

The UNIVERSITY OF BRITISH
COLUMBIA
DEPARTMENT OF STATISTICS
TECHNICAL REPORT #250

BAYESIAN EMPIRICAL
ORTHOGONAL FUNCTIONS

BY

YIPING DOU
NHU D LE
JAMES V ZIDEK

JANUARY 2009

Bayesian Empirical Orthogonal Functions *

Yiping Dou, Nhu D Le and James V Zidek

July 20, 2009

Abstract

This report presents a Bayesian version of empirical orthogonal functions (EOFs) and thereby overcomes a difficulty with the classical version of these functions when an environmental process is autocorrelated as most are. The approach partitions the spatial variation into long and short scale variation, the latter including measurement errors. The most general version of our results incorporates the generalized Wishart prior distribution for the unknown covariance matrices and thereby gains considerable flexibility. In particular, the method can contend with situations in which the data exhibit a monotone pattern of missingness. The report includes a simulation study that demonstrates how the proposed method better characterizes spatial pattern by removing the deleterious effect of autocorrelation.

Keywords: Hierarchical Bayesian models; space–time fields; empirical orthogonal functions; orthogonal expansions.

1 Introduction

Empirical orthogonal functions (EOFs) have been used extensively to represent spatial patterns in environmental processes (Wikle 2002). However the common method for estimating these functions described in Section 2 can produce seriously distorted patterns when a space–time process is strongly autocorrelated. This paper presents a method for reducing those distortions.

To derive that method, some new theory is developed as an extension of a Bayesian spatial prediction approach of Dou et al.(2009a). That approach partitions the spatial variation into those with long scales like those captured by the EOFs and

*The research reported in this paper was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

those with short scales that include measurement errors. The most general version of our results incorporates the generalized inverted Wishart (GIW) prior distribution for the unknown covariance matrix and thereby gains considerable flexibility. In particular, the method can contend with situations where the data exhibit a monotone missing data pattern. However, the full generality of that extension is not needed in this report, which presents the results of simulation studies that demonstrate both the severity of distortions that can occur as well as how the proposed method better characterizes spatial pattern by removing those distortions.

As a summary of the report, Section 2 describes the common method for computing EOFs from environmental process data and discusses potential difficulties with this method. One of its subsections presents Simulation Study 1, which shows how distorted the spatial patterns can become. Another presents an alternative method that produces corrected EOFs for a known temporal covariance in a separable space–time covariance structure. Finally at the end of that section, Simulation Study 1 is revisited and the new method is seen to reduce those distortions. However, in practice uncertainty about the autocorrelation structure leads in Section 3 to the Bayesian EOF, the main contribution of this paper. In a subsection, those results are extended by use of the generalized inverted Wishart prior distribution. Another presents the results of Simulation Study 2 where the Bayesian approach is shown to deal successfully with the unknown autocorrelation. Finally, Section 5 summarizes the results and states conclusions.

2 Empirical orthogonal functions

This section reviews the common method of constructing the empirical orthogonal functions as estimates of components of a Karhunen–Loève (KL) expansion for a continuous spatial process observed at discrete time points. Unless otherwise stated, EOF will refer to the empirical orthogonal functions constructed by that method, although for clarity they are in called “classical” or “traditional” in some cases to distinguish them from other EOFs proposed below. To begin consider an arbitrary spatio–temporal process $\{Z(\mathbf{s}, t) : \mathbf{s} \in \mathcal{D}\}$, where \mathbf{s} represents for the spatial location in the finite domain $\mathcal{D} = \{\mathbf{s}_1, \dots, \mathbf{s}_p\}$ of interest while $t = 1, \dots, n$ represents a time point. Assume $E(Z(\mathbf{s}, t)) = 0$ and $Cov(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) = C_0(\mathbf{s}_i, \mathbf{s}_j)$ for all t . The KL expansion represents the covariance function as an infinite linear combination of orthogonal functions, that is,

$$C_0(\mathbf{u}, \mathbf{v}) = \sum_{j=1}^{\infty} \lambda_j \phi_j(\mathbf{u}) \phi_j(\mathbf{v}), \quad (1)$$

where $\{\lambda_j : j = 1, \dots, \infty\}$ are the eigenvalues and $\{\phi_j(\cdot) : j = 1, \dots, \infty\}$, the orthogonal eigenfunctions. For the complete set of orthonormal basis functions

$\{\phi_j(\cdot) : j = 1, \dots, \infty\}$, the response can be represented as follows:

$$Z(\mathbf{s}, t) = \sum_{j=1}^{\infty} a_j(t) \phi_j(\mathbf{s}), \quad (2)$$

where $Var(a_j(t)) > Var(a_{j+1}(t))$ for $j = 1, 2, \dots$, and $Cov(a_i(t), a_j(t)) = 0$ for $i \neq j$. In other words, the KL expansion allows one to represent the process by an infinite set of separable orthonormal basis functions that optimally minimize the mean square variance. The $\{\phi_j(\cdot)\}$ successively represent different spatial patterns of diminishing importance while the associated time series $\{a_j(\cdot)\}$ are projections of the process on the basis functions. The time series coefficients are called *principal time series components* (Wikle 2002) or *expansion coefficients*, the terminology in this paper.

Note that in the literature of climatology some authors define EOF differently. Going further, any one of the infinitely many orthonormal bases could be used to construct some version of the EOF. However the one we describe above is the only one that leads to the principal components PCA property, that is, the one having expansion coefficients of the EOFs that are uncorrelated (Björnsson and Venegas, 1997).

The KL integral equations can be solved approximately as a finite sum over equal sized grid cells. In that case the KL equations reduce to a system of linear equations. Their solution leads to a decomposition that is quite familiar in statistical science and at the heart of principal component analysis (PCA), albeit with a covariance matrix multiplied by the grid cell size.

However difficulties arise in practice since the continuous process $Y(\mathbf{s}, t)$ is usually observed not at the center of the grid cells above, but at an irregularly distributed discrete set of spatial points $\mathbf{s}_i \in D$, $i = 1, \dots, p$. Nevertheless PCA can be and is applied even though the result can no longer produce estimates of the (approximated) eigenvectors from the KL decomposition. According to Wikle (2002), this can give erroneous results in EOF analysis. In fact this

“distinction between EOFs on a regular grid and on an irregular grid is the source of many incorrect applications of the technique in the literature,”

according to Wikle. However this difficulty can also be avoided by appropriately weighting the estimated covariance matrix. Therefore this is not the issue addressed in this report.

Instead this paper addresses a flaw in the method of computing EOFs that derives from the way the covariance matrix is estimated. Usually this is done by implicitly assuming the independence of the time sequence of observed replicates of vectors, whose coordinates represent the responses at sites in a discretized spatial field. As a simulation study in this report shows, these EOFs can greatly misrepresent spatial

pattern when the vectors are strongly autocorrelated. This report also presents a Bayesian method that addresses this difficulty.

To describe the flaw, we first note that EOFs are computed from data anomalies rather than the values themselves. These anomalies are found by subtracting at each site $i \in \mathcal{D}$, the time average of measured responses at that site. More precisely, let $Z_i(t) \equiv Z(\mathbf{s}_i, t)$ for simplicity and the anomaly at site s_i and time t be defined as

$$Y_i(t) = Z_i(t) - \frac{1}{n} \sum_{t=1}^n Z_i(t), \quad (3)$$

for $i = 1, \dots, p$ and $t = 1, \dots, n$. Furthermore let $\mathbf{Y}_t = (Y_t(\mathbf{s}_1), \dots, Y_t(\mathbf{s}_p))' : p \times 1$ represent the anomalies vector at t across all the spatial sites in the region. Finally let $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n) : p \times n$ be the anomaly matrix.

The covariance estimate uses the anomalies described above, computed for a time sequence of successive realizations of a vector of responses over a discrete spatial domain. The estimate is just the usual covariance estimate obtained when the time sequence is regarded as unautocorrelated. The resulting estimate of $\{\phi_j(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$, obtained by PCA, is called the j^{th} EOF and $\{a_j(t) : t = 1, \dots, n\}$, the expansion coefficients corresponding to the j^{th} EOF.

Finding EOFs

Let $\mathbf{Z}_t = (Z_1(t), \dots, Z_p(t))' : p \times 1$ be the response vector at time t . Assume the matrix-variate response variable $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n) : p \times n$ follows a matrix-normal distribution with a separable covariance structure in space and time, that is, $\mathbf{Z} \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \Sigma_T)$, where $\Sigma_S : p \times p$ represents the spatial covariance matrix and $\Sigma_T : n \times n$, the temporal covariance matrix. This separable covariance structure implies no space-time interaction in spatial-temporal processes.

Based on these assumptions, the spatial covariance matrix can be estimated by $\hat{\Sigma}_S = \frac{1}{n} \mathbf{Y} \mathbf{Y}'$, where Y is the anomaly matrix defined in (3). The spectral decomposition theorem implies the existence of a unique decomposition for $\hat{\Sigma}_S$ such that $\hat{\Sigma}_S = \hat{\Psi} \hat{\Lambda}^2 \hat{\Psi}'$, where $\hat{\Lambda}^2 = \text{diag}\{\hat{\lambda}_1^2, \dots, \hat{\lambda}_p^2\}$ where $\hat{\lambda}_1 > \dots > \hat{\lambda}_p > 0$ are the eigenvalues for $\mathbf{Z} \mathbf{Z}'$ while each column of $\hat{\Psi}$ is the eigenvector corresponding to the associated eigenvalue. Hence, we can represent Σ_S as

$$\begin{aligned} \Sigma_S &= \Psi \Lambda^2 \Psi' \\ &= (\Psi \Lambda)(\Psi \Lambda)' \\ &= \Phi \Phi', \end{aligned} \quad (4)$$

a form of the KL expansion. We then obtain the EOFs from $\hat{\Phi} = \hat{\Psi} \hat{\Lambda}$.

However as noted above EOFs do not efficiently estimate the population level counter parts in Σ_S without temporal independence, an unrealistic assumption in most cases. The effects are demonstrated in the simulation study that follows.

Simulation study 1

This subsection presents the results of a simulation study involving simulated, separable state–space process with specified EOFs and temporal covariance matrices. These results show that as commonly calculated, EOFs can severely misrepresent spatial patterns.

For the study we need to construct an orthogonal matrix \mathbf{O} in a spectral decomposition, the Gram–Schmidt procedure being used to obtain an orthogonal basis. The EOFs specified in the simulation study can then be constructed to have a specified diagonal matrix $\mathbf{\Lambda}$ and an orthonormal basis function $\mathbf{\Phi}$. The construction starts with any given set of orthogonal vectors, say $\mathbf{O}_1, \dots, \mathbf{O}_k$, for $1 \leq k \leq n - 1$, and $k \in \mathcal{Z}$. Lemma 1 gives the details.

Lemma 1 *Given the orthogonal vectors $\mathbf{G}_j : p \times 1$, $j = 1, \dots, k$, we obtain an orthogonal matrix $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p) : p \times p$ by repeating steps (i)–(iii) for $j = k + 1, \dots, p$:*

(i) *Generate a realization \mathbf{y}_j from $N(\mathbf{0}, \mathbf{I}_p)$.*

(ii) *Fit the linear regression model:*

$$\mathbf{y}_j = \mathbf{A}_0 + \sum_{i=1}^{j-1} \mathbf{A}_i \mathbf{G}_i + \varepsilon_i,$$

and obtain the estimated coefficients $\{\hat{\mathbf{A}}_i : i = 0, \dots, j - 1\}$.

(iii) *Let \mathbf{G}_j be the fitted residuals $\mathbf{y}_j - \hat{\mathbf{A}}_0 - \sum_{i=1}^{j-1} \hat{\mathbf{A}}_i \mathbf{G}_i$ such that $\mathbf{G}_j \perp \{\mathbf{G}_1, \dots, \mathbf{G}_{j-1}\}$.*

This lemma gives us the orthonormal basis function by normalizing the generated orthogonal vectors $\mathbf{G}_1, \dots, \mathbf{G}_p$. We hereafter use \mathbf{O} to represent the orthonormal basis matrix using the Gram–Schmidt type expansion. Since $\mathbf{\Sigma}_S = \mathbf{O}\mathbf{\Lambda}^2\mathbf{O}'$, we then obtain the spatial covariance function $\mathbf{\Sigma}_S$ using the above constructed EOFs.

