NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

University of Wollongong

Working Paper

02-15

Capturing Multivariate Spatial Dependence: Model, Estimate, and then Predict

Noel Cressie, Sandy Burden, Walter Davis, Pavel Krivitsky, Payam Mokhtarian, Thomas Suesse and Andrew Zammit-Mangion

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Capturing Multivariate Spatial Dependence: Model, Estimate, and then Predict

Noel Cressie* **and Sandy Burden and Walter Davis and Pavel Krivitsky and Payam Mokhtarian and Thomas Suesse and Andrew Zammit-Mangion**

National Institute for Applied Statistics Research Australia (NIASRA)
University of Wollongong, AUSTRALIA

We would like to thank Marc Genton and William Kleiber (hereafter, GK) for their informative review, "Cross-covariance functions for multivariate geostatistics" (forthcoming in *Statistical Science*), and the editor of *Statistical Science* for the opportunity to contribute to the discussion.

Physical processes rarely occur in isolation, rather they influence and interact with one another. Thus, there is great benefit in modelling potential dependence between both spatial locations and different processes. It is the interaction between these two dependencies that is the focus of GK.

We see the problem of ensuring that the matrix given in GK-(2) is nonnegative definite (nnd) as important, but we also see it as a means to an end. That "end" is solving the scientific problem of predicting a multivariate field of, say, temperature and rainfall, based on noisy and spatially incomplete data from weather stations in a region of interest. There is also scientific interest in the behaviour of the measures of cross-spatial dependence (e.g., cross-covariance functions), but usually spatial prediction is the ultimate goal.

Of course, an interim goal is estimation of the means, covariances, and cross-covariances, but not a lot of GK's review was devoted to this. The nonparametric estimators given by GK-(6) and GK-(11) are useful for recognising which parametric *class* of valid cross-covariance functions might represent the multivariate spatial dependence in the data. Estimation of the parameters in this class is usually obtained by weighted least squares or maximum likelihood. Optimal spatial prediction in practice proceeds by substituting these parameter estimates into the model and computing the optimal data weights as if the parameters were known. Because of this, predictors and their standard errors are biased. These and other issues (e.g., change-of-support) are well known in the univariate spatial setting and they clearly also arise in the multivariate spatial setting.

*National Institute for Applied Statistics Research Australia (NIASRA), University of Wollongong, NSW 2522, AUSTRALIA (e-mail: ncressie@uow.edu.au).

One problem that arises in multivariate spatial statistics, but not discussed very much by GK, is collocation (or not) of spatial data from the different variables. This might be viewed as a missing-data problem, for which a hierarchical multivariate spatial statistical model offers a path forward. Being hierarchical does not necessarily mean being Bayesian, as the next section on latent modelling demonstrates.

1. THE LATENT PROCESS IS WHERE THE SPATIAL DEPENDENCE IS USUALLY MODELLED

The multivariate spatial models in GK do not account for the measurement error that exists for all physical observations. Their multivariate spatial processes are written as $\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$, and valid spatial-covariance models $\{\mathbf{C}(\mathbf{s}, \mathbf{u}) : \mathbf{s}, \mathbf{u} \in \mathbb{R}^d\}$ are constructed for them. GK's models are all "smooth," in the sense that

$$\lim_{\mathbf{u}\to\mathbf{s}} \mathbf{C}(\mathbf{s}, \mathbf{u}) = \mathbf{C}(\mathbf{s}, \mathbf{s}). \tag{1}$$

However, an observation or potential observation is observed with error, since no measuring instrument is perfect. Therefore, if the observations are $\mathbf{Z}(\mathbf{s}_1), \ldots, \mathbf{Z}(\mathbf{s}_n)$, then there is a hidden (or latent) process $\{\mathbf{Y}(\mathbf{s}) : \mathbf{s} \in D\}$ such that

$$\mathbf{Z}(\mathbf{s}_i) = \mathbf{Y}(\mathbf{s}_i) + \boldsymbol{\varepsilon}(\mathbf{s}_i); \quad i = 1, \ldots, n, \tag{2}$$

where $\boldsymbol{\varepsilon}(\mathbf{s}_i)$ has mean zero and covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}(\mathbf{s}_i)) \equiv \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{s}_i)$. Further, the measurement process is independent of the latent process, and it is usually reasonable to assume that it is independent from one observation to another; that is, $\mathrm{cov}(\boldsymbol{\varepsilon}(\mathbf{s}_i), \boldsymbol{\varepsilon}(\mathbf{s}_j)) = \mathbf{0}$, for $i \neq j$.

