

Beyond axial symmetry: An improved class of models for global data

Stefano Castruccio^{a*} and Marc G. Genton^a

Received 3 February 2014; Accepted 14 February 2014

An important class of models for data on a spherical domain, called axially symmetric, assumes stationarity across longitudes but not across latitudes. The main aim of this work is to introduce a new and more flexible class of models by relaxing the assumption of longitudinal stationarity in the context of regularly gridded climate model output. In this investigation, two other related topics are discussed: the lack of fit of an axially symmetric parametric model compared with a non-parametric model and to longitudinally reversible processes, an important subclass of axially symmetric models. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: axial symmetry; longitudinal reversibility; massive dataset; non-stationarity; sphere

1 Introduction

In comparison with theoretical and practical developments of data on the plane, data on a spherical domain have received considerably less attention from the space–time statistics community. Recently, Gneiting (2013) provided an overview of theoretical challenges for some classes of data in this domain, with a list of open problems that still need to be addressed. On the applied side, the most common assumption for data on spherical domain is *axial symmetry* (Jones, 1963); that is, data are stationary across longitudes. To generate positive definite covariance functions with axial symmetry, Jun & Stein (2007, 2008) proposed to define an isotropic process in \mathbb{R}^3 , restrict it over the sphere, and apply derivatives to obtain non-stationarity across latitudes. This approach has been applied to climate model output (Jun et al., 2008) and has been extended to the multivariate setting (Jun, 2011). More recently, Castruccio & Stein (2013a) proposed a spectral-based model in the special case of gridded data.

In this work, we aim to address some fundamental issues in the use of axially symmetric processes in the gridded setting. In particular, we demonstrate that in the case of temperature simulated by a climate model, the fit can be greatly improved if a broader class of models is considered. Moreover, we show that the likelihood for this class can be obtained in a computationally convenient form, and a dataset of 25 million points can be fit. We also show how the parametric assumption of an axially symmetric model implies a small loss of information compared with a fully non-parametric model, and how restricting the analysis to a subclass of models with symmetry with respect to longitude (longitudinal reversibility; Stein (2007)) is an adequate practice with temperature data.

The remainder of this paper proceeds as follows: Section 2 describes the dataset; Section 3 reviews some previous results that will be used in the model fit; Section 4 introduces the steps of the model fit and, for the spatial part, it discusses and compares different classes of models. Section 5 concludes with a discussion.

^aCEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

*Email: stefano.castruccio@kaust.edu.sa

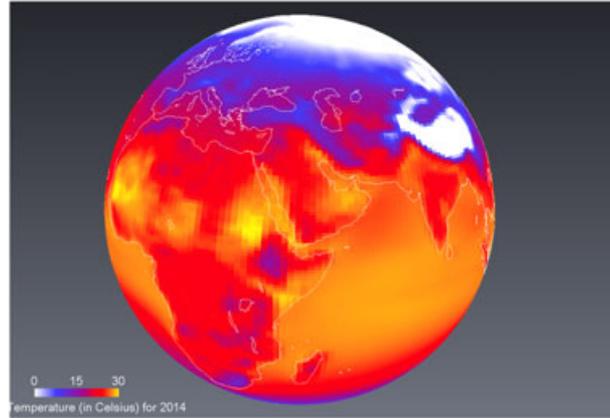


Figure 1. Annual mean temperature (in Celsius) for the first run for the year 2014.

2 The dataset

We focus on the Coupled Model Intercomparison Project Phase 5 (Taylor et al., 2012), the reference multi-model ensemble for the fifth Intergovernmental Panel on Climate Change Assessment Report (Cubasch et al., 2013). In particular, we focus on yearly temperature at the Earth's surface (at a reference height of 2 m above ground level) of the National Center for Atmospheric Research Community Climate System Model 4 (NCAR CCSM4; Gent et al. (2011)). The output is considered for the Representative Concentration Pathway 85 scenario (Van Vuuren et al., 2011). The data are on a regular latitude-by-longitude grid with 192 latitudinal points and 288 longitudinal points. The scenario comprises 95 years, between 2006 and 2100, and the ensemble consists of six runs with different initial conditions. As the statistical behaviour in the Antarctic region is very different from the rest of the data and as latitude bands near the poles are physically very close and this leads to numerical instabilities, we remove the data south of -62° latitude and north of 82° latitude. The dataset thus consists of 155 latitudinal bands, for a total of $155 \times 288 \times 95 \times 6 \approx 25$ million data points. In Figure 1, we see an example of a temperature field for the first run for the year 2014.

