Cross-covariance Functions for Tangent Vector Fields on the Sphere

Minjie Fan[1] Tomoko Matsuo[2]

[1]Department of Statistics
University of California, Davis

[2]Cooperative Institute for Research in Environmental Sciences
University of Colorado, Boulder

ICSA/Graybill, 2015

Outline

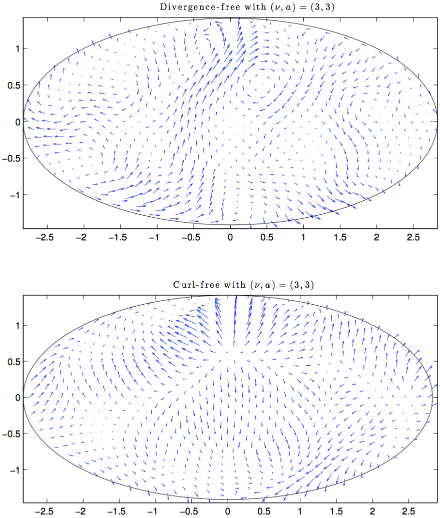
Tangent Vector Fields on the Sphere

- Suppose $\boldsymbol{Y}(\boldsymbol{s}) \in \mathbb{R}^3, \boldsymbol{s} \in \mathbb{S}^2$ is a random tangent vector field defined on the sphere.
- Let $\boldsymbol{Y}(\boldsymbol{s}) = \boldsymbol{\mu}(\boldsymbol{s}) + \boldsymbol{e}(\boldsymbol{s})$,
  $\boldsymbol{\mu}(\boldsymbol{s})$: non-random mean; $\boldsymbol{e}(\boldsymbol{s})$: zero-mean random process.
- Cross-covariance function: $\boldsymbol{C}(\boldsymbol{s}, \boldsymbol{t}) = \mathrm{Cov}(\boldsymbol{e}(\boldsymbol{s}), \boldsymbol{e}(\boldsymbol{t}))$, $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{S}^2$.

The Helmholtz-Hodge Decomposition

- Any tangent vector field can be uniquely decomposed into the sum of a divergence-free component and a curl-free component.
- Suppose the fields are velocity fields of fluids.
- Divergence-free: incompressible flow.
- Curl-free: irrotational flow.

An Example


Divergence-free with $(\nu, a) = (3, 3)$
1
0.5
0
−0.5
−1
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5
Curl-free with $(\nu, a) = (3, 3)$
1
0.5
0
−0.5
−1
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5

The Surface Gradient and the Surface Curl

- Suppose f is a scalar function defined on the sphere.
- The surface gradient of f at location $\boldsymbol{s}$,

$$\nabla_{\boldsymbol{s}}^{*} f =: \boldsymbol{P}_{\boldsymbol{s}} \nabla_{\boldsymbol{s}} f,$$

  where $\boldsymbol{P}_{\boldsymbol{s}} = \boldsymbol{I}_3 - \boldsymbol{s}\boldsymbol{s}^T$, and $\nabla_{\boldsymbol{s}}$ is the usual gradient on $\mathbb{R}^3$.
- The surface curl of f at location $\boldsymbol{s}$,

$$L_{\boldsymbol{s}}^{*} f =: \boldsymbol{Q}_{\boldsymbol{s}} \nabla_{\boldsymbol{s}} f,$$

  where

$$\boldsymbol{Q}_{\boldsymbol{s}} = \begin{pmatrix} 0 & -s_3 & s_2 \\ s_3 & 0 & -s_1 \\ -s_2 & s_1 & 0 \end{pmatrix}.$$

- $\nabla_{\boldsymbol{s}}^{*} f$ is curl-free and $L_{\boldsymbol{s}}^{*} f$ is divergence-free.

Outline

Main Idea

- Suppose $Z(\boldsymbol{s})$ is a sufficiently smooth univariate random field defined on the sphere. Besides, $Z(\boldsymbol{s})$ is stationary with mean zero and covariance function

  $$C(\boldsymbol{h}) = \mathrm{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{t})),$$

  where $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{t}$.
- Apply the surface gradient (or the surface curl) operator to the sample paths of $Z(\boldsymbol{s})$. The resulting random tangent vector field is curl-free (or divergence-free) in terms of its sample paths.