Simulated data

In this study the spatial region D is a square lattice consisting of 324 grid cells, each of size 6 km squared. The zero mean spatial–temporal process over D has a separable spatio–temporal covariance structure. After constructing the EOF matrix using Lemma 1 and specifying a diagonal matrix, the spatial covariance function is constructed from the spectral decomposition theorem. The temporal covariance function is a causal, invertible AR(1) process with variance σ_v^2 and AR coefficient, ϕ . The autocorrelation function (ACF) for the AR(1) process is given by

$$\gamma(h) = \begin{cases} \frac{\sigma_v^2}{1-\phi^2} & h = 0 \\ \frac{\sigma_v^2}{1-\phi^2} \phi^{|h|} & h \geq 1. \end{cases}$$

We begin with two primary EOFs specified as p - dimensional orthogonal vectors $\mathbf{G}^{(1)}$ and $\mathbf{G}^{(2)}$ where $p = 4q^2$ for some positive integer q . Purely for expository purposes we think of the first EOF as representing the north-south pattern in the northern hemisphere’s ground level winter temperature while the second, lesser EOF is the high - low contribution due to elevation. Thus

$$G_i^{(1)} = \begin{cases} 1, & i \in [1, \frac{p}{2}] \\ -1, & i \in [\frac{p}{2} + 1, p] \\ 0, & \text{otherwise.} \end{cases}$$

In contrast \mathbf{G}_2 has $(\mathbf{1}_{\sqrt{p}/2}, -\mathbf{1}_{\sqrt{p}/2})$ repeated \sqrt{p} times, $\mathbf{1}_{\sqrt{p}/2}$ being a $\sqrt{p}/2$ -dimensional vector of 1s. Then application of Lemma 1 yields the orthogonal matrix needed for our construction of the process covariance.

For the diagonal matrix $\mathbf{\Lambda}^2$ needed to complete that construction, we take the first four diagonal entries in the specified as 40, 20, 15, and 10, respectively. The remaining diagonal entries are a decreasing sequence which sums to 15 and has a minimum (final) value of 0.023. The specific choices we have made for our construction are pretty much arbitrary and qualitatively similar results to those below would be obtained for any choices following the same patterns.

Figures 1 and 2 plot the contours for simulated process data at $t = 5$ and $t = 28$, respectively in the regions we study. These graphs show the north–south and high–low spatial patterns. Moreover, this spatio–temporal field varies over space and time.

To examine the temporal variations in the simulated data, we randomly select four grid cells and plot their histograms, ACFs and PACFs in Figure 3. As expected these graphs show a very strong autocorrelation in the time series data in all cells.

Results and comparisons

We first compute the EOFs and compare them with the true EOFs. Figure 4 plots the contours for the true EOFs. As expected the first shows the north–south spatial pattern and the second, the high–low elevation spatial pattern as the two principal determinants of surface temperatures. However Figure 5 tells quite a different story. There the estimated contours deviate substantially from the true EOFs. For example while the values for the first true EOF in the northern region are close to be 0.35, the first EOF puts them close to 0.6 in the northeastern region and 1.0 at the northwestern region. Similar things are seen in comparing the true second EOF against second EOF found in the traditional way. We conclude that EOFs can fail to capture the true spatial patterns in a space–time field when the autocorrelation is high. The latter can intertwine itself with and distort the spatial pattern in this case.

The next section provides what we call “corrected” EOFs, a fix when the temporal covariance matrix, $\mathbf{\Sigma}_T$.

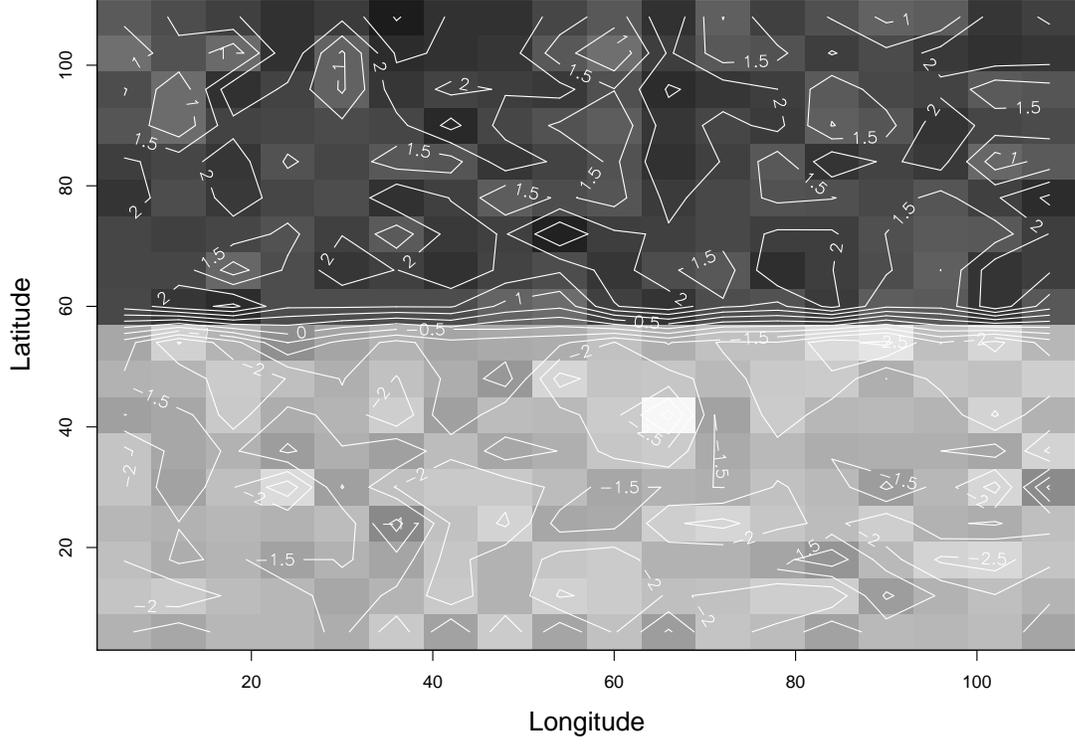


Figure 1: Contour plot for the simulated process data at day $t = 5$ over the 324 grid cell locations. The AR coefficient for the simulated data is $\phi = 0.9$. (White=-4.0; Black=4.0)

Corrected EOFs

If the autocovariance Σ_T were known, an alternative to the EOF that we call the “corrected EOF” would present itself. While that assumption is unrealistic, the analysis below based on it shows the serious distortions possible with the EOF, can be eliminated by filtering out that autocorrelation. In other words, the results demonstrate that those distortions result from effects that autocorrelation has on the spatial structure. Finally it was this analysis that suggested the methodology presented in the sequel.

To find the corrected EOFs, suppose given a matrix-variate normal distribution for $\mathbf{Y} \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \Sigma_T)$. Then $\mathbf{Y}^* = \mathbf{Y}\Sigma_T^{-1/2} \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \mathbf{I}_n)$, as a standard property of the matrix-variate normal distribution. It is then straightforward to estimate Σ_S using $\frac{1}{n} \sum_{t=1}^n \mathbf{Y}_t^* (\mathbf{Y}_t^*)' = \frac{1}{n} \mathbf{Y}^* (\mathbf{Y}^*)'$. In other words,

$$\hat{\Sigma}_S = \frac{1}{n} \mathbf{Y} \Sigma_T^{-1} \mathbf{Y}'. \quad (5)$$

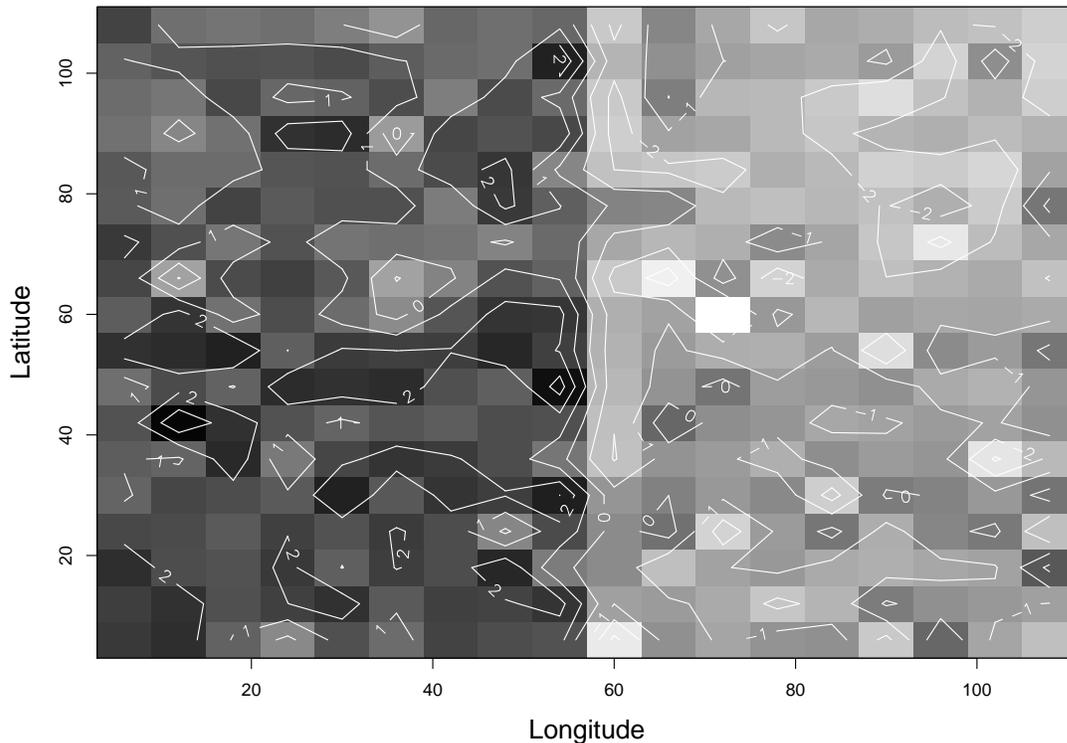


Figure 2: Contour plot for the simulated data at day $t = 28$ in the 18×18 grid locations. The AR coefficient in the simulated data is set to be $\phi = 0.9$.

The corrected EOFs are then constructed using the spectral decomposition theorem, as for EOFs. To obtain unique EOFs, we restrict the eigenvectors to form an orthonormal matrix that has positive elements in its first row.

Simulation study 1 revisited

This subsection revisits the example in Section 2 to compare the corrected EOFs with the EOFs with respect to the true EOFs. The goal is to see if the corrected EOFs represent better, the principal spatial patterns in the simulated data. Figures 4–5 show the first two EOFs for true, classical and corrected, respectively. These figures suggest that the corrected EOFs capture spatial patterns in the two true EOFs much better than the EOFs.

To get a more objective comparison, Table 1 presents the percentage of the spatial variation ascribable to the first ten EOFs by the true, traditional and corrected methods. The latter are much closer to the true values than the traditional ones. A different approach, which defines the distance between matrices through

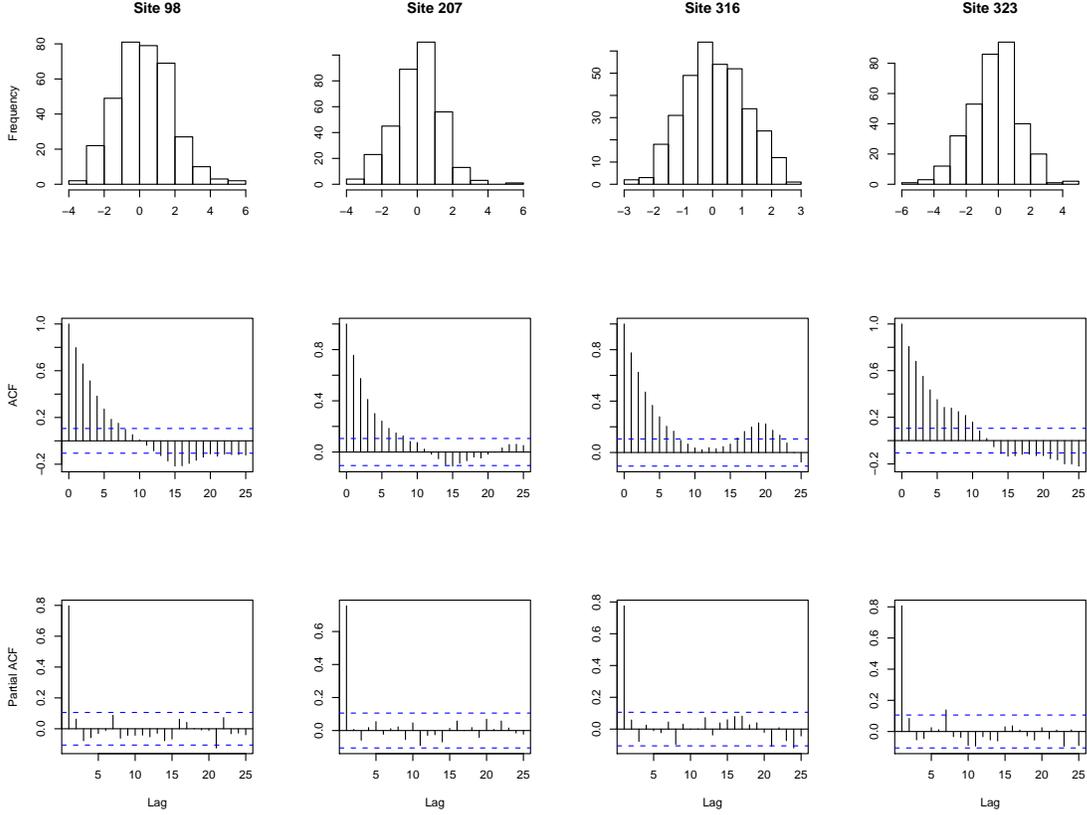


Figure 3: The histogram (first row), ACFs (second row), and PACFs (third row) for the simulated data at four randomly selected grid cells in the region. The AR coefficient for the simulated data is $\phi = 0.9$.

the Frobenius norm, leads to Table 2. There the separation between the traditional as well as corrected EOFs relative to the true EOFs, all represented as matrices, are tabulated. Again, the corrected EOFs prove to be closer than the classical ones to the true EOFs. Thus overall these quantitative assessments both point to the superiority of the corrected EOFs over the EOFs.

