Paper Review:
"Nonstationary covariance models for global data" by Mikyoung Jun and Michael L. Stein, 2008[*]

Sungkyu Jung

Statistics & O. R., UNC Chapel Hill

April 27, 2009

1 Introduction

The paper by Jun and Stein discusses a class of covariance models that are designed to capture nonstationary covariance structures. In particular, they focus on data that cover a large portion of the Earth. One example of those data is the Level 3 Total Ozone Mapping Spectrometer (TOMS) data, which gives the daily total column ozone levels. In this data set, the ozone level is evaluated at the spatially regular grid. See Section 4.1 (of the paper) for detailed description of the data. Such data show a particular pattern in the covariance structure, which can be seen in Fig. 1 of the paper. Fig. 1 (a), (b) illustrate empirical standard deviations of the TOMS data calculated along the longitude and the latitude. It is clear by investigation of Fig. 1 that there is a strong dependence of the covariance structure along the latitude but not so clear along the longitude. These indicated features of the data will be regarded when building the general covariance model (See section 2).

We begin by defining distance functions on the Earth. A location on the Earth can be written as the 2-tuple of latitude (L) and longitude (l). There are two notions of distances for points on the surface of the Earth: the Euclidean distance and the great circle distance. *Chordal distance* is the Euclidean distance when the Earth is thought of as an embedded sphere in $\mathbb{R}^3$, and is defined by

$$d[(L_1, l_1), (L_2, l_2)] = 2R \left\{ \sin^2 \left(\frac{L_1 - L_2}{2} \right) + \cos L_1 \cos L_2 \sin^2 \left(\frac{l_1 - l_2}{2} \right) \right\}^{1/2}, \quad (1)$$

where R denotes the radius of the Earth. The great circle distance is the length of the shortest path on the surface and is given by

$$gc[(L_1, l_1), (L_2, l_2)] = 2R \arcsin \left(\frac{d[(L_1, l_1), (L_2, l_2)]}{2R} \right). \quad (2)$$

Note that since arcsin is one-to-one, these two distances are equivalent. In the paper, chordal distance is used mainly because it is easier to use a covariance structure defined in $\mathbb{R}^3$.

The class of *Legendre polynomials* is a class of orthogonal functions that can be defined on a sphere. Let P_i be the Legendre polynomial of order i such that $P_0(x) = 1$, $P_1(x) = (x+1)/2$, and so on for $x \in [-1, 1]$. Graphs of Legendre polynomials can be found, for example, at `http://en.wikipedia.org/wiki/Legendre_polynomials`. Define $P(L; p_0, \ldots, p_m) = \sum_{i=0}^m p_i P_i(\sin L)$, where $\sin L \in [-1, 1]$ for latitude L.

The rest of this manuscript is divided in two parts. First, a class of nonstationary covariance functions is defined for the spatial data covering a wide range of the Earth. Second, various models defined in Section 2 are fitted to level 3 TOMS data, and different models are compared in rather informal fashion.

2   Covariance models

A homogenous (i.e. stationary and isotropic) covariance model shall be defined first. Let $Z_0(L, l)$ be a spatial process on (the surface of) a sphere. Z_0 is said to be homogeneous if its covariance function K_0 only depends on the distance. I.e.

$$\text{Cov}(Z_0(L_1, l_1), Z_0(L_2, l_2)) = K_0(d[(L_1, l_1), (L_2, l_2)]), \tag{3}$$

where $d(\cdot, \cdot)$ is the chordal distance function. Moreover, set Z_0 as a mean zero process and let K_0 be a Matérn covariance function on $\mathbb{R}^3$, that is,

$$K_0(d[(L_1, l_1), (L_2, l_2)]) = \alpha(d/\beta)^\nu \mathcal{K}_\nu(d/\beta), \tag{4}$$

where α, β and ν denote the sill, range and smoothness parameter of the Matérn function, respectively, and $\mathcal{K}_\nu$ is the modified Bessel function of the third kind of order ν.

As can be seen in Fig. 1 (of the paper), a desirable property of the covariance function is the capability to capture the non-stationarity along latitude. A spatial process Z is said to be *axially symmetric* if Z is stationary with respect to longitude. The simplest model for axially symmetric Z is a rescaled version of Z_0 plus a nugget effect, i.e.

$$Z(L, l) = P(L; k_0, \ldots, k_m) Z_0(L, l) + \psi(L, l),$$

where ψ is a white noise process (with variance ϵ), and $P(L; k_0, \ldots, k_m)$ is a linear combination of *Legendre polynomials* up to order m, as defined in the previous section. The nugget effect is added on account of the measurement error. The covariance function of Z is

$$\begin{aligned}
&K_1(L_1, L_2, l; \alpha, \beta, \nu, \epsilon, k_0, \ldots, k_m) \\
&= P(L_1; k_0, \ldots, k_m) P(L_2; k_0, \ldots, k_m) K_0(d; \alpha, \beta, \nu) + \epsilon \mathbf{1}_{(L_1 - L_2 = l = 0)}.
\end{aligned} \tag{5}$$

By multiplying $P(L)$, K_1 can capture the strong dependence of variance along latitude L in a fairly simple model.

