Cross–Covariance Functions for Multivariate Geostatistics

Marc G. Genton

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

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
- Recruiting students, postdocs and faculty in statistics
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
  \[ Z(s) = \{Z_1(s), \ldots, Z_p(s)\}^T \] defined on $\mathbb{R}^d$, $d \geq 1$

- Gaussian with $\mu(s) = E\{Z(s)\}$ and cross-covariance matrix function $C(s_1, s_2) = cov\{Z(s_1), Z(s_2)\} = \{C_{ij}(s_1, s_2)\}_{i,j=1}^p$
  composed of functions $C_{ij}(s_1, s_2) = cov\{Z_i(s_1), Z_j(s_2)\}$

- The covariance matrix $\Sigma$ of $\{Z(s_1)^T, \ldots, Z(s_n)^T\}^T \in \mathbb{R}^{np}$:
  \[
  \Sigma = \begin{pmatrix}
  C(s_1, s_1) & C(s_1, s_2) & \cdots & C(s_1, s_n) \\
  C(s_2, s_1) & C(s_2, s_2) & \cdots & C(s_2, s_n) \\
  \vdots & \vdots & \ddots & \vdots \\
  C(s_n, s_1) & C(s_n, s_2) & \cdots & C(s_n, s_n)
  \end{pmatrix}
  \]
  should be nonnegative definite: $a^T \Sigma a \geq 0$ for any vector $a \in \mathbb{R}^{np}$, any spatial locations $s_1, \ldots, s_n$, and any integer $n$

- **Second-order stationarity:** $cov\{Z_i(s_1), Z_j(s_2)\} = C_{ij}(h)$

- **Isotropy:** $cov\{Z_i(s_1), Z_j(s_2)\} = C_{ij}(\|h\|)$

- Statistical tests of cross-covariance structure
Properties

- $\Sigma$ must be symmetric, hence matrix functions must satisfy $C(s_1, s_2) = C(s_2, s_1)^T$, or $C(h) = C(-h)^T$ under stationarity.

- Thus cross-covariance matrix functions usually not symmetric: $C_{ij}(s_1, s_2) = \text{cov}\{Z_i(s_1), Z_j(s_2)\} \neq \text{cov}\{Z_j(s_1), Z_i(s_2)\} = C_{ji}(s_1, s_2)$.

- Collocated matrices $C(s, s)$, or $C(0)$ under stationarity, are symmetric and nonnegative definite.

- $|C_{ij}(s_1, s_2)|^2 \leq C_{ii}(s_1, s_1)C_{jj}(s_2, s_2)$, or $|C_{ij}(h)|^2 \leq C_{ii}(0)C_{jj}(0)$.

- But $|C_{ij}(s_1, s_2)|$ need not be less than or equal to $C_{ij}(s_1, s_1)$, or $|C_{ij}(h)|$ need not be less than or equal to $C_{ij}(0)$.

- This is because the maximum value of $C_{ij}(h)$ is not restricted to occur at $h = 0$, unless $i = j$.

- **Separability:** $C_{ij}(s_1, s_2) = \rho(s_1, s_2)R_{ij}$

  where $\rho(s_1, s_2)$ is a valid, nonstationary or stationary, correlation function and $R_{ij} = \text{cov}(Z_i, Z_j)$ 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:
  \[
  C_{ij}(h) = \sum_{k=1}^{r} \rho_k(h)A_{ik}A_{jk}, \quad h \in \mathbb{R}^d, \text{ for an integer } 1 \leq r \leq p,
  \]
  where \( \rho_k(\cdot) \) are valid stationary correlation functions and \( A = (A_{ij})_{i,j=1}^{p,r} \) is a \( p \times r \) full rank matrix

- Only \( r \) univariate covariances \( \rho_k(h) \) 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:

\[ C_{ij}(h) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} k_i(v_1) k_j(v_2) \rho(v_1 - v_2 + h) dv_1 dv_2 \]

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 \( Z_i(s) \) are generated by the same underlying process, which is restrictive.

Except some special cases, requires Monte Carlo integration.

Covariance convolution method:

\[ C_{ij}(h) = \int_{\mathbb{R}^d} C_i(h - k) C_j(k) dk \]

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.
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 \( \mathbf{Z}(s) \) is represented as a point \( \xi_i = (\xi_{i1}, \ldots, \xi_{ik})^T \) in \( \mathbb{R}^k \), \( i = 1, \ldots, p \), for an integer \( 1 \leq k \leq p \)

Then: \( C_{ij}(s_1, s_2) = C\{(s_1, \xi_i), (s_2, \xi_j)\} \), \( s_1, s_2 \in \mathbb{R}^d \) where \( C \) is a valid univariate covariance function on \( \mathbb{R}^{d+k} \)