Conclusions

The standard method of finding the EOFs may lead to severe distortions in the primary spatial patterns of space–time fields. The corrected EOFs behave much better than the standard ones as we showed in Example 2. However, since the autocorrelations needed to make that correction will generally not be known, the method used above will not be of practical value. So how can the temporal patterns be taken into account in this case? In other words, can we find a good estimated substitute for Σ^T or possibly, a MCMC method whose posterior samples of Σ^T

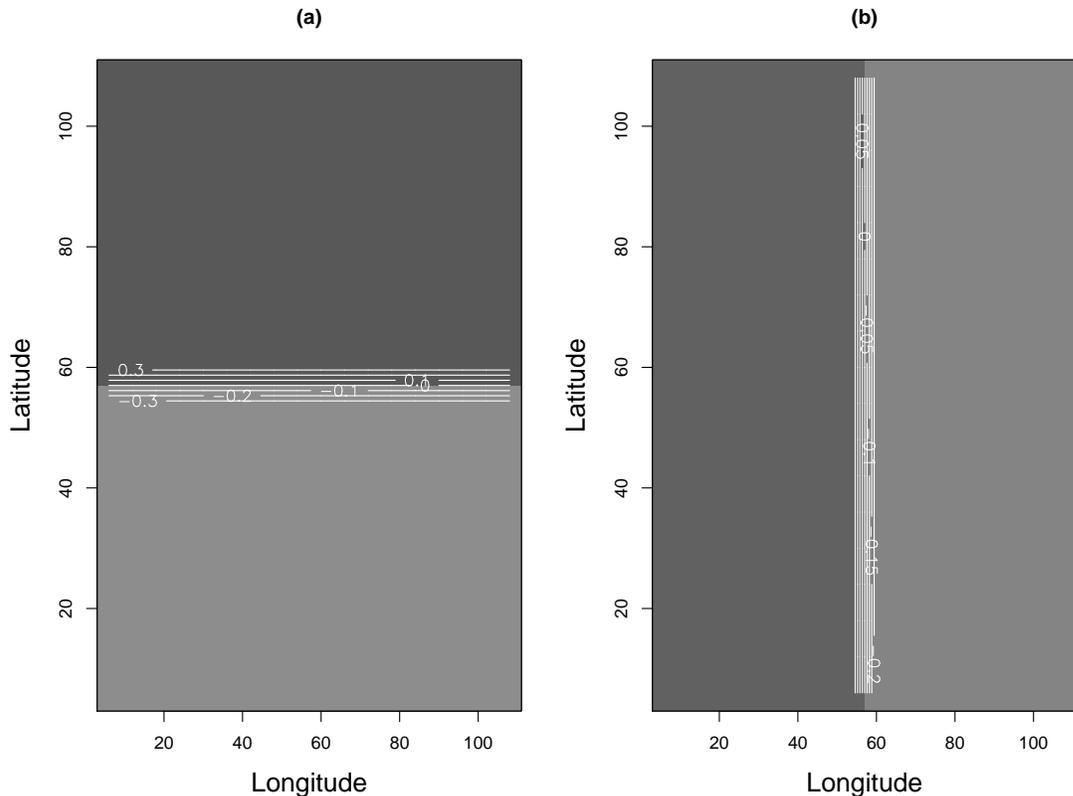


Figure 4: The contour plots for the first two true EOF vectors: (a) the first EOF; (b) the second EOF. (White=-1.7; Black=1.4)

converge to its true autocorrelation matrix value as the number of iterations goes to infinity? An answer is provided in the next section.

3 Bayesian EOFs

In practice the corrected EOFs are incomputable because of uncertainty about the required temporal covariance matrix for the spatial-temporal matrix-variate response $\mathbf{Z} : p \times n$, where p is the total number of spatial locations in the domain of interest and n , the total number of time points. Hence this section develops a Bayesian version of the corrected EOFs.

Underlying model

The Bayesian EOF model we adopt here is given by

$$\mathbf{Z} = \boldsymbol{\mu} \otimes \mathbf{1}'_n + \boldsymbol{\Phi} \mathbf{X} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_{p \times n}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{\Sigma}_\epsilon) \quad (6)$$

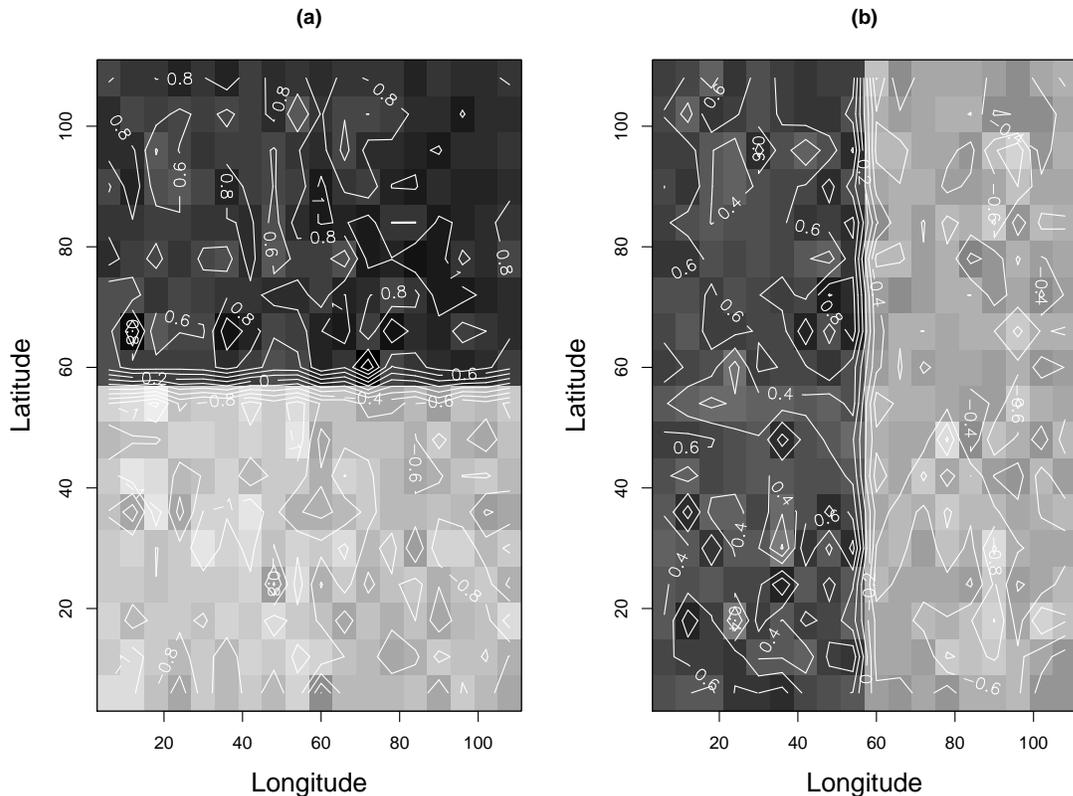


Figure 5: The contour plots for the first two EOFs: (a) the first EOF; (b) the second EOF. (White=-1.7; Black=1.4)

and

$$\mathbf{X} | \theta \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \Sigma_T) \quad (7)$$

where we assume both spatial and temporal covariance matrices Σ_S and Σ_T have full rank, that is, are nonsingular. Moreover, we consider two cases for the temporal covariance matrix Σ_T : (i) it is randomly distributed with an inverted Wishart distribution; and (ii) it has a semi-parametric form, that is, $\Sigma_T = \sigma^2 \rho(\cdot, \theta)$, with a known temporal correlation form for $\rho(\cdot, \cdot)$ but unknown parameters σ^2 and θ . We assume that the temporal correlation matrix, $\rho(\cdot, \theta)$, decreases as the difference between two time points increases, such that $\rho(t_i - t_j, \theta) \rightarrow 1$ as $t_i - t_j \rightarrow 0$. Moreover we are forced to assume $\sigma^2 = 1$ due to the non-identifiability property of the Kronecker product. The semi-parametric form for $\rho(\cdot, \theta)$ is then estimable using the MCMC method. For simplicity, we assume no small-scale spatial variation and measurement error, that is, no ϵ term in (6), deferring such refinements to future work.

In (6), Φ can be treated as a constant matrix for a known Σ_S by the Karhunen-

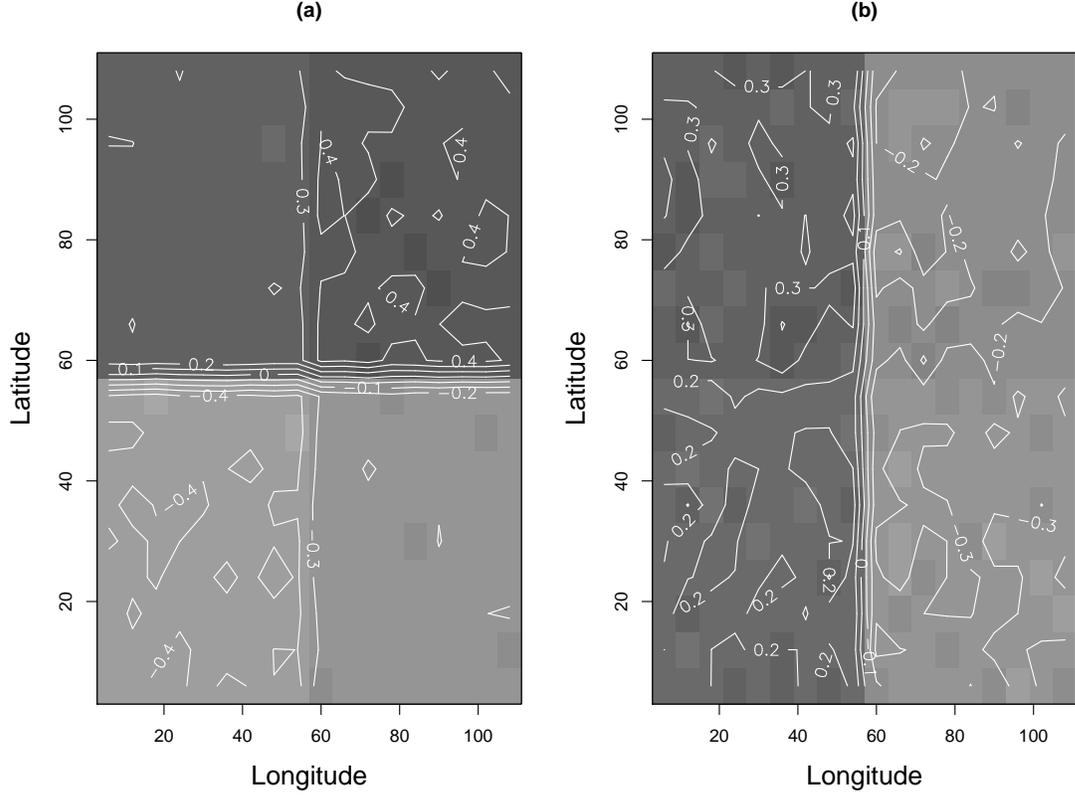


Figure 6: The contour plots for the first two corrected EOFs: (a) the first EOF; (b) the second EOF. (White=-1.7; Black=1.4)

Loeve (KL) expansion. Given the positive definite spatial covariance function $\Sigma_S : p \times p$, we have the unitary orthogonal matrix $\mathbf{O} : p \times p$, with its first row being positive, and diagonal matrix $\mathbf{\Lambda} = \text{diag} \{ \lambda_1, \dots, \lambda_p \} : p \times p$, with λ^2 's being the eigenvalues for Σ_S satisfying $\lambda_1 > \dots > \lambda_p > 0$, such that

$$\Sigma_S^{-1} = \mathbf{O} \mathbf{\Lambda}^{-2} \mathbf{O}'.$$

Hence, we have the Karhunen–Loeve (KL) expansion as follows:

$$\begin{aligned} \Sigma_S^{-1} &= (\lambda_1^{-1} \mathbf{O}^1, \dots, \lambda_p^{-1} \mathbf{O}^p) (\lambda_1^{-1} \mathbf{O}^1, \dots, \lambda_p^{-1} \mathbf{O}^p)' \\ &= (\Phi^1, \dots, \Phi^p) (\Phi^1, \dots, \Phi^p)' \\ &= \sum_{j=1}^p \Phi^j \Phi^{j'} \\ &= \Phi \Phi', \end{aligned}$$

where $\Phi_j = \lambda_j^{-1} \mathbf{O}^j$, for $j = 1, \dots, p$, and $\Phi = (\Phi^1, \dots, \Phi^p) : p \times p$.