It appears that GK build multivariate spatial covariance models for the latent process, yet the definition of $\hat{\mathbf{C}}(\mathbf{h})$ in GK-(6) is based on observations that come with measurement error. Hence, $\hat{\mathbf{C}}(\mathbf{h})$ is estimating a $\mathbf{C}_Z(\mathbf{h})$ that satisfies:

$$\mathbf{C}_Z(\mathbf{0}) - \lim_{\mathbf{h}\to\mathbf{0}} \mathbf{C}_Z(\mathbf{h}) \text{ is nnd.} \tag{3}$$

This difference of the two matrices above is in fact the measurement-error covariance matrix, which we denote as $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{0})$ in the stationary case (i.e., where $\mathbf{C}(\mathbf{s}, \mathbf{s}+\mathbf{h})$ depends only on $\mathbf{h}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{s}) = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{0})$ for all $\mathbf{s}$).

This mismatch between (3) and the stationary version of (1), namely $\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{C}(\mathbf{h}) = \mathbf{C}(\mathbf{0})$, can be resolved once one realises that GK are really building models for a latent $\mathbf{Y}$-process, and that a full multivariate spatial covariance function for the $\mathbf{Z}$-process (i.e., the observations) is obtained by additionally modelling the measurement-error covariance, $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{0})$. In their second example (GK-Section 6.2), GK recognise the need for $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{0})$: "Due to the fact that the data are observational, we augment each process' covariance with a nugget effect." However, there are potentially nonzero off-diagonal terms in $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{0})$. Notice that the two variables in GK-Section 6.2 are observed maximum and minimum temperatures obtained from the *same* instrument at location $\mathbf{s}_i$, say, and hence the measurement

error for the maximum, $\varepsilon_1(\mathbf{s}_i)$, and the measurement error for the minimum, $\varepsilon_2(\mathbf{s}_i)$, should be correlated. GK's choice of a diagonal matrix for $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{0})$ does not reflect this.

While it may not be obvious, there are also circumstances where a "measurement error" component is needed when modelling deterministic spatio-temporal output from computer experiments, such as those used in GK-Section 6.1. This component is actually a spatio-temporal interaction that "hides" the latent spatial process (Kang and Cressie, 2013).

Finally, it is possible that a latent $\mathbf{Y}$-process is itself not "smooth." In this case, we can write the latent process as

$$\mathbf{Y}(\mathbf{s}) = \mathbf{W}(\mathbf{s}) + \boldsymbol{\xi}(\mathbf{s}); \quad \mathbf{s} \in D, \tag{4}$$

where the $\mathbf{W}$-process is smooth (i.e., satisfies (1)), and where the $\boldsymbol{\xi}$-process is independent of the $\mathbf{W}$-process and has mean zero. Often, it is assumed that $\boldsymbol{\xi}(\mathbf{s})$ is independent of $\boldsymbol{\xi}(\mathbf{u})$ for any $\mathbf{s} \neq \mathbf{u}$ and, when $\mathbf{u} = \mathbf{s}$, $\text{cov}(\boldsymbol{\xi}(\mathbf{s}), \boldsymbol{\xi}(\mathbf{u})) \equiv \mathbf{\Sigma}_{\boldsymbol{\xi}}(\mathbf{0})$.

