Cross–Covariance Functions for Multivariate Geostatistics

Marc G. Genton

marc.genton@kaust.edu.sa
stsda.kaust.edu.sa
stat.kaust.edu.sa

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Based on joint work with William Kleiber

September 19, 2014

1. KAUST

- New graduate-level university located 50 miles north of Jeddah
- On the Red Sea
- Western style campus (14 miles^2) and encourages cultural diversity
- First classes in Fall 2009
- About 900 students & 130 faculty (will grow to 2000 & 220)
- More at: www.kaust.edu.sa
- Partnership with TAMU through IAMCS
- Past President of CalTech is new President of KAUST since July 1, 2013
- New faculty in statistics: Prof. Ying Sun (es.kaust.edu.sa)
- New website: stat.kaust.edu.sa
- Recruting students, postdocs and faculty in statistics

2. Motivation, Definitions, Properties

- Continuously indexed datasets with multiple variables have become ubiquitous in the geophysical, ecological, environmental and climate sciences
- **Example:** in environmental and climate sciences, monitors collect information on multiple variables such as temperature, pressure, wind speed and direction, and various pollutants
- **Example:** the output of climate models generate multiple variables, and there are multiple distinct climate models
- **Example:** physical models in computer experiments often involve multiple processes that are indexed by not only space and time, but also parameter settings
- **Key difficulty:** specifying the cross-covariance function, responsible for the relationship between distinct variables
- Cross-covariance functions must be consistent with marginal covariance functions such that the second order structure always yields a **nonnegative definite covariance matrix**

Definitions

- p-dimensional multivariate random field $\mathbf{Z}(\mathbf{s}) = \{Z_1(\mathbf{s}), \ldots, Z_p(\mathbf{s})\}^{\mathrm{T}}$ defined on $\mathbb{R}^d$, $d \geq 1$
- Gaussian with $\boldsymbol{\mu}(\mathbf{s}) = \mathrm{E}\{\mathbf{Z}(\mathbf{s})\}$ and cross-covariance matrix function $\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{cov}\{\mathbf{Z}(\mathbf{s}_1), \mathbf{Z}(\mathbf{s}_2)\} = \{C_{ij}(\mathbf{s}_1, \mathbf{s}_2)\}_{i,j=1}^{p}$ composed of functions $C_{ij}(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{cov}\{Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)\}$
- The covariance matrix $\boldsymbol{\Sigma}$ of $\{\mathbf{Z}(\mathbf{s}_1)^{\mathrm{T}}, \ldots, \mathbf{Z}(\mathbf{s}_n)^{\mathrm{T}}\}^{\mathrm{T}} \in \mathbb{R}^{np}$:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{C}(\mathbf{s}_1, \mathbf{s}_1) & \mathbf{C}(\mathbf{s}_1, \mathbf{s}_2) & \cdots & \mathbf{C}(\mathbf{s}_1, \mathbf{s}_n) \\ \mathbf{C}(\mathbf{s}_2, \mathbf{s}_1) & \mathbf{C}(\mathbf{s}_2, \mathbf{s}_2) & \cdots & \mathbf{C}(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}(\mathbf{s}_n, \mathbf{s}_1) & \mathbf{C}(\mathbf{s}_n, \mathbf{s}_2) & \cdots & \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n) \end{pmatrix}$$

  should be nonnegative definite: $\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a} \geq 0$ for any vector $\mathbf{a} \in \mathbb{R}^{np}$, any spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$, and any integer n
- **Second-order stationarity:** $\mathrm{cov}\{Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)\} = C_{ij}(\mathbf{h})$
- **Isotropy:** $\mathrm{cov}\{Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)\} = C_{ij}(\|\mathbf{h}\|)$
- Statistical tests of cross-covariance structure