At this point, the authors define a new component that makes use of differential operators. A process Z is defined by applying differential operators to the homogenous process Z_0 (see (3)),

$$Z(L, l) = \left\{ A(L) \frac{\partial}{\partial L} + B(L) \frac{\partial}{\partial l} \right\} Z_0(L, l), \qquad (6)$$

where $A(L)$ and $B(L)$ are linear combinations of Legendre polynomials. The equation can be expressed and understood in stochastic integration form. It is not clearly specified how to calculate covariance function of this process exactly. The authors did not clearly present on this matter. However, a list of reference was available. Here are some observations from the references.

- The covariance function can be written in terms of derivatives of K_0. (Stein (2005))

- That A, B only depending on L makes Z axially symmetric. (Jun and Stein (2007))

- Z has different covariance structure along different latitudes.

These can be witnessed by some calculations. First note that $d[(L_1, l_1), (L_2, l_2)] = d'(L_1, L_2, l) = d'(L_1, L_2, -l)$ for $l = l_1 - l_2$. Moreover, $K_0(d[(L_1, l_1), (L_2, l_2)]) = K'_0(L_1, L_2, l)$. Then

$$\frac{\partial}{\partial l_i} K'_0(L_1, L_2, l) = \frac{\partial}{\partial l} K'_0(L_1, L_2, l),$$

for $i = 1, 2$, and

$$\frac{\partial^2}{\partial l_1 \partial l_2} K'_0(L_1, L_2, l) = \frac{\partial^2}{\partial l^2} K'_0(L_1, L_2, l).$$

Thus,

$$\begin{aligned}
&\mathrm{Cov}(Z(L_1, l_1), Z(L_2, l_2)) \\
&= A(L_1) A(L_2) \frac{\partial}{\partial L_1} \frac{\partial}{\partial L_2} K'_0(L_1, L_2, l) + A(L_1) B(L_2) \frac{\partial}{\partial L_1} \frac{\partial}{\partial l} K'_0(L_1, L_2, l) \\
&\quad + B(L_1) A(L_2) \frac{\partial}{\partial l} \frac{\partial}{\partial L_2} K'_0(L_1, L_2, l) + B(L_1) B(L_2) \frac{\partial^2}{\partial l^2} K'_0(L_1, L_2, l) \\
&\doteq K_Z(L_1, L_2, l)
\end{aligned}$$

The general axially symmetric nonstationary covariance function is defined by adding a covariance function K_Z to (5), i.e.

$$\begin{aligned}
&K_2(L_1, L_2, l; \alpha, \beta, \nu, \epsilon, k_0, \ldots, k_m, \alpha_1, \beta_1, \nu_1, a_0, \ldots, a_{n_1}, b_0, \ldots, b_{n_2}) \\
&= K_Z(L_1, L_2, l; \alpha_1, \beta_1, \nu_1, a_0, \ldots, a_{n_1}, b_0, \ldots, b_{n_2}) \\
&\quad + K_1(L_1, L_2, l; \alpha, \beta, \nu, \epsilon, k_0, \ldots, k_m) \qquad (7)
\end{aligned}$$

Then the number of parameters, or the complexity of the model, is controlled by m, n_1, and n_2.

3 Application: Level 3 TOMS data

Different values of m, n_1, and n_2 are used to fit Level 3 TOMS data of a particular day (May 14-15, 1990) by the likelihood approach. Different models are compared by 1) maximized likelihood values and by 2) comparison between empirical values and fitted values. A pre-processing is applied to the raw data, that is, the mean structure is subtracted and the residuals are tapered to have continuity.

The rescaled symmetric model (model B with $(m, n_1, n_2) = (3, 0, 0)$) increases likelihood by a significant amount, compared to the simplest homogenous model (model A with $(m, n_1, n_2) = (0, 0, 0)$). A bigger value of m does not help much in terms of likelihood. However, adding K_Z term increases likelihood also by a significant amount (e.g. model F with $(m, n_1, n_2) = (3, 3, 3)$, or model H with $(m, n_1, n_2) = (3, 6, 6)$). However model H compared to model F does not increase likelihood much, which can be understood that the required complexity of the Legendre polynomials is indeed fairly low.

Fig. 1 (a-c) (of the paper) shows clear improvements of fitting by adding more asymmetric terms in the model. Note that $K_2^{1/2}(\cdot, \cdot, l)$, averaged over latitude, can be added to panel (a), but is omitted since it is a constant function of l and is not a concern of this paper. Fitted curves of $K_2^{1/2}(L, L, 0)$ is illustrated with empirical values on Panel (b). Model F (or H) that has $n_1, n_2 > 0$ gives the best fit. Standard deviations of lag-Δl difference along longitude $(K_2^{1/2}(L, L, \Delta l))$ are plotted in panel (c) and model F (or H) gives the best fit.

We may need to develop more formal criteria (e.g. AIC, BIC) to penalize the number of parameters. However, the illustrated examples in the paper are enough to show the use of the axially symmetric covariance model for global data. These can naturally be extended to a spatial-temporal process as done in Jun and Stein (2007).

References

1 Jun, Mikyoung and Stein, Michael L. (2007). An approach to producing space-time covariance functions on spheres. *Technometrics* **49** 468–479.

2 Stein, Michael L. (2005). Space-time covariance functions. *J. Amer. Statist. Assoc.* **100** 310–321.