When modelling univariate spatial data, there has been considerable inconsistency in the literature regarding how to handle the ξ-process (micro-scale variation) and the ε-process (measurement-error process). That confusion should also be avoided in the multivariate spatial setting. We need both $\mathbf{\Sigma}_{\boldsymbol{\xi}}(\mathbf{0})$ and $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}(\mathbf{0})$ for different purposes, and these different roles should be accounted for: We wish to filter out the $\boldsymbol{\varepsilon}$-process (since it is extraneous to the true, hidden $\mathbf{Y}$-process), but we wish to predict the $\boldsymbol{\xi}$-process (since it represents the scientific process at micro-scales). The presence of both processes is manifested in $\hat{\mathbf{C}}(\mathbf{h})$, for $\mathbf{h}$ near $\mathbf{0}$, but without more information than that supplied by the multivariate spatial data, the $\boldsymbol{\xi}$-process and the $\boldsymbol{\varepsilon}$-process are confounded.

2. ESTIMATE USING CROSS-VARIOGRAMS, THEN PREDICT USING CROSS-COVARIANCES

A small amount of GK's review of cross-covariances discusses cross-variograms and generalised covariance functions, since they are stationary when the process $\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D\}$ is differenced. Before differencing, the process is nonstationary. There are a number of ways to do the differencing in a multivariate context, leading to a lack of agreement among researchers of how to capture the cross-dependence between the processes $\{Z_q(\mathbf{s}) : \mathbf{s} \in D\}$ and $\{Z_r(\mathbf{u}) : \mathbf{u} \in D\}, 1 \leq q \neq r \leq p$. For many scientific purposes, the key goal is optimal multivariate spatial prediction. Therefore, the key measure of multivariate spatial dependence should be one that can be used without fail in kriging and co-kriging equations. That is, the optimal weights in the linear combination of the spatial data, $\{Z_q(\mathbf{s}_{qi}) : i = 1, \ldots, n_q, q = 1, \ldots, p\}$, should depend on this measure. If a measure sometimes yields non-optimal weights, we suggest that it is not as interesting as one that does. While GK-(1) and GK-(4) yield optimal weights, the (covariance-based) cross-variograms given by GK-(3) do not always.

Ver Hoef and Cressie (1993, 1994) gave an example where use of GK-(3) in spatial multivariate prediction yielded non-optimal weights. Indeed, it is GK-(3) that should have been tagged "pseudo" in the literature, not GK-(4). The article that gives the most general multivariate spatial dependence measure that is a function of $\mathbf{h} = \mathbf{s} - \mathbf{u}$, is Künsch, Papritz and Bassi (1997), who define the generalised cross-covariance functions. Certainly, researchers' familiarity with these more general forms of stationary cross-dependence is not high, but the interpretation of the appropriate cross-variograms given by GK-(4) is not difficult (Künsch, Papritz and Bassi, 1997; Majure and Cressie, 1997; Cressie and Wikle, 1998; Huang et al., 2009).

We have found that for *univariate* spatial-processes, the estimation of spatial-dependence parameters is achieved more stably through the variogram than the covariance function (Cressie, 1993, Section 2.4.1). On the other hand, because optimal spatial prediction using a valid covariance function can be used without fail (Cressie and Wikle, 2011, Section 4.1.2), we recommend the following inferential strategy for *multivariate* spatial processes: Put the cross-variogram at the core of parameter estimation and the cross-covariance function at the core of optimal multivariate spatial prediction.

We conclude that, for multivariate spatial processes, the appropriate cross-variograms given by GK-(4) have great potential for the purpose of estimation, and there is a need for a research program to pursue the interpretation and robust estimation of GK-(4), but *not* of the inappropriate GK-(3).

3. CAPTURING SPATIAL DEPENDENCE: A CONDITIONAL APPROACH

Let $[\cdot]$ denote the probability distribution of the argument within square brackets. GK tackle cross-covariance construction from the perspective of the *joint* distribution, $[Z_1(\cdot), Z_2(\cdot)]$. The *conditional* approach to constructing cross-covariance functions writes the joint distribution as the product, $[Z_2(\cdot)|Z_1(\cdot)][Z_1(\cdot)]$. Consider the space D discretised onto a fine-resolution grid, $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, such that the processes $Z_1(\cdot)$ and $Z_2(\cdot)$ are represented as n-dimensional vectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively. In practice, this is how a continuously indexed process is represented in a computer program. Then, from Cressie and Wikle (2011, p. 160), the conditional approach yields the bivariate spatial model,