Index of EOFs	True	Traditional	Corrected
1	40.000	40.833	39.507
2	20.000	20.230	20.178
3	15.000	15.890	15.072
4	10.000	9.165	10.267
5	0.023	0.544	0.189
6	0.024	0.519	0.183
7	0.024	0.481	0.178
8	0.024	0.434	0.178
9	0.024	0.431	0.173
10	0.024	0.405	0.171

Table 1: Percentage of spatial variation (%) captured by the first ten EOFs of the true, classical and corrected methods when the autocorrelation in the simulation study is high ($\rho = 0.9$).

Matrix discrepancies	
Traditional vs True	4.347
Corrected vs True	0.137

Table 2: Matrix discrepancies between the traditional and true EOFs as well as that discrepancy for the corrected EOFs.

However, when Σ_S is a random matrix, the EOFs represented by the columns of Φ are also random. Moreover, the orthogonal matrix can be treated as either constant or random in the KL expansion. The distribution for the random orthogonal matrix has been obtained by James (1954a), as an invariant uniform distribution on the Stiefel manifold. Moreover, he also obtained the independent distribution of the diagonal entries in Λ^2 in the KL expansion.

Then a Bayesian version of EOFs can be implemented either by using the MCMC method or through an empirical Bayes approach. The first level of a hierarchical model places no restriction on the form of our Bayesian EOFs, and so is a nonparametric approach. When the prior for the purely spatial covariance matrix, Σ_S , has been determined, Φ can be obtained using the KL expansion and Lemma 2 below.

Therefore, the Bayesian EOF model (6)–(7) is completed by specifying prior distributions for the model parameters: $p(\mu)$, $p(\Sigma_T)$ for Case (i) [or $p(\theta)$ for Case (ii)] and $p(\Sigma_S)$. Lacking specific prior information, we assume $p(\mu) \propto 1$ and $\Sigma_S^{-1} \sim W_p(\delta_S, \Xi_S)$, δ_S and Ξ_S being hyperparameters.

For Case (i), we assume $\Sigma_T^{-1} \sim W_n(\delta_T, \Xi_T)$. The collection of hyperparameters can be denoted by $\mathcal{H}_1 = \{\delta_S, \delta_T, \Xi_S, \Xi_T\}$. The joint posterior distribution we are interested in is given by $p(\mu, \Sigma_T, \Sigma_S | \mathbf{Y})$.

For Case (ii), we assume $\theta \sim N_k(\theta_0, \Sigma_0)$. The collection of hyperparameters here

can be denoted by $\mathcal{H}_2 = \{\delta_S, \Xi_S, \theta_0, \Sigma_0\}$. The joint posterior distribution we need for inference is then given by $p(\mu, \theta, \Sigma_S | \mathbf{Y})$.

In summary, we consider the following two Bayesian models to obtain the Bayesian EOFs for each of the two cases under consideration:

(i) The Bayesian model is given by

$$\mathbf{Z} = \mu \otimes \mathbf{1}'_n + \Phi \mathbf{X} \quad (8)$$

$$\mathbf{X} \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \Sigma_T) \quad (9)$$

$$\Phi = \mathbf{O} \Lambda^{-1}, \quad (10)$$

where $\Sigma_S^{-1} = \mathbf{O} \Lambda^{-2} \mathbf{O}'$ by the above KL expansion. And the priors for the model parameters are assumed mutually independent and given as follows:

$$p(\mu) \propto 1 \quad (11)$$

$$\Sigma_S^{-1} \sim W_p(\delta_S, \Xi_S) \quad (12)$$

$$\Sigma_T^{-1} \sim W_n(\delta_T, \Xi_T). \quad (13)$$

(ii) The Bayesian model is given as in (8)–(10), but Σ_T can be written as $\rho(\cdot, \theta)$, where $\rho(\cdot, \cdot)$ is the known temporal correlation matrix, decreasing as the difference between any two time points increases, and θ is an unknown parameter. The priors for μ and Σ_S^{-1} are given in (11)–(12), respectively, and they are mutually independent. Model specification is completed with the prior for the temporal covariance function given by:

$$\theta \sim N_k(\theta_0, \Sigma_0). \quad (14)$$

Notice that we assume both Σ_S and Σ_T are valid covariance matrices and non-singular, that is, the rank for Σ_S is p and for Σ_T , n .

Decomposing $\mathbf{Y} \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \mathbf{I}_n)$

We first describe the celebrated results of James (1954a), who discovers that the distribution of the random matrix $\mathbf{Y} \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \mathbf{I}_n)$ can be uniquely decomposed into three independent parts, specifically, one part being Wishart distributed, one part being uniformly distributed on a Grassmann manifold and the last part being uniformly distributed on a Stiefel manifold, that is, an orthogonal group in this setting. He constructs invariant measures on the orthogonal group (that is, the Haar measure), the Grassmann and Stiefel manifolds. James also finds the distribution of a non-central Wishart distribution using the Haar measure (1954b) and the distribution of the latent roots for a covariance matrix (1960). We now introduce some basic notation from James (1954a) and Chikuse (2006).

Definition 1 The orthogonal group, $O(n)$, is the set of all orthogonal matrices with the operation of matrix multiplication.

Definition 2 The Stiefel manifold, $V_{k,n} = \{\mathbf{V} : n \times k; \mathbf{V}'\mathbf{V} = \mathbf{I}_k\}$, is a set of k ($k \leq n$) orthonormal vectors in \mathcal{R}^n .

Definition 3 The Grassmann manifold, $G_{k,n-k}$, is the set of all k -dimensional hyperplanes in \mathcal{R}^n that pass through the origin.

James notes that the Grassmann and Stiefel manifolds may be regarded as coset spaces of the orthogonal group (1954, p.63), an important property in their relationships. The main result from James is summarized by the following lemma.

Lemma 2 (James, 1945) Suppose $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n) \sim N_{p \times n}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \mathbf{I}_n)$. Then we have

$$\mathbf{Y} = \mathbf{O}\mathbf{L}\mathbf{P}, \tag{15}$$

where $\mathbf{O} : p \times p$ represents an orthogonal matrix that is uniformly distributed over the Grassmann manifold, $\mathbf{P} : p \times n$, a semi-orthogonal matrix that is uniformly distributed over the Stiefel manifold and $\mathbf{L} : p \times p$, a diagonal matrix with diagonal entries $\{l_1, \dots, l_p\}$ such that l_1^2, \dots, l_p^2 are the eigenvalues for $\mathbf{Y}\mathbf{Y}'$ with $l_1 > \dots > l_p > 0$.

Mardia and Khatri (1977) develop the exact and asymptotic distributions for the random matrix uniformly distributed on a Stiefel manifold. They also discuss the matrix form of the von Mises–Fisher distribution on a Stiefel manifold. We will not discuss those applications of the work of James in this report but leave the construction of the posterior distributions on Stiefel and Grassmann manifolds for future research. Instead we use just the basic elements of his decomposition.

Theoretical results

We now present inference for Bayesian EOFs in the following lemmas and theorems. All proofs in this section are given in Appendix A.1.

Lemma 2 and the KL expansion below provide the basis for the theoretical results needed to obtain Bayesian EOFs. It leads to the prior distribution as a special case when $\boldsymbol{\Xi}_S = \mathbf{I}_p$, as shown in following lemma.

Lemma 3 If $\boldsymbol{\Sigma}_S^{-1} \sim W_p(n, \mathbf{I}_p)$ for some $n \in \mathcal{Z}^+$, by the KL expansion and Lemma 2, we have that the λ^{-2} 's are mutually independently distributed with $\lambda_j^{-2} \sim \chi_n^2$, for $j = 1, \dots, p$, and $\mathbf{O} : p \times p$ is uniformly distributed on the Stiefel manifold.

Lemma 3 provides one way to sample the random matrix Σ_S^{-1} from its prior distribution $W_p(\delta_S, \mathbf{I}_p)$.

Theorem 1 Consider the data matrix $\mathbf{Y} : p \times n \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \Sigma_T)$. Given the nonsingular spatial and temporal covariance matrices Σ_S and Σ_T , let $\mathbf{Y}^* = \mathbf{Y}\Sigma_T^{-1/2}$. Then $\mathbf{Y}^* \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \mathbf{I}_n)$. Consequently, $\mathbf{Y}^* = \Sigma_S^{1/2} \mathbf{O} \mathbf{L} \mathbf{P}$, where \mathbf{O} , \mathbf{L} , and \mathbf{P} are given in Lemma 2. The Bayesian EOFs are then obtained as $\mathbf{W} = \frac{1}{n} \Sigma_S^{1/2} \mathbf{O} \mathbf{L}$.

Theorem 2 Consider the random matrix $\mathbf{Y} : p \times n \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \Sigma_T)$. Suppose the temporal covariance matrix Σ_T is nonsingular and known. Then we have $\mathbf{Y}^* = \mathbf{Y}\Sigma_T^{-1/2} \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \mathbf{I}_n)$. Assume $\Sigma_S^{-1} \sim W_p(\delta_S, \Xi_S)$. The posterior distribution for the spatial precision matrix Σ_S^{-1} is given as follows:

$$\Sigma_S^{-1} | \mathbf{Y} \sim W_p(\delta^o, \Xi^o), \quad (16)$$

where $\delta^o = \delta_S + n$, and

$$\Xi^o = \Xi_S - \Xi_S \mathbf{Y} (\mathbf{Y}' \Xi_S \mathbf{Y} + \Sigma_T)^{-1} \mathbf{Y}' \Xi_S. \quad (17)$$

The Bayesian EOFs can be obtained when Σ_S is estimated or sampled from its posterior distribution.

Consequently from Theorem 2, we can either obtain the estimates for Σ_S using an empirical method such as that of the Sampson–Guttorp (SG) or sample it from its posterior distribution in (16). In other words, we can either use empirical Bayes or hierarchical Bayesian methods to obtain the estimates for the model parameters.

If the spatial covariance matrix were known, valid and nonsingular, we would have a similar result for the posterior distribution of the temporal covariance matrix, Σ_T . The next theorem tells us its posterior distribution if the prior for Σ_T^{-1} is assumed to be Wishart distributed, that is, Case (i).