3 The setting

Our model assumes that, conditionally on the knowledge of the mean, the output variability can be considered independent across initial conditions, and we therefore refer to the runs as realizations. This assumption relies on the deterministically chaotic nature of the primitive equations in climate models (Lorenz, 1963), and its validity has been discussed in several works (Collins & Allen, 2002; Collins, 2002; Branstator & Teng, 2010). Denote by \mathbf{T}_r the vector of temperatures for realization r , by $\boldsymbol{\mu}$ its expected value across realizations, by ℓ_n the longitude, by L_m the latitude, and by t_k the time, where $n = 1, \dots, N$, $m = 1, \dots, M$ and $k = 1, \dots, K$. We assume that

$$\mathbf{T}_r = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_r, \quad \boldsymbol{\varepsilon}_r \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \Sigma), \quad (1)$$

for $r = 1, \dots, R$, and

$$\mathbf{T}_r = \{\mathbf{T}_r(L_1, \ell_1, t_1), \dots, \mathbf{T}_r(L_M, \ell_1, t_1), \mathbf{T}_r(L_1, \ell_2, t_1), \dots, \mathbf{T}_r(L_M, \ell_2, t_1), \dots, \mathbf{T}_r(L_M, \ell_N, t_k)\}^T.$$

Under this setting, if $r > 1$, it is possible to estimate the space–time structure Σ without estimating $\boldsymbol{\mu}$ since $\mathbf{T}_r - \mathbf{T}_{r'} \sim \mathcal{N}(\mathbf{0}, 2\Sigma)$ if $r \neq r'$. More formally, suppose that $\Sigma = \Sigma(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of parameters of the covariance structure. If we define $\mathbf{D}_r = \mathbf{T}_r - \frac{1}{R} \sum_{r'=1}^R \mathbf{T}_{r'}$, then we have

Result

Let $\mathbf{D} = (\mathbf{D}_1^\top, \dots, \mathbf{D}_R^\top)^\top$. The restricted loglikelihood for (1) is

$$l(\boldsymbol{\theta}; \mathbf{D}) = -\frac{1}{2}KMN(R-1) \log(2\pi) - \frac{1}{2}(R-1) \log |\Sigma(\boldsymbol{\theta})| - \frac{1}{2}KMN \log(R) - \frac{1}{2} \sum_{r'=1}^R \mathbf{D}_{r'}^\top \Sigma(\boldsymbol{\theta})^{-1} \mathbf{D}_{r'}. \quad (2)$$

Also, the corresponding estimator for $\boldsymbol{\mu}$, obtained by generalized least squares, is $\hat{\boldsymbol{\mu}} = \frac{1}{R} \sum_{r'=1}^R \mathbf{T}_{r'}$.

The proof can be found in the work of Castruccio & Stein (2013b). In this work, we obtain the parameters for all the models we present in the next sections by maximizing (2).

4 The model

In this section, we introduce the three-step model used to fit the data. Each step estimates the parameters along one dimension conditionally on the previous ones. While discussing the spatial part of the model, we relax the assumption of stationarity across longitudes and show the substantial improvement with respect to an axially symmetric model. We also discuss the appropriateness of parametric models and a subclass of axially symmetric processes in terms of lack of fit compared with a fully non-parametric model.

4.1. Temporal part

We assume that the model has an autoregressive AR(2) structure, with different parameters for each grid point. A diagnostic analysis similar to that of Castruccio & Stein (2013a) has shown that this is a reasonable assumption. Denote by $\boldsymbol{\varepsilon}(t; r)$ the component for time t and realization r . Then,

$$\begin{aligned} \boldsymbol{\varepsilon}(t; r) &= \Phi_1 \boldsymbol{\varepsilon}(t-1; r) + \Phi_2 \boldsymbol{\varepsilon}(t-2; r) + \boldsymbol{\eta}(t; r), \\ \boldsymbol{\eta}(t; r) &\stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{SCS}), \end{aligned}$$