$$\text{cov}(\mathbf{Z}_2) = \boldsymbol{\Sigma}_{2|1} + \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}', \tag{5}$$

$$\text{cov}(\mathbf{Z}_1, \mathbf{Z}_2) = \boldsymbol{\Sigma}_{11}\mathbf{B}', \tag{6}$$

$$\text{cov}(\mathbf{Z}_1) = \boldsymbol{\Sigma}_{11}, \tag{7}$$

where $\boldsymbol{\Sigma}_{2|1}$ and $\boldsymbol{\Sigma}_{11}$ are nnd matrices obtained from univariate spatial processes, and the $n \times n$ matrix $\mathbf{B}$ of real-valued entries is unrestricted. Critically, the joint matrix, $\text{cov}\big((\mathbf{Z}_1', \mathbf{Z}_2')'\big)$ is always nnd.

Equations (5)–(7) are obtained from,

$$\mathrm{E}(\mathbf{Z}_2|\mathbf{Z}_1) = \mathbf{B}\mathbf{Z}_1, \tag{8}$$

$$\mathrm{cov}(\mathbf{Z}_2|\mathbf{Z}_1) = \boldsymbol{\Sigma}_{2|1}, \tag{9}$$

$$\mathrm{cov}(\mathbf{Z}_1) = \boldsymbol{\Sigma}_{11}, \tag{10}$$

which only involve univariate spatial processes. It should be emphasised that the conditioning in (8) and (9) is on the whole process $\mathbf{Z}_1$. In contrast to what has been stated elsewhere (Banerjee, Carlin and Gelfand, 2015, p. 273), there is no attempt in the conditional approach to build a joint distribution solely from $Z_2(\mathbf{s}_i)|Z_1(\mathbf{s}_i)$, for $i = 1, \ldots, n$. Indeed,

$$[Z_2(\mathbf{s}_i)|Z_1(\mathbf{s}_i)] = \int \ldots \int \frac{[Z_2(\mathbf{s}_i)|\mathbf{Z}_1]\,[\mathbf{Z}_1]}{[Z_1(\mathbf{s}_i)]} d\mathbf{Z}_{1,-i},$$

where

$$d\mathbf{Z}_{1,-i} = dZ_1(\mathbf{s}_1)\ldots dZ_1(\mathbf{s}_{i-1})dZ_1(\mathbf{s}_{i+1})\ldots dZ_1(\mathbf{s}_n).$$

The order of the variables $\mathbf{Z}_1$ and $\mathbf{Z}_2$ in the conditional approach is a choice that is generally driven by the underlying science (e.g., Royle et al., 1999). When more variables are involved, the order may not always be obvious but, if the goal is to construct valid covariance and cross-covariance functions, the different orderings can be viewed as enlarging the space of valid models.

4. CAPTURING SPATIAL DEPENDENCE THROUGH "FACTOR" PROCESSES: A JOINT APPROACH

Any univariate covariance function, $C(\mathbf{s}, \mathbf{u})$, that satisfies mild integrability conditions has a Karhunen-Loéve representation (Papoulis, 1991):

$$C(\mathbf{s}, \mathbf{u}) = \sum_{a=1}^{\infty} \lambda_a P_a(\mathbf{s}) P_a(\mathbf{u}),$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$, and $\{P_a(\cdot) : a = 1, 2, \ldots\}$ are orthogonal eigenvectors obtained by solving a Fredholm integral equation. After truncation, the function,

$$C^{(b)}(\mathbf{s}, \mathbf{u}) \equiv \sum_{a=1}^{b} \lambda_a P_a(\mathbf{s}) P_a(\mathbf{u}), \tag{11}$$

is still nnd. Indeed, an equivalent way to write (11) is in terms of a spatial process,

$$Z^{(b)}(\mathbf{s}) \equiv \sum_{a=1}^{b} \eta_a P_a(\mathbf{s}), \tag{12}$$

where $\boldsymbol{\eta} \equiv (\eta_1, \ldots, \eta_b)'$ is the random vector with $\mathrm{cov}(\boldsymbol{\eta}) = \mathrm{diag}(\lambda_1, \ldots, \lambda_b)$.