Properties

- $\mathbf{\Sigma}$ must be symmetric, hence matrix functions must satisfy $\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{C}(\mathbf{s}_2, \mathbf{s}_1)^{\mathrm{T}}$, or $\mathbf{C}(\mathbf{h}) = \mathbf{C}(-\mathbf{h})^{\mathrm{T}}$ under stationarity
- Thus cross-covariance matrix functions usually not symmetric: $C_{ij}(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{cov}\{Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)\} \neq \mathrm{cov}\{Z_j(\mathbf{s}_1), Z_i(\mathbf{s}_2)\} = C_{ji}(\mathbf{s}_1, \mathbf{s}_2)$
- Collocated matrices $\mathbf{C}(\mathbf{s}, \mathbf{s})$, or $\mathbf{C}(\mathbf{0})$ under stationarity, are symmetric and nonnegative definite
- $|C_{ij}(\mathbf{s}_1, \mathbf{s}_2)|^2 \leq C_{ii}(\mathbf{s}_1, \mathbf{s}_1) C_{jj}(\mathbf{s}_2, \mathbf{s}_2)$, or $|C_{ij}(\mathbf{h})|^2 \leq C_{ii}(\mathbf{0}) C_{jj}(\mathbf{0})$
- But $|C_{ij}(\mathbf{s}_1, \mathbf{s}_2)|$ need not be less than or equal to $C_{ij}(\mathbf{s}_1, \mathbf{s}_1)$, or $|C_{ij}(\mathbf{h})|$ need not be less than or equal to $C_{ij}(\mathbf{0})$
- This is because the maximum value of $C_{ij}(\mathbf{h})$ is not restricted to occur at $\mathbf{h} = \mathbf{0}$, unless $i = j$
- **Separability:** $C_{ij}(\mathbf{s}_1, \mathbf{s}_2) = \rho(\mathbf{s}_1, \mathbf{s}_2) R_{ij}$ where $\rho(\mathbf{s}_1, \mathbf{s}_2)$ is a valid, nonstationary or stationary, correlation function and $R_{ij} = \mathrm{cov}(Z_i, Z_j)$ is the nonspatial covariance between variables i and j

3. Cross-Covariances built from Univariate Models

- **3.1 Linear model of coregionalization**
- Representation of the multivariate random field as a linear combination of r independent univariate random fields
- Resulting cross-covariance functions: $C_{ij}(\mathbf{h}) = \sum_{k=1}^{r} \rho_k(\mathbf{h}) A_{ik} A_{jk}, \quad \mathbf{h} \in \mathbb{R}^d,$ for an integer $1 \leq r \leq p$, where $\rho_k(\cdot)$ are valid stationary correlation functions and $\mathbf{A} = (A_{ij})_{i,j=1}^{p,r}$ is a $p \times r$ full rank matrix
- Only r univariate covariances $\rho_k(\mathbf{h})$ must be specified, thus avoiding direct specification of a valid cross-covariance matrix function
- When $r = 1$, the cross-covariance function is separable
- With a large number of processes, the number of parameters can quickly become unwieldy and resulting estimation difficult
- Smoothness of any component of multivariate random field restricted to that of the roughest underlying univariate process

- **3.2 Convolution methods**
- **Kernel convolution** method:

$$C_{ij}(\mathbf{h}) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k_i(\mathbf{v}_1)k_j(\mathbf{v}_2)\rho(\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{h})\mathrm{d}\mathbf{v}_1\mathrm{d}\mathbf{v}_2$$

  where the k_i are square integrable kernel functions and $\rho(\cdot)$ is a valid stationary correlation function
- This approach assumes that all the spatial processes $Z_i(\mathbf{s})$ are generated by the same underlying process, which is restrictive
- Except some special cases, requires Monte Carlo integration
- **Covariance convolution** method:

$$C_{ij}(\mathbf{h}) = \int_{\mathbb{R}^d} C_i(\mathbf{h} - \mathbf{k})C_j(\mathbf{k})\mathrm{d}\mathbf{k}$$

  where C_i are square integrable functions
- Except some special cases, requires Monte Carlo integration
- Example: when the C_i are Matérn correlation functions with common scale parameters, they are closed under convolution and this yields special case of the multivariate Matérn model