where Φ_1 and Φ_2 are $NM \times NM$ diagonal matrices with the autoregressive coefficients, \mathbf{S} is a $NM \times NM$ diagonal matrix with the standard deviations for each grid point, and \mathbf{C} is the correlation matrix. We use (2) to estimate the components for Φ_1 , Φ_2 , and \mathbf{S} separately for every n and m . Each fit is performed for the six independent time series of length 95 at the chosen locations. As our workstation has multiple processors, we perform the maximization of (2) in parallel, a procedure that greatly reduces the total computational time. Throughout this work, we define $\mathbf{H}(t; r) = \{\boldsymbol{\varepsilon}(t; r) - \Phi_1 \boldsymbol{\varepsilon}(t-1; r) - \Phi_2 \boldsymbol{\varepsilon}(t-2; r)\} \mathbf{S}^{-1}$ and $\hat{\mathbf{H}}(t; r) = \{\mathbf{D}(t; r) - \hat{\Phi}_1 \mathbf{D}(t-1; r) - \hat{\Phi}_2 \mathbf{D}(t-2; r)\} \hat{\mathbf{S}}^{-1}$. The next two sections describe several models for \mathbf{C} , which are estimated using $\hat{\mathbf{H}}(t; r)$. As the distributions of $\mathbf{H}(t; r)$ and $\hat{\mathbf{H}}(t; r)$ do not depend on time or realization, these two indexes will be dropped for simplicity.

4.2. Latitudinal bands: stationarity and beyond

We first consider a model for $\mathbf{H}_m = \{\mathbf{H}(L_m, \ell_1), \dots, \mathbf{H}(L_m, \ell_N)\}^\top$, the vector of the components of \mathbf{H} across a single band. A convenient assumption for the process is stationarity across longitude (*axial symmetry*; Jones (1963)). In this setting, with points equally spaced on a circle, the matrix is exactly circulant (Davis, 1979), and it is more natural

to model the process in the spectral domain, as the wavenumbers are uncorrelated. We denote by $\tilde{\mathbf{H}}_m$ the band-wise Fourier transformed vector and by $f_m(c) = \text{var}\{\tilde{\mathbf{H}}_m(c)\}$ the spectral density at wavenumber $c = 0, \dots, N - 1$.

In past work (Castruccio & Stein, 2013a), a parametric model for f_m was proposed. This model is similar to the Matérn family but accounts for the circular geometry and avoids loss of regularity at high wavenumbers:

$$f_m(c; \phi_m, \alpha_m, \nu_m) = \frac{\phi_m}{\{\alpha_m^2 + 4 \sin^2(\frac{c}{N}\pi)\}^{\nu_m+1/2}}, \quad c = 0, \dots, N - 1. \tag{3}$$

In this section, we seek to understand

- if (3) is a reasonable approximation of the periodogram;
- if the stationarity assumption is valid.

For the first task, as independent replicates of $\hat{\mathbf{H}}_m$ occur across realizations and time, the periodogram can be estimated as the sample variance of its band-wise Fourier transform (denoted by $\tilde{\mathbf{H}}_m$), across these two dimensions at each wavenumber c . In Figure 2, we compare the restricted loglikelihoods (2) between (3) and the model with the empirical periodogram, for all bands (in blue and red, respectively). It is apparent from this plot that the choice of a parametric periodogram implies no loss of information in the model for all latitudes other than at 50°S, where a small drop in likelihood appears. This comparison confirms the diagnostic in Castruccio & Stein (2013a) who argued that a parametric model exhibits good fit by visually comparing the fitted periodogram with the empirical estimate.

For the second task, instead of just considering the variance, we consider the full sample covariance matrix of $\tilde{\mathbf{H}}_m$ averaged over time and realizations. As the stationarity assumption is equivalent to independence across wavenumbers in the spectral domain, if this assumption is reasonable, then the covariance matrix should be approximately diagonal, and the corresponding likelihood should not show a noticeable change. In Figure 2, the restricted loglikelihood (2) is shown (in black) in relation to the two previously introduced models. The stationarity assumption is remarkably not adequate, as the difference in likelihood is overwhelmingly large compared with that of non-parametric versus