Because the original covariance has been truncated, one way to capture the lost covariation is to add back a simple random process:

$$Z(\mathbf{s}) \equiv \sum_{a=1}^{b} \eta_a P_a(\mathbf{s}) + \xi(\mathbf{s}). \tag{13}$$

It is common to choose $\xi(\cdot)$ to be a white-noise process, but it is straightforward to maintain some spatial structure in $\xi(\cdot)$ (Berliner, Wikle and Cressie, 2000). Clearly, the expression (13) could be thought of as a spatial-factor-analysis model (Christensen and Amemiya, 2001; Lopes, Salazar and Gamerman, 2008), although there are important differences in what is assumed known and what is estimated.

The definition (13) is directly expressed in terms of the random components of the model, and it is a very fertile way of constructing covariance functions: Specifically, replace $\{P_a(\cdot)\}$ with any set of known basis functions $\{S_a(\cdot)\}$, orthogonal or not; replace $\text{cov}(\boldsymbol{\eta}) = \text{diag}(\lambda_1, \ldots, \lambda_b)$ with any $b \times b$ positive definite (pd) matrix $\mathbf{K}$; and write $\text{cov}(\xi(\mathbf{s}), \xi(\mathbf{u})) = \sigma^2_\xi I(\mathbf{u} = \mathbf{s})$. Then

$$C(\mathbf{s}, \mathbf{u}) \equiv \mathbf{S}(\mathbf{s})'\mathbf{K}\mathbf{S}(\mathbf{u}) + \sigma^2_\xi I(\mathbf{u} = \mathbf{s}), \tag{14}$$

is a valid nonstationary univariate covariance model, where $\mathbf{S}(\cdot) \equiv (S_1(\cdot), \ldots, S_b(\cdot))'$. Cressie and Johannesson (2008) call this a Spatial Random Effects (SRE) model.

The generalisation of (14) to multivariate spatial processes is easiest to obtain from its expression in terms of random components. Here, the bivariate case shows its potential:

$$\begin{aligned} Z_1(\mathbf{s}) &= \mathbf{S}^{(1)}(\mathbf{s})'\boldsymbol{\eta}_1 + \xi_1(\mathbf{s}); \quad \mathbf{s} \in D, \\ Z_2(\mathbf{s}) &= \mathbf{S}^{(2)}(\mathbf{s})'\boldsymbol{\eta}_2 + \xi_2(\mathbf{s}); \quad \mathbf{s} \in D, \end{aligned} \tag{15}$$

where $\mathbf{S}^{(1)}(\cdot)$ and $\mathbf{S}^{(2)}(\cdot)$ are given spatial basis functions that are quite likely to be different for $Z_1(\cdot)$ and for $Z_2(\cdot)$, and $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ may have non-zero $\text{cov}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$; see Bradley, Holan and Wikle (2014). Note that (15) could be viewed as an errors-in-variables parameterisation (Christensen and Amemiya, 2002, 2003). Clearly, the implied covariance and cross-covariance functions are nonstationary, but their parameters can still be estimated from the multivariate spatial data.

Representations using "factor" processes, such as in (15), generalise many of the constructions outlined in GK-Section 2. Let $\{U_r(\cdot) : r = 1, \ldots, p\}$ be a set of independent univariate processes with mean 0, variance 1, and stationary correlation functions $\{\rho_r(\mathbf{h})\}$. Suppose further that $\{g_{qr}(\mathbf{h}) : q, r = 1, \ldots, p\}$ are square-integrable kernels; then a very general factor representation is,

$$Z_q(\mathbf{s}) \equiv \sum_{r=1}^{p} \int g_{qr}(\mathbf{u} - \mathbf{s}) U_r(\mathbf{u}) d\mathbf{u}. \tag{16}$$

The cross-covariances implied by (16) are:

$$C_{qr}(\mathbf{h}) \equiv \sum_{k=1}^{p} \iint g_{qk}(\mathbf{v}_1) g_{rk}(\mathbf{v}_2) \rho_k(\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{h}) d\mathbf{v}_1 d\mathbf{v}_2. \tag{17}$$