Cross-Covariances built from Univariate Models

- **3.3 Latent dimensions**
- **Idea:** create additional latent dimensions that represent the various variables to be modeled
- Each component i of the multivariate random field $\mathbf{Z}(\mathbf{s})$ is represented as a point $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{ik})^{\mathrm{T}}$ in $\mathbb{R}^k$, $i = 1, \ldots, p$, for an integer $1 \leq k \leq p$
- Then: $C_{ij}(\mathbf{s}_1, \mathbf{s}_2) = C\{(\mathbf{s}_1, \boldsymbol{\xi}_i), (\mathbf{s}_2, \boldsymbol{\xi}_j)\}, \quad \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^d$ where C is a valid univariate covariance function on $\mathbb{R}^{d+k}$
- If C is stationary: $C_{ij}(\mathbf{h}) = C(\mathbf{h}, \boldsymbol{\xi}_i - \boldsymbol{\xi}_j)$
- Example:
  $C_{ij}(\mathbf{h}) = \frac{\sigma_i \sigma_j}{\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\| + 1} \exp\left\{ \frac{-\alpha \|\mathbf{h}\|}{(\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\| + 1)^{\beta/2}} \right\} + \tau^2 I(i = j) I(\mathbf{h} = \mathbf{0})$
  where $\sigma_i > 0$ are marginal standard deviations, $\tau \geq 0$ is a nugget effect, and $\alpha > 0$ is a length scale
- Here $\beta \in [0, 1]$ controls the non-separability between space and variables, with $\beta = 0$ being the separable case

4. Matérn Cross-Covariance Functions

- Matérn class of positive definite functions has become the standard covariance model for univariate fields
- $\mathrm{M}(\mathbf{h} \,|\, \nu, a) = \frac{2^{1-\nu}}{\Gamma(\nu)}(a\|\mathbf{h}\|)^{\nu}\mathrm{K}_{\nu}(a\|\mathbf{h}\|)$
  where K_{ν} is modified Bessel function of order ν, $a > 0$ is length scale parameter that controls rate of decay of correlation at larger distances, while $\nu > 0$ is smoothness parameter that controls behavior of correlation near the origin
- Multivariate Matérn:
  $\rho_{ii}(\mathbf{h}) = \mathrm{M}(\mathbf{h} \,|\, \nu_i, a_i)$ and $\rho_{ij}(\mathbf{h}) = \beta_{ij}\mathrm{M}(\mathbf{h} \,|\, \nu_{ij}, a_{ij})$
- β_{ij} is a collocated cross-correlation coefficient i.e. strength of correlation between Z_i and Z_j at same location, $\mathbf{h} = \mathbf{0}$
- Conditions on model parameters $\nu_i, \nu_{ij}, a_i, a_{ij}$ and β_{ij} that result in a valid multivariate covariance class
- Parsimonious Matérn: $a_i = a_{ij} = a$, $\nu_{ij} = (\nu_i + \nu_j)/2$
- Full Matérn: $p = 2$ characterization, $p > 2$

5. Nonstationary Cross-Covariance Functions

- Geophysical, environmental and ecological spatial processes often exhibit spatial dependence that depends on fixed geographical features such as terrain or land use type, or dynamical environments such as prevailing winds
- Need nonstationary models: $\text{cov}\{Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)\} = C_{ij}(\mathbf{s}_1, \mathbf{s}_2)$
- Nonstationary LMC: $C_{ij}(\mathbf{s}_1, \mathbf{s}_2) = \sum_{k=1}^{r} \rho_k(\mathbf{s}_1, \mathbf{s}_2) A_{ik} A_{jk}$ or $C_{ij}(\mathbf{s}_1, \mathbf{s}_2) = \sum_{k=1}^{r} \rho_k(\mathbf{s}_1 - \mathbf{s}_2) A_{ik}(\mathbf{s}_1) A_{jk}(\mathbf{s}_2)$
- Nonstationary multivariate Matérn:

$$\rho_{ii}(\mathbf{s}_1, \mathbf{s}_2) \propto \mathrm{M}(\mathbf{s}_1, \mathbf{s}_2 \mid \nu_i(\mathbf{s}_1, \mathbf{s}_2), a_i(\mathbf{s}_1, \mathbf{s}_2))$$

$$\rho_{ij}(\mathbf{s}_1, \mathbf{s}_2) \propto \beta_{ij}(\mathbf{s}_1, \mathbf{s}_2) \mathrm{M}(\mathbf{s}_1, \mathbf{s}_2 \mid \nu_{ij}(\mathbf{s}_1, \mathbf{s}_2), a_{ij}(\mathbf{s}_1, \mathbf{s}_2))$$