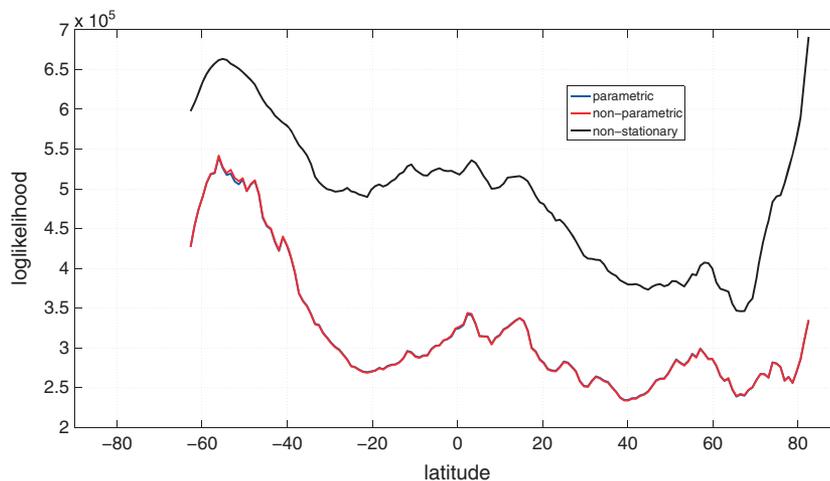


Figure 2. Loglikelihood comparison of the parametric, non-parametric, and non-stationary models. The blue and red curves are almost superimposed.

parametric stationary models. This result stresses the importance of defining models for global temperature data that are able to relax this assumption.

4.3. Multiple bands: parametric and non-parametric forms and longitudinal reversibility

Once the model for single bands is defined, the relationship among bands must be specified. If the process is axially symmetric, then Jun & Stein (2008) proved the following:

$$\text{cov} \left\{ \tilde{\mathbf{H}}_m(c), \tilde{\mathbf{H}}_{m'}(c') \right\} = 0, \quad \text{for all } c \neq c', \tag{4}$$

for all $m, m' = 1, \dots, M$. This allows for a block diagonal structure in the spectral domain since only the coherence across Fourier transformed bands at the same wavenumber needs to be defined. When $c = c'$ in (4), Castruccio & Stein (2013a) assumed that

$$\begin{aligned} \left| \text{corr} \left(\tilde{\mathbf{H}}_m, \tilde{\mathbf{H}}_{m'} \right) (c; \xi, \tau) \right| &= \left[\frac{\xi}{\{1 + 4 \sin^2(\frac{\xi}{N} \pi)\}^\tau} \right]^{|m-m'|}, \quad c = 0, \dots, N-1, \\ \arg \left\{ \text{corr} \left(\tilde{\mathbf{H}}_m, \tilde{\mathbf{H}}_{m'} \right) (c) \right\} &= 0, \end{aligned} \tag{5}$$

where $\arg\{z\}$ is the argument of the complex number z . The second equation is equivalent to the process being *longitudinally reversible* (see Stein (2007) and North et al. (2011) for a similar class of models emerging from energy balance models considerations); that is, $\text{cov} \{ \mathbf{H}(L_m, \ell_n), \mathbf{H}(L_{m'}, \ell_{n'}) \} = \mathbf{g}(L_m, L_{m'}, |\ell_n - \ell_{n'}|)$.

In this section, we seek to understand

- if (5) is a reasonable approximation of the cross-periodogram;
- if the assumption of longitudinal reversibility is appropriate.

For the first point, we compare the model with the optimal parameters in (5) with that obtained from a block diagonal covariance matrix, each block corresponding to a wavenumber and consisting of the *real* part of $\text{corr} \left(\tilde{\mathbf{H}}_m, \tilde{\mathbf{H}}_{m'} \right) (c)$, averaged over time and realizations; in other words, we compare the best parametric and non-parametric models in the class of axially symmetric, longitudinally reversible processes. Table I shows the comparison between these models in terms of number of parameters, restricted loglikelihood difference (2) normalized by the number of points, and Akaike information criterion (AIC; Akaike (1974)). The non-parametric model shows an improvement in the fit, but at the expense of a greatly increased number of parameters. In fact, while the parametric model requires only $3M + 2 = 467$ parameters, the non-parametric model requires $M(M + 1)/2 \approx 12,000$ parameters of the empirical cross-periodogram for each wavenumber. Despite this increase, the AIC shows that the non-parametric model is preferable.