Clearly, asymmetry is present in all but the simplest cases, and the linear model of coregionalisation in GK-Section 2.1 is recovered by setting $g_{qr}(\mathbf{u}-\mathbf{s}) = A_{qr}\delta(\mathbf{u}-\mathbf{s})$ in (16), where $\delta(\cdot)$ is the Dirac delta function. The cross-covariance function given at the beginning of GK-Section 2.2 is also recovered by setting $\rho_k(\mathbf{h}) \equiv \rho(\mathbf{h})/p$, $g_{qk}(\mathbf{h}) \equiv k_q(\mathbf{h})$, and $g_{rk}(\mathbf{h}) \equiv k_r(\mathbf{h})$. There are many familiar special cases, and the "factor" processes $\{U_r(\cdot)\}$ in (16) do not even need to be independent.

5. COMMENTS ON THE DATA EXAMPLES

In GK-Section 6, several bivariate spatial models are implemented on data describing pressure and temperature. GK compare these models and assess which ones are best able to capture the dependence within and between the processes. This may be the first time such an exercise has been carried out; nevertheless, there are aspects of the analyses we would modify.

First, the various models under consideration contain different numbers of free parameters. For example, the parsimonious Matérn has six free parameters, while the nonstationary parsimonious Matérn has several hundred. In this context, the log-likelihood does not provide a useful comparison of model fit. We suggest that the Aikake Information Criterion (AIC) and its corrected version (AICc) are preferable when the number of free parameters differ across models (Hoeting et al., 2006; Lee and Ghosh, 2009).

Second, the other summaries (RMSE and CRPS) may not be indicative of model performance, since parameters estimated from the entire dataset were used. In our view, the flexibility, adaptability, and utility of a model can only be assessed (in the context of cross-validation) using data that has not been used for parameter estimation.

Finally, as we mentioned earlier (Section 1), the so-called nugget-effect matrix that consists of both measurement-error and micro-scale matrices needs to be modelled in both examples, and the possibility of non-diagonal contributions should be considered in GK-Section 6.2.

6. BIBLIOGRAPHIC NOTES

It is a big task to review multivariate geostatistics, and we appreciate that GK had limits on what they could cover. They selected a few topics that went beyond their core goal of reviewing cross-covariance functions, and to some of these topics we add the following bibliographic notes.

Nonstationarity for Factor Processes

Wikle (2010): Recall from Section 4 that in the univariate setting, reduced-rank covariance functions are very useful for big spatial data:

$$\text{cov}(Z(\mathbf{s}), Z(\mathbf{u})) = \mathbf{S}(\mathbf{s})'\mathbf{K}\mathbf{S}(\mathbf{u}) + v(\mathbf{s})I(\mathbf{u}=\mathbf{s}),$$

where $\mathbf{S}(\cdot)$ is a given b-dimensional ($b \ll n$) vector of spatial basis functions, $\mathbf{K}$ is an unknown $b \times b$ pd matrix, and $v(\mathbf{s}) > 0$. In Wikle's review of these rank-b covariance models, he points out their computational advantages for spatial prediction. A generalisation to multivariate covariance functions is straightforward; see Section 4. When considering global processes, such as in remote sensing applications, the nonstationarity of these models is an advantage.

Asymmetric Cross-Covariance Functions

The asymmetry in multivariate spatial processes may come from, say, preferable mineralisation of metals in an ore body or deposition of lighter particulate matter in the environment. Cross-covariance models should be able to detect such phenomena. The "shifted-lag model" is a natural way to capture sources of asymmetry and has a longer history than that indicated by the reference to Li and Zhang (2011) in GK-Section 5.1:

Ver Hoef and Cressie (1993, 1994): The shifted-lag model given in GK-(12) was proposed.

Majure and Cressie (1997): The shifted-lag model was estimated from variance-based cross-variograms.

Christensen and Amemiya (2001, 2002): A latent variable factor analysis model for a multivariate spatial process was based on the shifted-lag model.

Spatio-Temporal Covariance Functions

From the point of view of building covariance functions, the temporal dimension could be viewed simply as an extra "spatial" dimension. However, an alternative approach, based on the dynamical evolution of spatial processes in time, often allows optimal prediction to be carried out without explicitly constructing covariance models.