- $\beta_{ij}(\mathbf{s}, \mathbf{s})$ is proportional to the collocated cross-correlation coefficient $\mathrm{cor}\{Z_i(\mathbf{s}), Z_j(\mathbf{s})\}$
- Covariance and kernel convolution can also be extended to result in nonstationary matrix functions

- **6.1 Asymmetric cross-covariance functions**
- All the stationary models described so far are symmetric i.e. $C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h})$, or equivalently, $C_{ij}(\mathbf{h}) = C_{ij}(-\mathbf{h})$.
- Although $C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$ by definition, the aforementioned properties may not hold in general
- **Key idea:** If $\mathbf{Z}(\mathbf{s}) = \{Z_1(\mathbf{s}), \ldots, Z_p(\mathbf{s})\}^{\mathrm{T}}$ has cross-covariance functions $C_{ij}(\mathbf{h})$, then $\{Z_1(\mathbf{s} - \mathbf{a}_1), \ldots, Z_p(\mathbf{s} - \mathbf{a}_p)\}^{\mathrm{T}}$ has cross-covariance functions $C_{ij}^{a}(\mathbf{h}) = C_{ij}(\mathbf{h} + \mathbf{a}_i - \mathbf{a}_j)$
- Constraint $\mathbf{a}_1 + \cdots + \mathbf{a}_p = \mathbf{0}$ or $\mathbf{a}_1 = \mathbf{0}$ to ensure identifiability
- Can render any stationary symmetric cross-covariance function asymmetric
- Asymmetric cross-covariance functions, when required, can achieve remarkable improvements in prediction over symmetric models

- **6.2 Compactly supported cross-covariance functions**
- Computational issues in the face of large datasets is a major problem in any spatial analysis, even more so in multivariate case, including likelihood calculations and/or co-kriging
- One approach is to induce sparsity in the covariance matrix, either by using a compactly supported covariance function as the model, or by covariance tapering
- Scale mixtures of the form: $C_{ij}(\mathbf{h}) = \int (1 - \|\mathbf{h}\|/x)_+^{\nu} g_{ij}(x) \mathrm{d}x$ where $\nu \geq (d+1)/2$ and $\{g_{ij}(x)\}_{i,j=1}^{p}$ forms a valid cross-covariance matrix function
- For instance, with $g_{ij}(x) = x^{\nu}(1 - x/b)_+^{\gamma_{ij}}$ where $\gamma_i > 0$ and $\gamma_{ij} = (\gamma_i + \gamma_j)/2$ we have the multivariate Askey taper

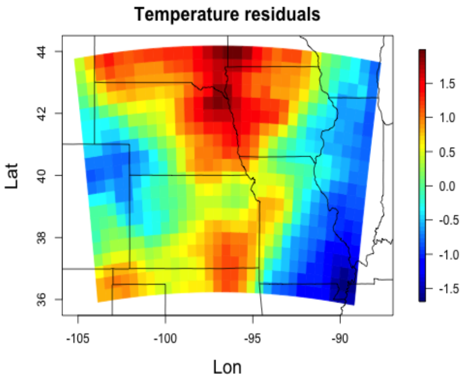
$$C_{ij}(\mathbf{h}) = b^{\nu+1} B(\gamma_{ij} + 1, \nu + 1) \left(1 - \frac{\|\mathbf{h}\|}{b}\right)^{\nu+\gamma_{ij}+1}, \quad \|\mathbf{h}\| < b$$