Theorem 3 Consider the data matrix $\mathbf{Y} : p \times n \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \Sigma_T)$. Suppose the spatial covariance matrix Σ_S is known, valid and nonsingular. Assume the prior for Σ_T^{-1} is $W_n(\delta_T, \Xi_T)$. Then the posterior distribution for Σ_T^{-1} is given by

$$\Sigma_T^{-1} | \mathbf{Y} \sim W_n(\delta^*, \Xi^*), \quad (18)$$

where $\delta^* = \delta_T + p$ and

$$\Xi^* = \Xi_T - \Xi_T \mathbf{Y}' (\mathbf{Y} \Xi_T \mathbf{Y}' + \Sigma_S)^{-1} \mathbf{Y} \Xi_T. \quad (19)$$

In Case (ii) where the temporal covariance matrix is assumed to have a known parametric form but unknown parameters, we can obtain the posterior distribution for these parameters given a valid and known spatial covariance matrix. This posterior distribution is given in the following theorem.

Theorem 4 *Given the same condition as in Theorem 3 and the fully Bayesian model in Case (ii) with the priors for θ in (14), the conditional posterior distributions for θ are given as follows:*

$$p(\theta | \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} \left[\frac{1}{\sigma^2} \text{tr}(\mathbf{V}' \mathbf{V} \rho(\cdot, \theta)^{-1}) + (\theta - \theta_0)' \boldsymbol{\Sigma}_0^{-1} (\theta - \theta_0) \right] \right\}, \quad (20)$$

where $\mathbf{V} = \boldsymbol{\Sigma}_S^{-1/2} \mathbf{Y}$.

In practice, both temporal and spatial covariance matrices are unknown, and so conditions in Theorems 1 and 4 do not hold. Although the nonsingularity condition for these two covariance matrices might be difficult to verify due to challenging numerical problems, we need the condition to obtain the posterior samples for both matrices in this report. Future research will be devoted to improving on these results.

To obtain posterior samples for both parameters in the MCMC framework, we can either use a mixture of MCMC and empirical Bayes methods or use pure MCMC runs. We will illustrate the algorithm for both cases next. But the idea here is to obtain the conditional posterior samples for μ , $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_T$. The empirical Bayes method will be used to obtain the estimate for $\boldsymbol{\Sigma}_S$ given the data matrix, μ , and $\boldsymbol{\Sigma}_T$. We use the Gelman and Rubin R statistics as a device to check the convergence of the Markov chains (Gelman et al., 2004, p. 296–297). The estimates for the model parameters are then obtained as the mean of the posterior samples after the burn-in period.

We next develop the posterior conditional distributions for the model parameters for Cases (i) and (ii). The Bayesian EOFs can be obtained by Theorem 1 in which the mean field, spatial and temporal covariance matrices are estimated from their posterior distributions or by the empirical Bayes method.

Posterior conditional distributions

The posterior conditional distributions for μ , $\boldsymbol{\Sigma}_S^{-1}$, and $\boldsymbol{\Sigma}_T^{-1}$ can be obtained for Cases (i) and (ii) in the fully Bayesian framework. The proofs of the theorems in this section are presented in Appendix A.1. We first present the conditional posterior distributions for model parameters for Case (i) in the following theorem.

Theorem 5 *Given the Bayesian hierarchical model in (8)–(13), the posterior conditional distributions for these model parameters are given as follows:*

(i) *The conditional posterior distribution for μ is given by*

$$\mu | \mathbf{Z}, \boldsymbol{\Sigma}_S, \boldsymbol{\Sigma}_T \sim N_p(\mathbf{M}, \Sigma^* \boldsymbol{\Sigma}_S) \quad (21)$$

where $\Sigma^* = \{ \text{tr}(\mathbf{1}'_n \mathbf{1}_n \boldsymbol{\Sigma}_T^{-1}) \}^{-1}$ and $\mathbf{M} = \mathbf{Z} \boldsymbol{\Sigma}_T^{-1} \mathbf{1}'_n \Sigma^*$.

(ii) The conditional posterior distribution for Σ_S^{-1} is given by

$$\Sigma_S^{-1} | \mathbf{Z}, \mu, \Sigma_T \sim W_p(\delta_1, \Xi_1), \quad (22)$$

where $\delta_1 = \delta_S + n$,

$$\Xi_1 = \Xi_s - \Xi_s \mathbf{Y} (\mathbf{Y}' \Xi_s \mathbf{Y} + \Sigma_T)^{-1} \mathbf{Y}' \Xi_s, \quad (23)$$

and $\mathbf{Y} = \mathbf{Z} - \mu \otimes \mathbf{1}'_n$.

(iii) The conditional posterior distribution for Σ_T^{-1} is given by

$$\Sigma_T^{-1} | \mathbf{Z}, \mu, \Sigma_S \sim W_n(\delta_2, \Xi_2), \quad (24)$$

where $\delta_2 = \delta_T + p$,

$$\Xi_2 = \Xi_T - \Xi_T \mathbf{Y}' (\mathbf{Y} \Xi_T \mathbf{Y}' + \Sigma_S) \mathbf{Y} \Xi_T, \quad (25)$$

and $\mathbf{Y} = \mathbf{Z} - \mu \otimes \mathbf{1}'_n$.

In the same way, the posterior conditional distributions for the model parameters for Case (ii) are obtained in the next theorem.

Theorem 6 *Given the Bayesian hierarchical model in (8)–(12) and (14), the posterior conditional distributions for μ and Σ_S^{-1} are given in Theorem 5. Let $\mathbf{V} = \Sigma_S^{-1/2}(\mathbf{Z} - \mu \otimes \mathbf{1}'_n)$. Moreover, the posterior conditional distribution for θ is given by:*

$$p(\theta | \mathbf{Z}, \mu, \Sigma_S) \propto \exp \left\{ -\frac{1}{2} \left[\text{tr}(\mathbf{V} \mathbf{V}' \rho(\cdot, \theta)^{-1}) + (\theta - \theta_0)' \Sigma_0^{-1} (\theta - \theta_0) \right] \right\}. \quad (26)$$

After obtaining the estimates for the spatial and temporal covariance matrices, as well as other model parameters for both cases, the Bayesian EOFs can then be obtained by Theorem 1, Lemma 2 and the KL expansion.

Next we illustrate how the MCMC algorithm can be used to obtain the posterior samples from the joint posterior distribution $p(\mu, \Sigma_S^{-1}, \Sigma_T^{-1} | \mathbf{Z})$ and $p(\mu, \Sigma_S^{-1}, \theta | \mathbf{Z})$ for Cases (i) and (ii), respectively.

MCMC algorithms

To obtain the posterior samples from the joint posterior distribution of model parameters, μ, Σ_S and Σ_T (or θ), the MCMC algorithm is used to draw samples from their posterior distributions. For Case (i), we use Gibbs sampling method based on the full conditional distributions we obtained before. For Case (ii), a Metropolis–within–Gibbs algorithm is used because the posterior conditional distribution for θ does not have any closed form.

Algorithm 1 For Case (i), Gibbs sampling can be used to draw the posterior samples from $p(\mu, \Sigma_S, \Sigma_T | \mathbf{Z})$:

1. Initialization: set

$$\mu^{(1)} = \overline{\mathbf{Z}}^{row},$$

sample

$$\Sigma_S^{-1(1)} \sim W_p(\delta_S, \Xi_S),$$

and

$$\Sigma_T^{-1(1)} \sim W_n(\delta_T, \Xi_T).$$

2. Given the $(j-1)^{th}$ values, $\mu^{(j-1)}$, $\Sigma_S^{-1(j-1)}$, $\Sigma_T^{-1(j-1)}$, and \mathbf{Z} :

(1) Sample $\mu^{(j)}$ from $p(\mu | \mathbf{Z}, \Sigma_S^{(j-1)}, \Sigma_T^{(j-1)})$ from (21).

(2) Sample $\Sigma_S^{-1(j)}$ from $p(\Sigma_S^{-1} | \mathbf{Z}, \mu^{(j)}, \Sigma_T^{(j-1)})$ from (22).

(3) Sample $\Sigma_T^{-1(j)}$ from $p(\Sigma_T^{-1} | \mathbf{Z}, \mu^{(j)}, \Sigma_S^{(j)})$ from (24).

3. Repeat until convergence.

The Metropolis–within–Gibbs algorithm is omitted here because we present a similar result in Dou et al. (2007, 2008). In this section, we give the algorithm for a very special case when the temporal process is assumed to be an AR(1) process. Hence we have ϕ as the parameter that characterizes the AR(1) process. Assume that $\phi \sim N(\phi_0, \sigma_{\phi_0}^2)$. Then the collection of hyperparameters can be denoted by $\mathcal{H} = \{\phi_0, \sigma_{\phi_0}^2, \delta_S, \Xi_S\}$. Since there is no closed form for the posterior conditional distribution in (26), the Metropolis–within–Gibbs algorithm could then be used to draw posterior samples of interest.

The next section includes a straightforward extension on the Bayesian EOFs results where its spatial covariance is assumed to have a GIW distribution instead of IW (see Le and Zidek (2006), for example).

Extension: Bayesian EOFs

We can extend the above results about Bayesian EOFs to incorporate the GIW prior for the spatial covariance structure in such a way that the GIW prior reflects some characteristics of the data matrix. Le & Zidek (1992–2006) develop theoretical results for modelling spatio–temporal processes, i.e., the BSP approach in Dou et al. (2008, 2009a).

In such a Gaussian GIW framework, the spatial covariance matrix Σ_S can be estimated by the SG–method or Damian SG–method (Damian et al., 2002). Therefore, the estimates for the Σ_S can be updated at each iteration in the MCMC sampling. This will be carried out in future work.

The next section includes two simulation examples that help assess the performance for the Bayesian, classical and corrected EOFs.

4 Simulation study 2

The Bayesian EOF models we consider above have a very general structure. Note that ϵ in (6) represents small scale spatial variation or measurement error. If ϵ is close to $\mathbf{0}$, or equivalently to a very small value for Σ_ϵ , we then have approximately $\mathbf{Y} = \mu \otimes \mathbf{1}'_n + \Phi \mathbf{X}$, where Φ 's columns are the EOFs for Σ_S and $\mathbf{X}|\theta \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \Sigma_T(\theta))$. If $\Sigma_T(\theta) = \mathbf{I}_n$, then we have the classical EOFs in Section 2. If $\Sigma_T(\theta) \neq \mathbf{I}_n$ but known, we then have the corrected EOFs in Section 2. If $\Sigma_T(\theta)$ is unknown, we can use the Bayesian EOFs obtained in Section 3.

This section compares three different types of EOFs for a separable state–space process with specified spatial and temporal covariance matrices. To do that, we first simulate the matrix–variate data set. We then compute these three EOFs and compare them with the true EOFs by contour plots and the matrix discrepancies or separations based on the distance obtained from the Frobenious norm.

To briefly review these three types of EOFs, suppose $\mathbf{Y} : p \times n$ represents the anomaly matrix for p sites and n time points, and follows a multivariate normal distribution $N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \Sigma_T)$. Recall that the classical EOFs estimate the sample spatial covariance matrix by $\frac{1}{n} \mathbf{Y} \mathbf{Y}'$. Given the temporal dependence structure, that is, the temporal covariance matrix, the corrected EOFs estimate the sample spatial covariance matrix by $\mathbf{Y} \Sigma_T^{-1} \mathbf{Y}'$. Given the priors for the spatial and temporal covariance matrix, the Bayesian EOFs estimate the sample spatial covariance matrix in the hierarchical model by means of the corresponding posterior mode.

We consider two cases in this section to assess the performance of the EOFs for two different temporal dependence structures. For both cases, we assume a separable space time covariance structure, that is, an exponential spatial covariance function and an AR(1) temporal covariance function. In particular, the spatial covariance function is given by

$$(\Sigma_S)_{ij} = \exp(-V_{ij}/\lambda), \quad (27)$$

where V_{ij} is the Euclidean distance between \mathbf{s}_i and \mathbf{s}_j , for $i, j = 1, \dots, p$ and λ , a scale parameter. The temporal covariance function between t_k and t_l is given by $\sigma_v^2 \phi^{2||t_k - t_l||}$, for $t_k, t_l \in \{1, \dots, T\}$. Note that $|\phi| < 1$ corresponds to a causal AR(1) process. If $\phi \simeq 0$, then \mathbf{y}_t are approximately independent; if $\phi \simeq 1$, then

$\{\mathbf{y}_t : t = 1, \dots, n\}$ is a highly autocorrelated AR(1) process. We consider Case (i), $\phi = 0.1$, and Case (ii), $\phi = 0.9$.

The geographical region in this simulation study is set to be $[0.1, 1.0] \times [0.1, 1.0]$. We select 100 grid points in this region to be the locations of interest, i.e., $p = 100$. We then choose $n = 120$ time points at each of these 100 sites.