Wikle et al. (2001): This article uses conditional-probability modelling (in space and time) and reduced-rank models to achieve optimal spatio-temporal prediction. It bypasses the need for constructing spatio-temporal cross-covariance functions.

Cressie and Wikle (2011): In much of this book, cross-covariance functions are considered as derivative measures, from scientifically interpretable dynamical (multivariate) spatial models; see their pp. 418–425. A hierarchical dynamical approach is taken that yields optimal spatio-temporal prediction directly, without having to pass through multivariate covariance-function modelling.

In our view, hierarchical physical-statistical modelling of big, spatio-temporal, multivariate non-linear, non-Gaussian data will represent the next frontier.

REFERENCES

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, Second ed. CRC Chapman and Hall, Boca Raton, FL, USA.

Berliner, L. M., Wikle, C. K. and Cressie, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian Dynamic Modeling. *Journal of Climate* **13** 3953-3968.

Bradley, J. R., Holan, S. H. and Wikle, C. K. (2014). Mixed effects modeling for areal data that exhibit multivariate-spatio-temporal dependencies. arXiv: 1407.7479.

Christensen, W. F. and Amemiya, Y. (2001). Generalized shifted-factor analysis method for multivariate geo-referenced data. *Mathematical Geology* **33** 801-824.

Christensen, W. F. and Amemiya, Y. (2002). Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association* **97** 302-317.

Christensen, W. F. and Amemiya, Y. (2003). Modeling and prediction for multivariate spatial factor analysis. *Journal of Statistics and Inference* **115** 543-564.

Cressie, N. (1993). *Statistics for Spatial Data*, Revised ed. Wiley-Interscience, New York, NY, USA.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B* **70** 209-226.

Cressie, N. and Wikle, C. K. (1998). The variance-based cross-variogram: You can add apples and oranges. *Mathematical Geology* **30** 789-799.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ, USA.

Hoeting, J. A., Davis, R. A., Merton, A. A. and Thompson, S. E. (2006). Model selection for geostatistical models. *Ecological Applications* **16** 87-98.

Huang, C., Yao, Y., Cressie, N. and Hsing, T. (2009). Multivariate intrinsic random functions for cokriging. *Mathematical Geoscience* **41** 887-904.

Kang, E. L. and Cressie, N. (2013). Bayesian hierarchical ANOVA of regional climate-change projections from NARCCAP Phase II. *International Journal of Applied Earth Observation and Geoinformation* **22** 3-15.

Künsch, H. R., Papritz, A. and Bassi, F. (1997). Generalized cross-covariances and their estimation. *Mathematical Geology* **29** 779-799.

Lee, H. and Ghosh, S. K. (2009). Performance of information criteria for spatial models. *Journal of Statistical Computation and Simulation* **79** 93-106.

Li, B. and Zhang, H. (2011). An approach to modeling asymmetric multivariate spatial covariance structure. *Journal of Multivariate Analysis* **102** 1445-1453.

Lopes, H. F., Salazar, E. and Gamerman, D. (2008). Spatial dynamic factor analysis. *Bayesian Analysis* **3** 759-792.

Majure, J. and Cressie, N. (1997). Dynamic graphics for exploring spatial dependence in multivariate spatial data. *Geographical Systems* **4** 131-158.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Process*. McGraw-Hill, Columbus, OH, USA.

Royle, J. A., Berliner, L. M., Wikle, C. K. and Milliff, R. (1999). A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador sea. In *Case Studies in Bayesian Statistics IV* (R. E. Gatsonis, B. Kass, A. Carlin, A. Carriquiry, I. Gelman and M. Verdinelli, eds.) 367-382. Springer-Verlag, New York, NY, USA.

Ver Hoef, J. M. and Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology* **25** 219-240.

Ver Hoef, J. M. and Cressie, N. (1994). Errata: Multivariable spatial prediction. *Mathematical Geology* **26** 273-275.

Wikle, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 107-118. CRC Press, Boca Raton, FL, USA.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association* **96** 382-397.