial differential equations. Dept. Mathematical Sciences, NTNU, Norway. Available at arXiv:1307.1379.

HU, X., STEINSLAND, I., SIMPSON, D., MARTINO, S. and RUE, H. (2013c). Spatial modelling of temperature and humidity using systems of stochastic partial differential equations. Dept. Mathematical Sciences, NTNU, Norway. Available at arXiv:1307.1402.

LINDGREN, G. (2012). *Stationary Stochastic Processes*: *Theory and Applications*. Chapman & Hall/CRC, London.

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498.

SIMPSON, D., LINDGREN, F. and RUE, H. (2012). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics* **23** 65–74.

SIMPSON, D. P., MARTINS, T. G., RIEBLER, A., FUGLSTAD, G.-A., RUE, H. and SØRBYE, S. H. (2014). Penalising model component complexity: A principled, practical approach to constructing priors. Preprint. Available at arXiv:1403.4630.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261.