Main Idea (cont.)

- The cross-covariance functions of the tangent vector fields are
$$\boldsymbol{C}_{\mathrm{curl},Z}(\boldsymbol{s},\boldsymbol{t}) = -\boldsymbol{P}_{\boldsymbol{s}} \nabla_{\boldsymbol{h}} \nabla_{\boldsymbol{h}}^T C(\boldsymbol{h}) \Big|_{\boldsymbol{h}=\boldsymbol{s}-\boldsymbol{t}} \boldsymbol{P}_{\boldsymbol{t}}^T,$$
and
$$\boldsymbol{C}_{\mathrm{div},Z}(\boldsymbol{s},\boldsymbol{t}) = -\boldsymbol{Q}_{\boldsymbol{s}} \nabla_{\boldsymbol{h}} \nabla_{\boldsymbol{h}}^T C(\boldsymbol{h}) \Big|_{\boldsymbol{h}=\boldsymbol{s}-\boldsymbol{t}} \boldsymbol{Q}_{\boldsymbol{t}}^T.$$

Outline

Tangent Mixed Matérn

- Suppose $\boldsymbol{Z}(\boldsymbol{s}) = (Z_1(\boldsymbol{s}), Z_2(\boldsymbol{s}))^T$ is an isotropic bivariate Gaussian random field defined on the sphere with mean zero and cross-covariance function $\boldsymbol{C}(\|\boldsymbol{h}\|) = [C_{ij}(\|\boldsymbol{h}\|)]_{1\le i,j\le 2}$, where $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{t}$.
- The cross-covariance function satisfies a parsimonious bivaraite Matérn model [Gneiting et al., 2010].
- $C_{ii}(\|\boldsymbol{h}\|) = \sigma_i^2 M(\|\boldsymbol{h}\|; \nu_i, a), i = 1, 2,$
- $C_{12}(\|\boldsymbol{h}\|) = C_{21}(\|\boldsymbol{h}\|) = \rho_{12}\sigma_1\sigma_2 M(\|\boldsymbol{h}\|; (\nu_1 + \nu_2)/2, a)$.
- ρ_{12}, ν_1 and ν_2 satisfy a sufficient and necessary condition for non-negative definiteness.

Tangent Mixed Matérn (cont.)

- Based on the Helmholtz-Hodge decomposition, we construct a random tangent vector field on the sphere as

$$\underbrace{\nabla_{\boldsymbol{s}}^{*} Z_1(\boldsymbol{s})}_{\text{curl-free}} + \underbrace{L_{\boldsymbol{s}}^{*} Z_2(\boldsymbol{s})}_{\text{divergence-free}} .$$

- The cross-covariance function $\boldsymbol{C}_{\text{mix},\boldsymbol{Z}}(\boldsymbol{s},\boldsymbol{t})$ can be computed correspondingly.
- The smoothness parameters ν_1 and ν_2 are required to be larger than 1.

Propositions

- In climatology, tangent vector fields are often represented in terms of zonal and meridional components (i.e., u and v components).
- $$u(\theta, \phi) = \frac{1}{\sin\theta}\frac{\partial Z_1}{\partial \phi} + \frac{\partial Z_2}{\partial \theta} \quad \text{P-a.e.,}$$
  and
  $$v(\theta, \phi) = \frac{1}{\sin\theta}\frac{\partial Z_2}{\partial \phi} - \frac{\partial Z_1}{\partial \theta} \quad \text{P-a.e.}$$
- Axial symmetry: $\mathrm{Cov}(X(\theta_s, \phi_s), X(\theta_t, \phi_t)) = C(\theta_s, \theta_t, \phi_s - \phi_t)$.
- The u and v components are axially symmetric both marginally and jointly.

Propositions (cont.)

- Allows for negative covariances for the u and v components.
- This characteristic is very common in meteorological variables, such as wind fields [Daley, 1991].