For the second point, we consider a block diagonal covariance matrix, each block corresponding to a wavenumber and consisting of the *complex* $\text{corr} \left(\tilde{\mathbf{H}}_m, \tilde{\mathbf{H}}_{m'} \right) (c)$, averaged over time and realizations. From Table I, we see that this results in a small improvement (0.21 loglikelihood per point with respect to the parametric model, 0.06 with respect to the non-parametric model) but requires twice the number of parameters than in the longitudinally reversible case. This is because the cross-periodogram has an imaginary component, which doubles the number of parameters. This model shows a smaller AIC than that of the previous two models, thus confirming how the increase of parameters in the non-longitudinally reversible model is not balanced by a noticeable increase in flexibility. As non-longitudinally reversible models do not significantly improve the fit, we do not attempt to define a parametric model for this class.

Table 1. Comparison among different models in terms of number of parameters (without temporal parameters) restricted loglikelihood (2) and Akaike information criterion (AIC). In the first row, ax/nax = axially/non-axially symmetric, lr/nlr = longitudinally/non-longitudinally reversible, pa/npa = parametric/non-parametric model. The best two models, both non-axially symmetric, are in bold.

	Model				
	ax/lr/pa	ax/lr/npa	ax/nlr/npa	nax/pa	nax/npa
# parameters	467	1.7×10^6	3.5×10^6	12.9×10^6	14.6×10^6
$\Delta \log \text{likelihood} / \{NMT(R-1)\}$	0	0.15	0.21	1.11	1.24
AIC	-1.120×10^8	-1.169×10^8	-1.106×10^8	-1.322×10^8	-1.347×10^8

Finally, we compare the aforementioned models with a non-axially symmetric model with correlated wavenumbers at the same latitude as described in Section 4.2, but such that (4) would still hold for all $m \neq m'$. We define this model both with the parametric form (5) and with the matrix of the *real* corr $(\tilde{\mathbf{H}}_m, \tilde{\mathbf{H}}_{m'})$ (c). A computationally convenient evaluation of (2) is not straightforward since (4) must also hold for $m = m'$ to have a block diagonal covariance matrix. Nevertheless, it is possible to retrieve a simple expression of (2) if $\tilde{\mathbf{H}}_m$ is first decorrelated band-wise and the terms are rearranged subsequently. The expression can be found in the Appendix. The last two columns of Table 1 show the results for these fits. The number of parameters is considerably larger than those of the previous models, but the extent of the improvement is far larger than that in the previous steps. The AIC confirms a large improvement when considering a parametric model from the non-axially symmetric class and a further, smaller improvement for the non-parametric form. This result reinforces the findings in Section 4.2 and shows that relaxing the stationarity assumption across longitudes allows a quantum leap in model flexibility.

5 Conclusion

In this work, we investigated the validity of several assumptions for spatial processes on the spherical domain. The assumption of stationarity across longitude has been shown to be remarkably inadequate, at least for temperature data. We proposed a non-stationary model that decorrelates the spectral process for each band, and we derived an analytic expression of the likelihood that can be evaluated without noticeably increasing the computational complexity of the fitting algorithm. Despite the significantly increased number of parameters, the gain in fit is such that this model is preferred to the axially symmetric class according to the AIC criterion. To reduce the large number of parameters, more effort is needed to understand which subclasses of non-axially symmetric models are more appropriate and the factors driving the non-stationarity. We also showed how the non-parametric model for a single latitudinal band does not result in a noticeable improvement in the fit, thus confirming how a parametric, Matérn-like model is adequate within the subset of axially symmetric models. The non-parametric model for multiple bands instead results in a small improvement in the AIC. Finally, this work has shown how restriction to longitudinally reversible processes is an adequate practice: non-longitudinally reversible processes result in a significant increase in model parameters but do not improve significantly the fit.