The initial settings for the separable space–time covariance functions are given as follows: $\lambda = 0.4$, $\sigma_v^2 = 1.0$, and $\phi = 0.1$ for Case (i), and 0.9 for Case (ii).

We define $Y(\mathbf{s}, t) = Z(\mathbf{s}, t) - \hat{\mu}(\mathbf{s})$, the anomaly at site \mathbf{s} and time t , where $\hat{\mu}(\mathbf{s}) = \frac{1}{n} \sum_{t=1}^n Z(\mathbf{s}, t)$. We obtain the classical, corrected, and Bayesian EOFs using $Y(\mathbf{s}, t)$.

We now compare the corrected, classical and “true” EOFs in the these two cases: (i) $\phi = 0.1$, and (ii) $\phi = 0.9$, respectively.

Simulated data

Suppose $\mathbf{Y} : p \times n \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \Sigma_T)$. One way to generate the simulated data is by first simulating $\mathbf{Y}^* = \mathbf{Y} \Sigma_T^{-1/2} = (\mathbf{y}_1^*, \dots, \mathbf{y}_n^*)$. Thus, $\mathbf{y}_t^* \sim N_p(\mathbf{0}, \Sigma_S)$, independently for $t = 1, \dots, n$. We then generate \mathbf{Y} by $\mathbf{Y}^* \Sigma_T^{1/2}$.

An alternative for obtaining the simulated data uses James’s result and the KL expansion. Given both Σ_S and Σ_T , we first illustrate this method for generating the simulated data in any given region. Given the spatial covariance matrix Σ_S , the Karhunen–Loeve expansion gives the unique orthogonal matrix $\mathbf{O} : p \times p$ with a positive first row and the unique diagonal matrix $\Lambda^2 : p \times p$ with decreasing diagonal entries, the eigenvalues of Σ_S such that

$$\begin{aligned}
\Sigma_S &= \mathbf{O} \Lambda^2 \mathbf{O}' \\
&= (\Lambda \mathbf{O}')' \Lambda \mathbf{O}' \\
&= (\lambda_1 \mathbf{O}^{(1)}, \dots, \lambda_p \mathbf{O}^{(p)}) (\lambda_1 \mathbf{O}^{(1)}, \dots, \lambda_p \mathbf{O}^{(p)})' \\
&= (\Phi^{(1)}, \dots, \Phi^{(p)}) (\Phi^{(1)}, \dots, \Phi^{(p)})' \\
&= \sum_{j=1}^p \Phi^{(j)} (\Phi^{(j)})' \\
&= \Phi \Phi',
\end{aligned} \tag{28}$$

where $\Phi = (\Phi^{(1)}, \dots, \Phi^{(p)}) : p \times p$ and $\Phi^{(j)} = \lambda_j \mathbf{O}^{(j)}$ for $j = 1, \dots, p$.

Given $\Sigma_T : n \times n$, the corresponding Karhunen–Loeve expansion is given by

$$\begin{aligned}
\Sigma_T &= \mathbf{P} \mathbf{L}^2 \mathbf{P}' \\
&= \sum_{j=1}^p \Psi^{(j)} (\Psi^{(j)})' \\
&= \Psi \Psi',
\end{aligned} \tag{29}$$

where $\mathbf{P} : n \times n$ represents the orthogonal matrix with positive first row, \mathbf{L}^2 is a diagonal matrix with decreasing but positive diagonal entries l_1^2, \dots, l_n^2 ; $\Psi = (\Psi^{(1)}, \dots, \Psi^{(n)})$ and $\Psi^{(i)} = l_i \mathbf{P}^{(i)}$ for $i = 1, \dots, n$.

Consequently, if $\mathbf{Y} : p \times n \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \Sigma_T)$, and $\Xi \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \mathbf{I}_n)$. Thus $\mathbf{Y} = \Phi \Xi \Psi'$ where Φ and Ψ are given by (28) and (29), respectively. Let $\Xi^* = \Xi \Psi'$. Then $\Xi^* \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \Sigma_T)$ and $\mathbf{Y} = \Phi \Xi^*$.

In short, the simulated data can be generated for known spatial and temporal covariance matrices as follows:

- (i) Uniquely obtain Φ and Ψ as in (28) and (29) for known Σ_S and Σ_T , respectively.
- (ii) Generate $\Xi \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \mathbf{I}_n)$ by n samples independently distributed from the multivariate normal distribution $N_p(\mathbf{0}, \mathbf{I}_p)$.
- (iii) Obtain the simulated data matrix $\mathbf{Y} : p \times n$ by $\Phi \Xi \Psi'$.

Results for Case (i): $\rho = 0.1$

Table 3 presents the percentage of the spatial variance of each spatial pattern captured by the EOFs relative to the total spatial variance. It shows that these percentages are quite close to each other for all three EOFs the true, the classical and the corrected. Table 4 shows that the matrix discrepancy between the corrected and true EOFs closely resembles that for classical EOFs (with the classical doing slightly better).

Index of EOFs	True	Classical	Corrected
1	33.584	31.913	31.883
2	11.122	13.476	13.457
3	11.122	8.963	8.996
4	5.108	6.193	6.201
5	3.853	4.732	4.747
6	3.638	3.467	3.467
7	2.274	2.733	2.735
8	2.274	2.333	2.329
9	1.537	1.946	1.945
10	1.537	1.719	1.724

Table 3: Percentage of spatial variation (%) captured by the first ten EOFs by the true, classical and corrected methods when the autocorrelation is weak ($\rho = 0.1$).

Figures 7–9 plot the contours for the three type of EOFs. They show the classical EOFs are quite similar to the corrected ones when the autocorrelation is weak (i.e. ϕ is small).

Matrix discrepancies	
Classical v.s. True	0.230
Corrected v.s. True	0.233

Table 4: Matrix discrepancies between the classical and true EOFs as well as that discrepancy for the corrected EOFs when the autocorrelation is weak ($\rho = 0.1$).

These results validate our expectation that both EOFs would work well in this situation since the “true” data are approximately independent over time at all locations.

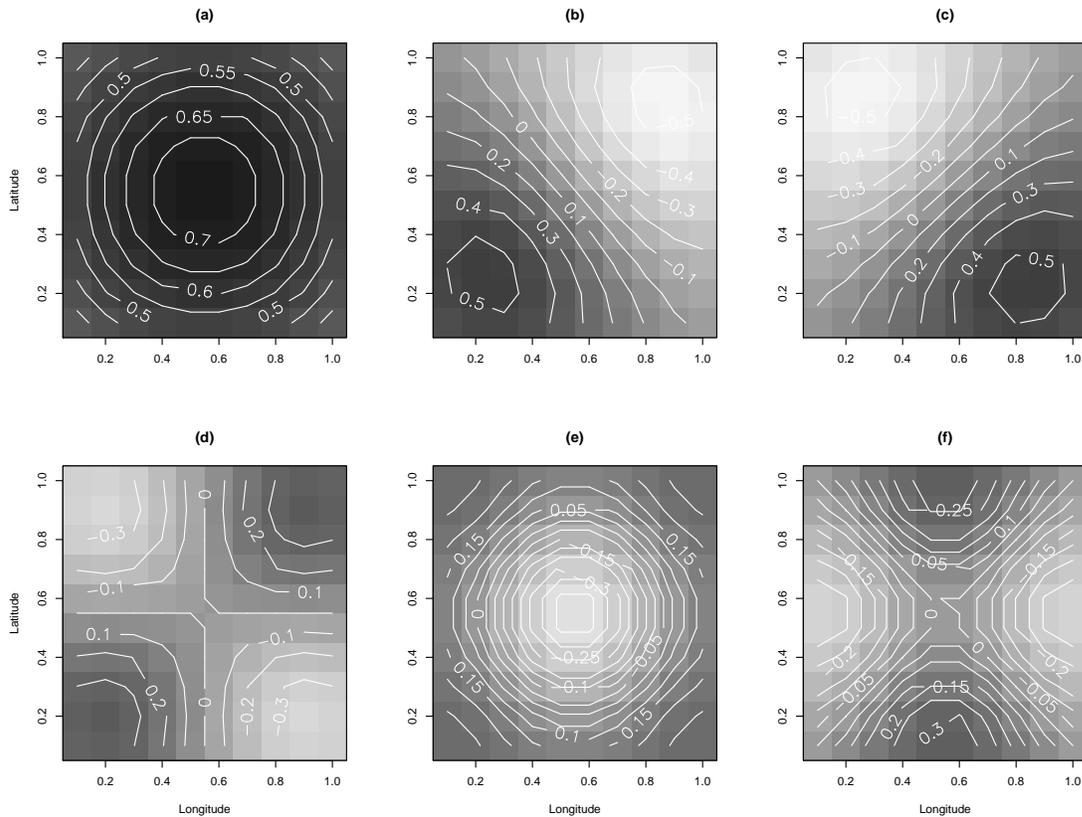


Figure 7: Contour plots for the first six true EOFs when the autocorrelation is weak ($\rho = 0.1$). (White=-0.6; Black=0.9)

Results for Case (ii): $\rho = 0.9$

Table 5 gives the percentages of the spatial variation captured by each of the three type EOFs: true, classical, and corrected. This graph shows that the corrected

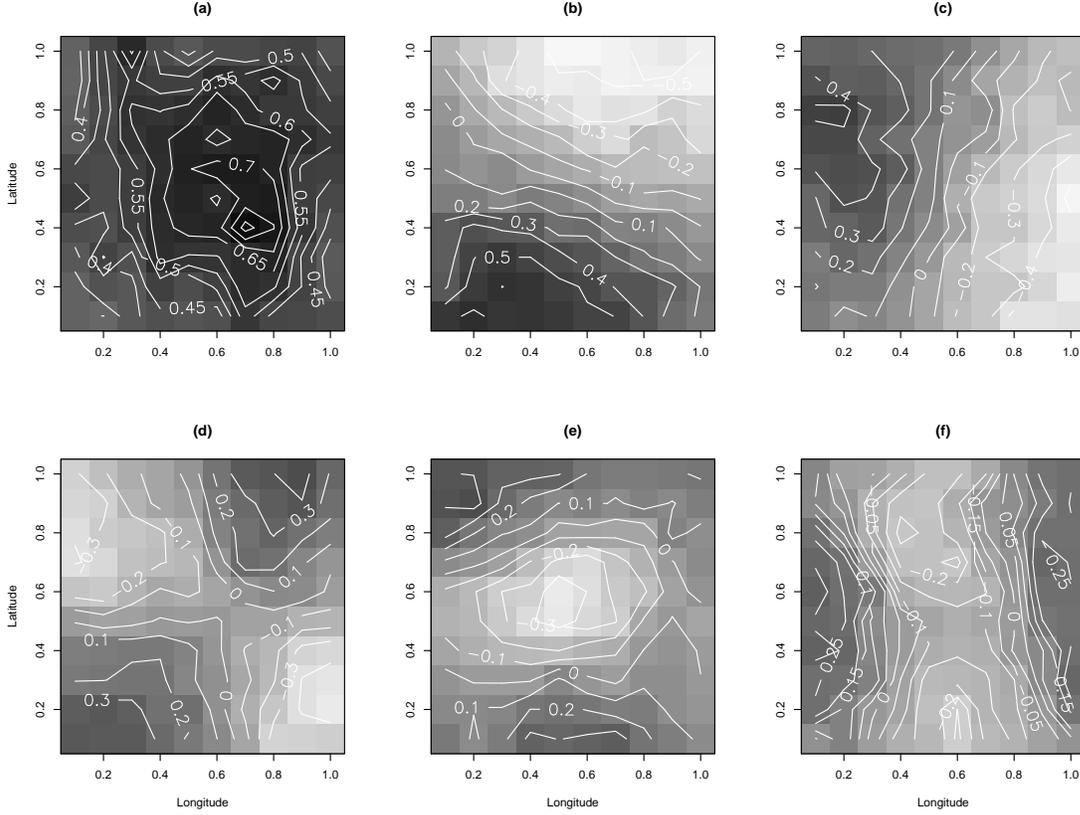


Figure 8: Contour plots for the first six classical EOFs when the autocorrelation is weak ($\rho = 0.1$). (White=-0.6; Black=0.9)

EOFs give more accurate estimates on the major diagonal elements of $\mathbf{\Lambda}$ in the KL expansion. Table 6 shows the matrix discrepancies between the corrected and the true EOFs to be much smaller than that discrepancy for classical EOFs, pointing to the former’s superiority in this case where that autocorrelation is strong.