Fast Computation

- The evaluation of the likelihood function requires $O(n^3)$ operations, where n is the number of sampling locations.
- If the observations are on a regular grid on the sphere, the discrete Fourier transform (DFT) can be used to speed up the computation [Jun, 2011].
- When $n_{\text{lat}} \sim n_{\text{lon}}$, the time complexity can be reduced to $O(n^2)$, where $n = n_{\text{lat}} n_{\text{lon}}$.
- $n_{\text{lat}} = 25$, $n_{\text{lon}} = 50$.
- Without using the DFT: 34.430 seconds.
- Using the DFT: 3.025 seconds.

Data Example

- Ocean surface wind dataset called QuikSCAT.
- Level 3 dataset.
- Monthly mean ocean surface winds from January 2000 to December 2008.
- The large-scale non-stationary variation is modeled by the method of empirical orthogonal function (EOF) analysis.
- Tangent Mixed Matérn is applied to the residual wind fields in the region of the Indian Ocean with a number of leading EOFs subtracted.

Residual Wind Fields

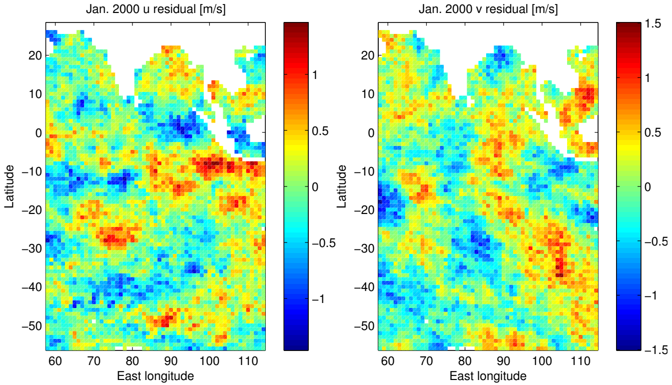


Figure : The u and v residual wind fields of January, 2000 in the region of the Indian Ocean.

List of Cross-covariance Models

- PARS-BM: parsimonious bivariate Matérn model.
- TMM: tangent Mixed Matérn model.

Table : Maximum likelihood estimates of parameters

Model	PARS-BM	TMM
σ_1^2	0.184 (2.85e-3)	-
σ_1	-	0.029 (4.25e-4)
σ_2^2	0.157 (2.24e-3)	-
σ_2	-	0.055 (8.05e-4)
ρ_{12}	-0.080 (6.58e-3)	0.281 (7.05e-3)
ν_1	1.239 (0.033)	1.758 (0.022)
ν_2	1.132 (0.032)	2.034 (0.020)
$1/a$	0.058 (1.24e-3)	-
a	-	9.472 (0.16)
τ_1	0.218 (1.3e-3)	0.210 (1.48e-3)
τ_2	0.203 (1.5e-3)	0.196 (1.48e-3)
Log-likelihood	-46995	**-45126**
# of parameters	8	8

Cokriging

- We randomly hold out the data at 170 locations for evaluation and estimate the parameters using the data at the remaining 900 locations.
- The cross-validation procedure is repeated 20 times.
- Scoring rules: the mean squared prediction error (MSPE) and the mean absolute error (MAE).

Table : Cokriging cross-validation results (20 times) in terms of the mean squared prediction error (MSPE) and the mean absolute error (MAE)

Model	Scoring Rule	U Residual Field [m/s]			V Residual Field [m/s]		
		Median	Max	Min	Median	Max	Min
PARS-BM	MSPE	0.0753	0.0803	0.0725	0.0680	0.0731	0.0651
	MAE	0.2157	0.2228	0.2117	0.2064	0.2120	0.2020
TMM	MSPE	0.0748	0.0803	0.0718	0.0671	0.0721	0.0642
	MAE	0.2150	0.2226	0.2107	0.2051	0.2106	0.2009

References

Roger Daley. *Atmospheric data analysis*. Cambridge: Cambridge University Press, 1991.

Tilmann Gneiting, William Kleiber, and Martin Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105 (491):1167–1177, 2010.

Mikyoung Jun. Non-stationary cross-covariance models for multivariate processes on a globe. *Scandinavian Journal of Statistics*, 38(4):726–747, 2011.