If \( C \) is stationary: \( C_{ij}(h) = C(h, \xi_i - \xi_j) \)

Example:
\[
C_{ij}(h) = \frac{\sigma_i \sigma_j}{\|\xi_i - \xi_j\| + 1} \exp \left\{ \frac{-\alpha \|h\|}{(\|\xi_i - \xi_j\| + 1)^{\beta/2}} \right\} + \tau^2 I(i = j) I(h = 0)
\]

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
  
  \[ M(h | \nu, a) = \frac{2^{1-\nu}}{\Gamma(\nu)} (a\|h\|)^\nu K_\nu(a\|h\|) \]

  where \( K_\nu \) is modified Bessel function of order \( \nu \), \( a > 0 \) 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:
  
  \[ \rho_{ii}(h) = M(h | \nu_i, a_i) \text{ and } \rho_{ij}(h) = \beta_{ij} M(h | \nu_{ij}, a_{ij}) \]

  \( \beta_{ij} \) is a collocated cross-correlation coefficient i.e. strength of correlation between \( Z_i \) and \( Z_j \) at same location, \( h = 0 \)

- Conditions on model parameters \( \nu_i, \nu_{ij}, a_i, a_{ij} \) and \( \beta_{ij} \) 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: \( \text{cov}\{Z_i(s_1), Z_j(s_2)\} = C_{ij}(s_1, s_2) \)
- Nonstationary LMC: \( C_{ij}(s_1, s_2) = \sum_{k=1}^{r} \rho_k(s_1, s_2) A_{ik} A_{jk} \) or \( C_{ij}(s_1, s_2) = \sum_{k=1}^{r} \rho_k(s_1 - s_2) A_{ik}(s_1) A_{jk}(s_2) \)
- Nonstationary multivariate Matérn:
  \[
  \rho_{ii}(s_1, s_2) \propto M(s_1, s_2 | \nu_i(s_1, s_2), a_i(s_1, s_2))
  \]
  \[
  \rho_{ij}(s_1, s_2) \propto \beta_{ij}(s_1, s_2) M(s_1, s_2 | \nu_{ij}(s_1, s_2), a_{ij}(s_1, s_2))
  \]
- \( \beta_{ij}(s, s) \) is proportional to the collocated cross-correlation coefficient \( \text{cor}\{Z_i(s), Z_j(s)\} \)
- Covariance and kernel convolution can also be extended to result in nonstationary matrix functions
6. Cross-Covariance Functions with Special Features

6.1 Asymmetric cross-covariance functions

All the stationary models described so far are symmetric i.e. $C_{ij}(h) = C_{ji}(h)$, or equivalently, $C_{ij}(h) = C_{ij}(-h)$.

Although $C_{ij}(h) = C_{ji}(-h)$ by definition, the aforementioned properties may not hold in general.

**Key idea:** If $Z(s) = \{Z_1(s), \ldots, Z_p(s)\}^T$ has cross-covariance functions $C_{ij}(h)$, then $\{Z_1(s - a_1), \ldots, Z_p(s - a_p)\}^T$ has cross-covariance functions $C_{ij}^a(h) = C_{ij}(h + a_i - a_j)$

Constraint $a_1 + \cdots + a_p = 0$ or $a_1 = 0$ 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:

\[ C_{ij}(h) = \int (1 - \|h\|/x)^\nu g_{ij}(x)dx \]

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.

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

and 0 otherwise, where \( B \) is the beta function; extend to \( b_{ij} \)
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 $S^2$ 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. Data Examples

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
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               | $\sigma_T$ | $\sigma_P$ | $\nu_T$ | $\nu_P$ | $1/a_T$ | $1/a_P$ | $1/a_{TP}$ | $\rho_{TP}$ |
|---------------------|------------|------------|---------|---------|---------|---------|------------|------------|
| 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

| Model                        | 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, $Z(s, t)$, has stationary cross-covariance functions $C_{ij}(h, u)$, 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

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

is a valid stationary covariance function on $\mathbb{R}^{d+1+k}$ that can be used to model cross-covariance functions with $v = \xi_i - \xi_j$.

First type of asymmetric spatio-temporal cross-covariance:

$C_{ij}^a(h, u) = C(h, u - \lambda_{\xi}(\xi_i - \xi_j), \xi_i - \xi_j)$

Second type of asymmetric spatio-temporal cross-covariance:

$C_{ij}^a(h, u) = C(h - \gamma_h u, u, \xi_i - \xi_j - \gamma_{\xi} u)$
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?
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

1. KAUST
2. Motivation, Definitions, Properties
3. Cross-Covariances built from Univariate Models
4. Matérn Cross-Covariance Functions
5. Nonstationary Cross-Covariance Functions
6. Cross-Covariance Functions with Special Features
7. Data Examples
8. Discussion