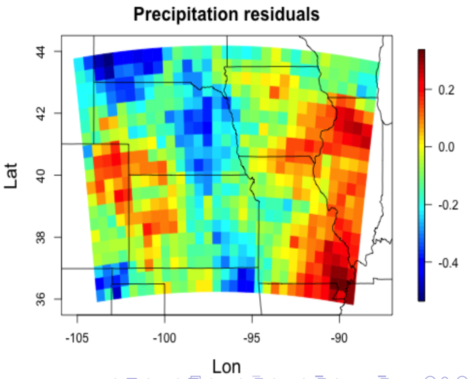
and 0 otherwise, where B is the beta function; extend to b_{ij}

Cross-Covariance Functions with Special Features

- **6.3 Cross-covariance functions on the sphere**
- Many multivariate datasets from environmental and climate sciences are collected over large portions of the Earth, for example by satellites, and therefore cross-covariance functions on the sphere $\mathbb{S}^2$ in $\mathbb{R}^3$ are in need
- Multivariate process on the sphere: $Z_i(L, l)$, $i = 1, \ldots, p$, with L=latitude and l=longitude
- Cross-covariance functions by applying differential operators with respect to latitude and longitude to process on the sphere
- Nonstationary models of cross-covariances with respect to latitude, so-called axially symmetric, and longitudinally irreversible cross-covariance functions: $\text{cov}\{Z_i(L_1, l_1), Z_j(L_2, l_2)\} \neq \text{cov}\{Z_i(L_1, l_2), Z_j(L_2, l_1)\}$
- Extensions from chordal distance to great circle distance

- **7.1 Climate model output data**
- North American Regional Climate Change Assessment Program (NARCCAP) climate modeling experiment
- Average summer (JJA) temperature and cube-root precipitation over a region of the midwest US
- 24 years (1981-2004) of residuals after removing spatially varying mean from each year's output for the two variables





Data Examples

- Temperature residuals smoother, precipitation rougher, both have similar correlation length scales
- Empirical correlation coefficient of -0.67

Table : Maximum likelihood estimates of parameters for full and parsimonious bivariate Matérn models, applied to the NARCCAP model data

Model	σ_T	σ_P	ν_T	ν_P	$1/a_T$	$1/a_P$	$1/a_{TP}$	ρ_{TP}
Full	1.63	0.19	1.31	0.55	384.3	361.6	420.1	-0.60
Parsimonious	1.61	0.19	1.33	0.54	367.1	-	-	-0.49

Table : Comparison of log likelihood values and pseudo cross-validation scores averaged over ten cross-validation replications for various multivariate models

	Log likelihood	RMSE (T)	CRPS (T)	RMSE (P)	CRPS (P)
Nonstationary parsimonious Matérn	53564.5	0.168	0.084	0.085	0.047
Parsimonious lagged Matérn	52563.7	0.179	0.090	0.087	0.048
Full Matérn	52560.1	0.178	0.090	0.087	0.048
Parsimonious Matérn	52556.9	0.179	0.090	0.087	0.048
Latent dimension	52028.8	0.180	0.091	0.088	0.049
LMC	51937.0	0.179	0.091	0.090	0.050
Independent Matérn	50354.5	0.180	0.091	0.088	0.049
Latent dimension example	48086.3	0.195	0.100	0.088	0.048

- **7.2 Observational temperature data**
- Bivariate minimum and maximum temperature data
- Observations available at stations from United States Historical Climatology Network over the state of Colorado
- Bivariate daily temperature residuals (having removed the state-wide mean) on September 19, 2004, a day with good network coverage with observations available at 94 stations
- Predictive RMSE and CRPS are improved by between $6 - 7\%$ when co-kriging using the parsimonious lagged Matérn, as compared to marginally kriging each variable

Table : Comparison of log likelihood values and pseudo cross-validation scores averaged over 100 cross-validation replications for various multivariate models

	Log likelihood	RMSE (min)	CRPS (min)	RMSE (max)	CRPS (max)
Parsimonious lagged Matérn	−414.0	3.18	1.83	3.14	1.79
Parsimonious Matérn	−414.9	3.22	1.85	3.16	1.80
LMC	−415.7	3.22	1.85	3.16	1.80
Latent dimension	−416.2	3.23	1.86	3.18	1.81
Latent dimension example	−419.1	3.24	1.86	3.17	1.81
Independent Matérn	−427.6	3.41	1.94	3.35	1.91