Figures 10–11 present the first six classical and corrected EOFs for this case. Comparing them with Figure 7, it is obvious that the corrected EOFs can estimate the main types of spatial patterns better than the classical ones. Moreover, the ratio of matrix discrepancies between the classical and true EOFs matrix is 5.79, while that between the corrected and true ones, is just 0.21. It shows that the classical EOFs are far from the “truth” for the highly autocorrelated data base.

5 Summary and conclusions

We have developed Bayesian EOFs in this report and given the corresponding theoretical results as well as the MCMC algorithm to obtain the posterior samples of

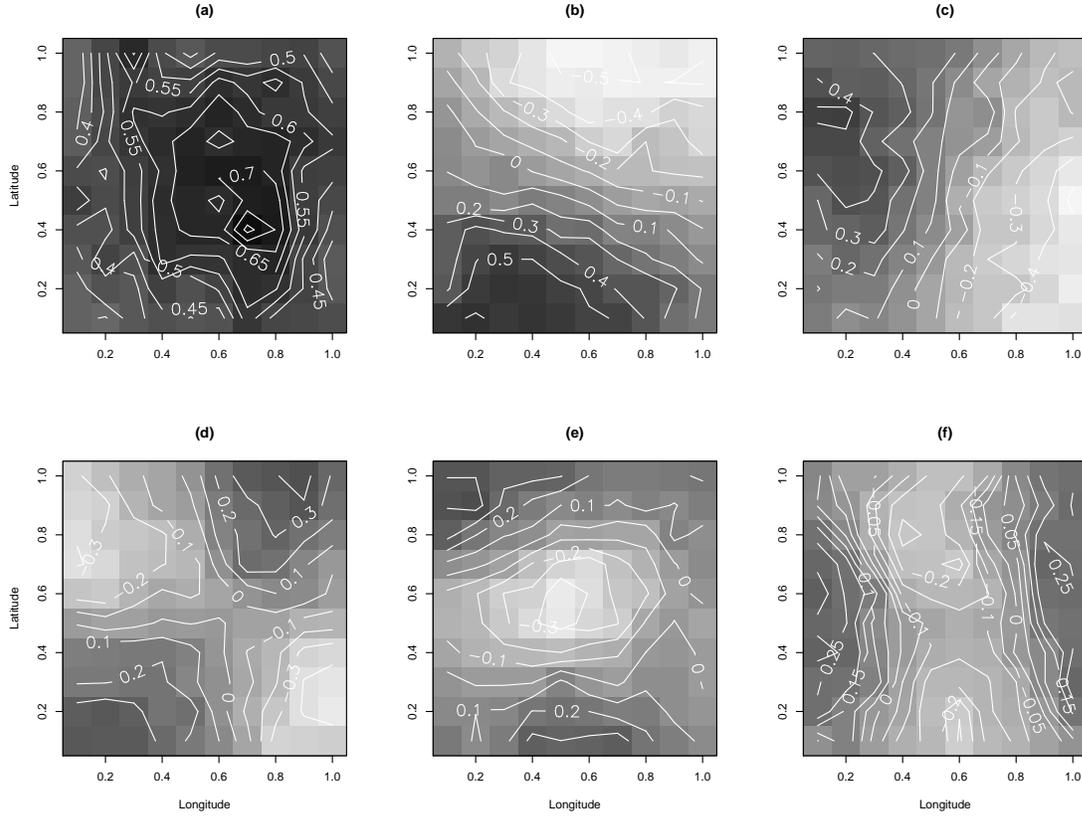


Figure 9: Contour plots for the first six corrected EOFs when the autocorrelation is weak ($\rho = 0.1$). (White=-0.6; Black=0.9)

the model parameters. Our two simulation studies have shown that the corrected EOF method can give much better representations of principal spatial patterns than classical EOFs. We conclude that classical EOFs can severely misrepresent spatial patterns in highly temporally correlated space-time processes. Corrected EOFs greatly improve the performance of the EOFs and capture the principal spatial patterns better than the classical EOFs. The BEOF offers a practical way of realizing those improvements in reality when the correction factors needed for the corrected BEOF are unknown. However, further improvements in the BEOF method are needed and will be the subject of future work, along with a more complete comparative assessment of the perform of the BEOFs against corrected and classical EOFs.

Index of EOFs	True	Classical	Corrected
1	33.584	33.572	37.702
2	11.122	20.789	11.868
3	11.122	9.660	11.005
4	5.108	7.835	6.204
5	3.853	4.341	4.344
6	3.638	3.302	3.405
7	2.274	2.870	2.038
8	2.274	2.287	1.910
9	1.537	1.746	1.521
10	1.537	1.583	1.460

Table 5: Percentage of spatial variation (%) for the first ten EOFs captured by the true, classical, and corrected methods when the autocorrelation is strong ($\rho = 0.9$).

Matrix discrepancies	
Classical vs True	4.978
Corrected vs True	0.332

Table 6: Matrix discrepancies between the traditional and true EOFs as well as that discrepancy for the corrected EOFs when the autocorrelation is strong ($\rho = 0.9$).

A Supplementary results

A.1 More results for Section 3

Definition 4 Suppose the random matrix response $\mathbf{X} : r \times q$ has a matrix normal distribution, denoted by $\mathbf{X} \sim N_{r \times q}(\mathbf{M}, \mathbf{C}, \mathbf{\Sigma})$, where $\mathbf{C} : r \times r > 0$, and $\mathbf{\Sigma} : q \times q > 0$. Then the probability density function of \mathbf{X} is given by

$$p(\mathbf{X}) = (2\pi)^{-rq/2} |\mathbf{C}|^{-q/2} |\mathbf{\Sigma}|^{-r/2} \exp\left\{-\frac{1}{2} \text{tr}[(\mathbf{X} - \mathbf{M})' \mathbf{C}^{-1} (\mathbf{X} - \mathbf{M}) \mathbf{\Sigma}^{-1}]\right\}. \quad (30)$$

Definition 5 Suppose the random matrix $\mathbf{X} : q \times q$ is symmetric, positive definite and follows an inverted Wishart distribution with degrees of freedom δ and scale matrix \mathbf{S} . Then the probability density function of \mathbf{X} is given by

$$p(\mathbf{X}) = k |\mathbf{X}|^{-(\frac{\delta}{2} + q)} \exp\left\{-\frac{1}{2} \text{tr}[\mathbf{S} \mathbf{X}^{-1}]\right\}, \quad (31)$$

where \mathbf{S} is positive definite and

$$k^{-1} = 2^{qv/2} \pi^{q(q-1)/4} \prod_{j=1}^q \Gamma\left(\frac{v-j-1}{2}\right) |\mathbf{S}|^{-v/2},$$

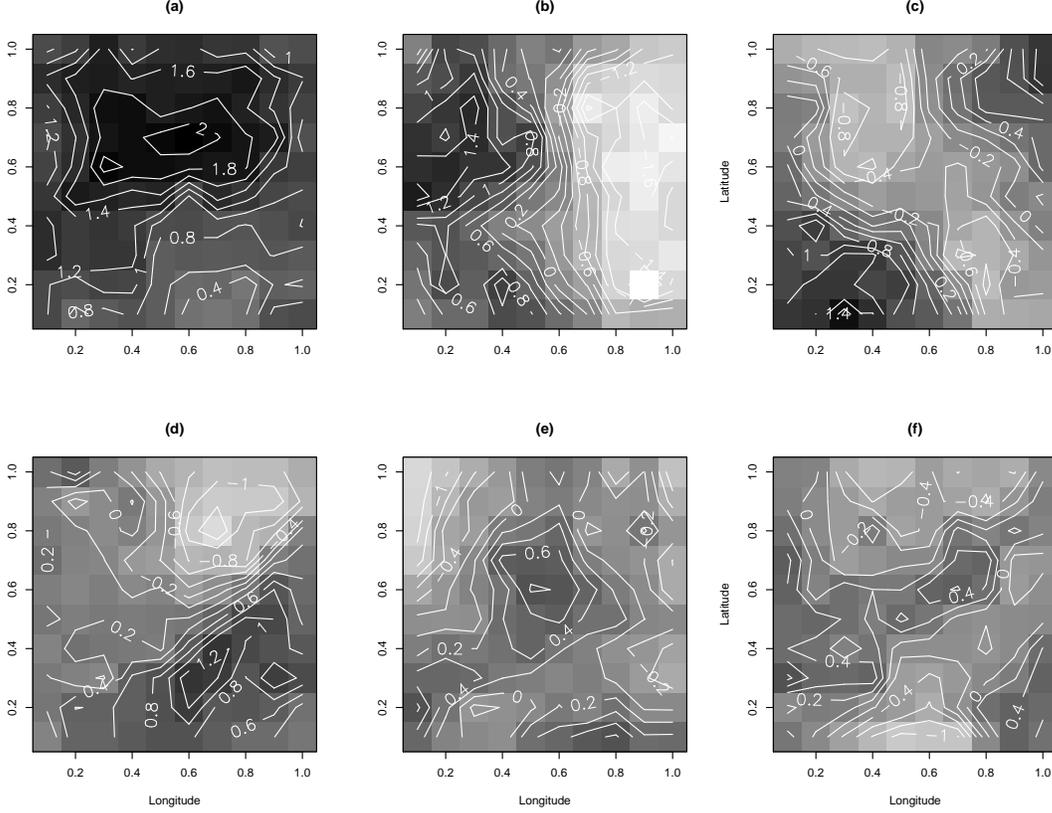


Figure 10: Contour plots for the first six classical EOFs when the autocorrelation is strong ($\rho = 0.9$). (White=-1.6; Black=2.2)

with $v = \delta + q - 1$.

Proof 1 (Lemma 3) By the KL expansion and Lemma 2, we have $\mathbf{\Lambda}^{-2} = \mathbf{O}\mathbf{\Sigma}_S^{-1}\mathbf{O}'$ and moreover, $\mathbf{\Lambda}^{-2} \sim W_p(n, \mathbf{I}_p)$. Hence, the $\{\lambda_j^{-2} : j = 1, \dots, p\}$ are mutually independent and $\lambda_j^{-2} \sim W_1(n, 1)$, that is, χ_n^2 , for $j = 1, \dots, p$.

Proof 2 (Theorem 1) Given $\mathbf{Y} : p \times n \sim N_{p \times n}(\mathbf{0}, \mathbf{\Sigma}_S \otimes \mathbf{\Sigma}_T)$, denote \mathbf{Y}^* to be $\mathbf{Y}\mathbf{\Sigma}_T^{-1/2}$. Consequently, $\mathbf{Y}^* \sim N_{p \times n}(\mathbf{0}, \mathbf{\Sigma}_S \otimes \mathbf{I}_n)$. Similarly, we have $\mathbf{\Sigma}_S^{-1/2}\mathbf{Y}^* \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \mathbf{I}_n)$.