8. Discussion

- **8.1 Specialized cross-covariance functions**
- Nonstationary construction that allows individual variables to be a spatially varying mixture of short and long range dependence
- Various approaches to produce valid cross-covariance functions based on differentiation of univariate covariance functions and on scale mixtures of covariance matrix functions
- Constructions of variogram matrix functions
- Approach to building variogram matrix functions based on a univariate variogram model
- Approach to generating valid matrix covariances by considering stochastic partial differential equations
- For example: systems of SPDEs to simultaneously model temperature and humidity, yielding computationally efficient means to analysis by approximating a Gaussian random field by a Gaussian Markov random field

- **8.2 Spatio-temporal cross-covariance functions**
- Spatio-temporal multivariate random field, $\mathbf{Z}(\mathbf{s}, t)$, has stationary cross-covariance functions $C_{ij}(\mathbf{h}, u)$, where u denotes a time lag
- If $\varphi_1(t)$, $t \geq 0$, is a completely monotone function and $\psi_1(t), \psi_2(t)$, $t \geq 0$, are positive functions with completely monotone derivatives, then

$$C(\mathbf{h}, u, \mathbf{v}) = \frac{\sigma^2}{[\psi_1\{u^2/\psi_2(\|\mathbf{v}\|^2)\}]^{d/2}\{\psi_2(\|\mathbf{v}\|^2)\}^{1/2}} \varphi_1\left[\frac{\|\mathbf{h}\|^2}{\psi_1\{u^2/\psi_2(\|\mathbf{v}\|^2)\}}\right]$$

  is a valid stationary covariance function on $\mathbb{R}^{d+1+k}$ that can be used to model cross-covariance functions with $\mathbf{v} = \boldsymbol{\xi}_i - \boldsymbol{\xi}_j$
- First type of asymmetric spatio-temporal cross-covariance: $C^a_{ij}(\mathbf{h}, u) = C(\mathbf{h}, u - \boldsymbol{\lambda}_\xi^{\mathrm{T}}(\boldsymbol{\xi}_i - \boldsymbol{\xi}_j), \boldsymbol{\xi}_i - \boldsymbol{\xi}_j)$
- Second type of asymmetric spatio-temporal cross-covariance: $C^a_{ij}(\mathbf{h}, u) = C(\mathbf{h} - \gamma_h u, u, \boldsymbol{\xi}_i - \boldsymbol{\xi}_j - \gamma_\xi u)$

Discussion

- **8.3 Physics-constrained cross-covariance functions**
- A number of physical processes, especially in fluid dynamics, involve fields with specialized restrictions such as being divergence free
- Matrix-valued covariance functions for divergence-free and curl-free random vector fields
- Framework for valid matrix-valued covariance functions when the constituent processes have known physical constraints relating their behavior
- Spatio-temporal correlations for temperature fields arising from simple energy-balance climate models, that is, white-noise-driven damped diffusion equations. The resulting spatial correlation on the plane is of Matérn type with smoothness parameter $\nu = 1$, although rougher temperature fields are expected due to terrain irregularities for example
- Extension to other variables such as pressure and wind fields, and possibly lead to Matérn cross-covariance models?

Discussion

- **8.4 Open problems**
- **Theoretical characterization of the allowable classes of multivariate covariances:** given two marginal covariances, what is the valid class of possible cross-covariances that still results in a nonnegative definite structure?
- **Utility of cross-covariance models:** for the purposes of co-kriging, in what situations are the use of nontrivial cross-covariances beneficial?
- **Validity** of multivariate version of power exponential class of covariances?
- **Extension** of spatial extremes to the case of multiple variables?
- Valid multivariate cross-covariance functions for **spatial data on a lattice**

Genton, M. G., and Kleiber, W. (2014), "Cross-covariance functions for multivariate geostatistics," *Statistical Science*, in press

Discussion