By Lemma 2, $\mathbf{\Sigma}_S^{-1/2}\mathbf{Y}^* = \mathbf{O}\mathbf{L}\mathbf{P}$, where \mathbf{O} represents an orthogonal matrix that is uniformly distributed over the Grassmann manifold, \mathbf{P} , an orthogonal frame that is uniformly distributed over the Stiefel manifold, and \mathbf{L} , a diagonal matrix with entries $\{l_1, \dots, l_p\}$ such that l_1^2, \dots, l_p^2 are the eigenvalues for $(\mathbf{\Sigma}_S^{-1/2}\mathbf{Y}^*)(\mathbf{\Sigma}_S^{-1/2}\mathbf{Y}^*)'$. Hence, $\mathbf{Y}^* = \mathbf{\Sigma}_S^{-1/2}\mathbf{O}\mathbf{L}\mathbf{P}$. Moreover, since

$$\begin{aligned} E[(\mathbf{Y}^*)(\mathbf{Y}^*)'] &= \mathbf{\Sigma}_S^{-1/2}E[\mathbf{O}\mathbf{L}^2\mathbf{O}']\mathbf{\Sigma}_S^{-1/2} \\ &= n\mathbf{\Sigma}_S, \end{aligned}$$

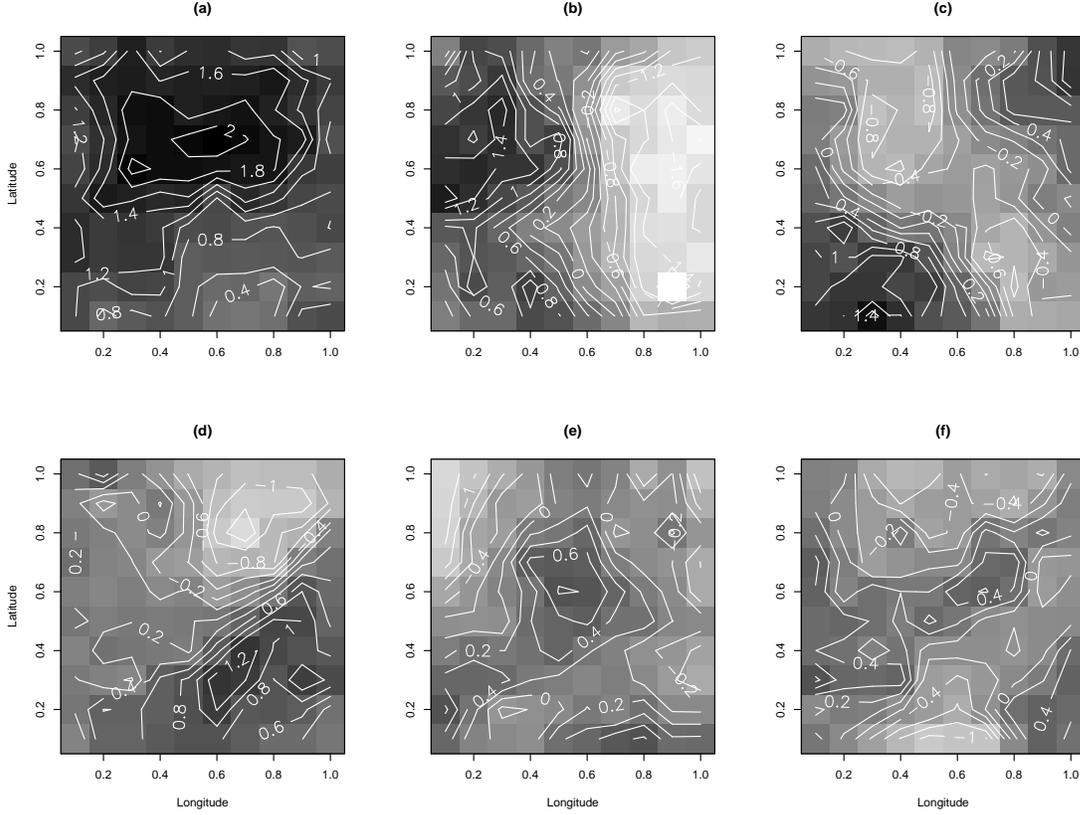


Figure 11: Contour plots for the first six corrected EOFs when the autocorrelation is strong ($\rho = 0.9$). (White=-1.6; Black=2.2)

the Bayesian EOFs can then be given by $\mathbf{W} = \frac{1}{n} \Sigma_S^{1/2} \mathbf{O} \mathbf{L}$.

Proof 3 (Theorem 2) By Definition 5, we have

$$p(\Sigma_S^{-1}) \propto |\Sigma_S|^{-\left(\frac{\delta_S}{2} + p\right)} \exp\left\{-\frac{1}{2} \text{tr}(\mathbf{\Xi}_S^{-1} \Sigma_S^{-1})\right\}.$$

Given Σ_T , $\mathbf{Y}^* = \mathbf{Y} \Sigma_T^{-1/2} \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \mathbf{I}_n)$. By Definition 4, we have

$$p(\mathbf{Y}^* | \Sigma_S) \propto |\Sigma_S|^{-n/2} \exp\left\{-\frac{1}{2} \text{tr}[\mathbf{Y}^* (\mathbf{Y}^*)' \Sigma_S^{-1}]\right\}.$$

Then the posterior distribution for Σ_S^{-1} given \mathbf{Y}^* , that is, \mathbf{Y} for known Σ_T is given as follows:

$$\begin{aligned} p(\Sigma_S^{-1} | \mathbf{Y}) &\propto p(\mathbf{Y}^* | \Sigma_S) p(\Sigma_S^{-1}) \\ &\propto |\Sigma_S|^{-\left(\frac{\delta_S + n}{2} + p\right)} \exp\left\{-\frac{1}{2} \text{tr}[(\mathbf{\Xi}_S^{-1} + \mathbf{Y}^* (\mathbf{Y}^*)') \Sigma_S^{-1}]\right\}. \end{aligned}$$

In other words, $\Sigma_S^{-1} | \mathbf{Y} \sim W_p(\delta^\circ, \Xi^\circ)$, where $\delta^\circ = \delta_S + n$ and

$$\begin{aligned}\Xi^\circ &= \{\Xi_S^{-1} + \mathbf{Y}^*(\mathbf{Y}^*)'\}^{-1} \\ &= (\Xi_S^{-1} + \mathbf{Y}\Sigma_T^{-1}\mathbf{Y}')^{-1} \\ &= \Xi_S - \Xi_S \mathbf{Y}(\mathbf{Y}'\Xi_S \mathbf{Y} + \Sigma_T)^{-1} \mathbf{Y}'\Xi_S.\end{aligned}$$

Proof 4 (Theorem 3) This proof resembles Proof 3, and so is omitted here.

Proof 5 (Theorem 4) Let $\mathbf{V} = \Sigma_S^{-1/2} \mathbf{Y}$. Then $\mathbf{V} \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \Sigma_T)$. Hence, we have

$$p(\mathbf{Y} | \Sigma_S, \mu, \theta) \propto \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{V}' \mathbf{V} \rho(\cdot, \theta)^{-1}] \right\}.$$

Given the prior for θ , $N_k(\theta, \Sigma_0)$, the posterior conditional density for θ can be represented by

$$\begin{aligned}p(\theta | \mathbf{Y}, \Sigma_S, \mu) &\propto p(\mathbf{Y} | \Sigma_S, \mu, \theta) p(\theta) \\ &\propto \exp \left\{ -\frac{1}{2} [\text{tr}(\mathbf{V}\mathbf{V}' \rho(\cdot, \theta)^{-1}) + (\theta - \theta_0)' \Sigma_0^{-1} (\theta - \theta_0)] \right\}.\end{aligned}$$

Proof 6 (Theorem 5)

(i) Since

$$\mathbf{Z} \sim N_{p \times n}(\mu \otimes \mathbf{1}'_n, \Sigma_S \otimes \Sigma_T),$$

and $p(\mu) \propto 1$, we have the posterior conditional distribution for μ as follows:

$$\begin{aligned}p(\mu | \mathbf{Z}, \Sigma_S, \Sigma_T) &\propto p(\mathbf{Z} | \mu, \Sigma_S, \Sigma_T) p(\mu) \\ &\propto \exp \left\{ -\frac{1}{2} \text{tr} [(\mu \otimes \mathbf{1}'_n - \mathbf{Z}) \Sigma_T^{-1} (\mu \otimes \mathbf{1}'_n - \mathbf{Z})' \Sigma_S^{-1}] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{tr} [(\mu \otimes \mathbf{1}'_n) \Sigma_T^{-1} (\mu \otimes \mathbf{1}'_n)' \Sigma_S^{-1} - (\mu \otimes \mathbf{1}'_n) \right. \\ &\quad \left. \times \Sigma_T^{-1} \mathbf{Z}' \Sigma_S^{-1} - \mathbf{Z} \Sigma_T^{-1} (\mu \otimes \mathbf{1}'_n)' \Sigma_S^{-1}] \right\} \\ &= \exp \left\{ -\frac{1}{2} \text{tr} [(\mu \mu' \text{tr}(\mathbf{1}'_n \mathbf{1}_n \Sigma_T^{-1}) - \mu \mathbf{1}_n \Sigma_T^{-1} \mathbf{Z}' \right. \\ &\quad \left. - \mathbf{Z} \Sigma_T^{-1} \mathbf{1}'_n \mu') \Sigma_S^{-1}] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{tr} [(\mu - \mathbf{M})(\Sigma^*)^{-1} (\mu - \mathbf{M})' \Sigma_S^{-1}] \right\},\end{aligned}$$

where $\Sigma^* = \{\text{tr}(\mathbf{1}'_n \mathbf{1}_n \Sigma_T^{-1})\}^{-1}$ and $\mathbf{M} = \mathbf{Z} \Sigma_T^{-1} \mathbf{1}'_n \Sigma^*$. Therefore, we have

$$\mu | \mathbf{Z}, \Sigma_S, \Sigma_T \sim N_{1 \times p}(\mathbf{M}, \Sigma^* \otimes \Sigma_S),$$

that is, $N_p(\mathbf{M}, \Sigma^* \Sigma_S)$ since Σ^* is a scalar.

(ii) Let $\mathbf{Y} = \mathbf{Z} - \mu \otimes \mathbf{1}'_n$. Consequently, we have $\mathbf{Y}\Sigma_T^{-1/2} \sim N_{p \times n}(\mathbf{0}, \Sigma_S \otimes \mathbf{I}_n)$. By Theorem 2, the posterior distribution for $\Sigma_S^{-1} | \mathbf{Z}, \Sigma_T$ is $W_p(\delta^o, \Xi^o)$, given by (16), where $\mathbf{Y} = \mathbf{Z} - \mu \otimes \mathbf{1}'_n$.

(iii) Similarly as in (ii), by Theorem 3 the posterior conditional distribution for $\Sigma_T^{-1} | \mathbf{Z}, \Sigma_S$ is given by (18), where $\mathbf{Y} = \mathbf{Z} - \mu \otimes \mathbf{1}'_n$.

Proof 7 (Theorem 6) The proof for this theorem follows Theorem 4 but letting $\mathbf{V} = \Sigma_S^{-1/2}(\mathbf{Z} - \mu \otimes \mathbf{1}'_n)$.

References

- [1] Aguilar, O., and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, 18, 338–357.
- [2] Björnsson, H. & Venegas, S.A. (1997). *A manual for EOF and SVD analyses of climatic data*. Centre for Climate and Global Change Research Report # 97-1, McGill U.
- [3] Brown, P.J., Le, N.D. and Zidek, J.V. (1994). Multivariate spatial interpolation and exposure to air pollutants, *Can. J. Statist.*, 22, 489–510.
- [4] Chikuse, Y. (2006). State space models on special manifolds. *Journal of Multivariate Analysis*, 97, 1284–1294.
- [5] Damian, D., Sampson, P.D., & Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, 12, 161–178.
- [6] Dou, Y.P., Le, N.D. and Zidek, J.V. (2007). A dynamic linear model for hourly ozone concentrations. *Technical Report #228*, Department of Statistics, UBC.
- [7] Dou, Y.P., Le, N.D. and Zidek, J.V. (2008). Modeling hourly ozone concentration fields. *Submitted*.
- [8] Dou, Y.P., Le, N.D. and Zidek, J.V. (2009a). Temporal prediction with a Bayesian spatial predictor: an application to ozone fields. *In preparation*.
- [9] Dou, Y.P., Le, N.D. and Zidek, J.V. (2009b). Generalized Bayesian spatial methodology. *In preparation*.
- [10] James A.T. (1954). Normal multivariate analysis and the orthogonal group.
- [11] James, A.T. (1960). The distribution of the latent roots of the covariance matrix. *The Annals of Mathematical Statistics*, 31, 151–158.
- [12] Hannachi, A., Jolliffe, I.T. and Stephenson, D.B. (2007). Empirical orthogonal functions and related techniques in atmospheric science, a review. *Int. J. Climatology*, 27, 1119–1152, DOI 10.1002/joc.1499.
- [13] Le, N.D., Sun, W. and Zidek, J.V. (1997). Bayesian multivariate spatial interpolation with data missing by design. *J. Roy. Stat. Soc., Ser B*, 59, 501–510.
- [14] Le, N.D. and Zidek, J.V. (1992). Interpolation with uncertain spatial covariance: a Bayesian alternative to kriging. *J. Mult. Anal.*, 43, 351–374.

- [15] Le, N.D., & Zidek, J.V. (2006). *Statistical analysis of environmental space-time processes*. Springer.
- [16] Mardia, K.V. (1977). Distributions on Stiefel and Grassman manifolds, and their applications. *Advances in Applied Probability*, 9, 435–436.
- [17] Wikle, C.K. (2002). Spatio-temporal methods in climatology. *To appear: Encyclopedia of Life Support Systems*. EOLSS Publishers Co. Ltd.
- [18] Zidek, J.V., Sun, L., Le, N.D., and Özkaynak, H. (2002). Contending with space-time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM₁₀ field. *Environmetrics*, 13, 1–19.