Appendix: Likelihood expression for the non-axially symmetric model

Recall that

$$\mathbf{H} = \{\mathbf{H}(L_1, \ell_1), \mathbf{H}(L_2, \ell_1), \dots, \mathbf{H}(L_M, \ell_1), \mathbf{H}(L_1, \ell_2), \dots, \mathbf{H}(L_M, \ell_N)\}^T.$$

Let \mathbf{P} be the permutation matrix such that

$$\mathbf{PH} = \{\mathbf{H}(L_1, \ell_1), \mathbf{H}(L_1, \ell_2), \dots, \mathbf{H}(L_1, \ell_N), \mathbf{H}(L_2, \ell_1), \dots, \mathbf{H}(L_M, \ell_N)\}^T.$$

Moreover, let \mathbf{F} be the matrix of the band-wise Fourier transform of \mathbf{H} , \mathbf{A} be the block diagonal matrix of \mathbf{FPH} , and $\mathbf{A}^{1/2}$ be its Cholesky decomposition. From (5), we know that

$$\mathbf{P}^T \mathbf{A}^{-1/2} \mathbf{FPH} \sim \mathcal{N}(\mathbf{0}, \mathbf{B}), \tag{6}$$

where \mathbf{B} is a block diagonal matrix with N blocks, each one being $M \times M$. From (6), we have that

$$\mathbf{H} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{P}^T \mathbf{F}^T \mathbf{A}^{1/2} \mathbf{P} \mathbf{B} \mathbf{P}^T \left(\mathbf{A}^{1/2}\right)^T \mathbf{F} \mathbf{P}\right).$$

As $|\mathbf{P}| = |\mathbf{F}| = 1$, we have

$$|\text{cov}(\mathbf{H})| = \left| \mathbf{P}^T \mathbf{F}^T \mathbf{A}^{1/2} \mathbf{P} \mathbf{B} \mathbf{P}^T \left(\mathbf{A}^{1/2}\right)^T \mathbf{F} \mathbf{P} \right| = \left| \mathbf{A}^{1/2} \mathbf{B} \left(\mathbf{A}^{1/2}\right)^T \right| = |\mathbf{B}| |\mathbf{A}|.$$

The negative loglikelihood for a single time and realization can therefore be computed as

$$\tilde{l}(\mathbf{H}) = \frac{1}{2} N M \log(2\pi) + \frac{1}{2} \log |\mathbf{B}| + \frac{1}{2} \log |\mathbf{A}| + \frac{1}{2} \left(\mathbf{P}^T \mathbf{A}^{-1/2} \mathbf{FPH} \right)^T \mathbf{B}^{-1} \mathbf{P}^T \mathbf{A}^{-1/2} \mathbf{FPH}.$$

By merging the last expression with (1), we have that the restricted negative loglikelihood is

$$\begin{aligned} l(\mathbf{D}) &= \frac{1}{2} K N M (R - 1) \log(2\pi) + \frac{1}{2} K N M \log(R) + \frac{1}{2} (R - 1) \log |\mathbf{B}| \\ &\quad + \frac{1}{2} (R - 1) \log |\hat{\mathbf{A}}| + \frac{1}{2} \sum_{r=1}^R \sum_{k=1}^K \mathbf{v}(t_k; r)^T \mathbf{B}^{-1} \mathbf{v}(t_k; r), \end{aligned} \tag{7}$$

where

$$\mathbf{v}(t_k; r) = \begin{cases} \mathbf{P}^T \hat{\mathbf{A}}^{-1/2} \mathbf{F} \mathbf{P} \mathbf{D}(t_1; r) \hat{\mathbf{S}}^{-1}, & k = 1, \\ \mathbf{P}^T \hat{\mathbf{A}}^{-1/2} \mathbf{F} \mathbf{P} \left\{ \mathbf{D}(t_2; r) - \hat{\Phi}_1 \mathbf{D}(t_1; r) \right\} \hat{\mathbf{S}}^{-1}, & k = 2, \\ \mathbf{P}^T \hat{\mathbf{A}}^{-1/2} \mathbf{F} \mathbf{P} \left\{ \mathbf{D}(t_k; r) - \hat{\Phi}_1 \mathbf{D}(t_{k-1}; r) - \hat{\Phi}_2 \mathbf{D}(t_{k-2}; r) \right\} \hat{\mathbf{S}}^{-1}, & k > 2. \end{cases}$$

The block diagonal matrix $\hat{\mathbf{A}}$ is obtained in Section 4.2 by pooling all the empirical covariance estimates for $\tilde{\mathbf{H}}_m$ for different m . The block diagonal matrix \mathbf{B} can be either parametrized by (5) and the optimal parameters obtained by maximizing (7) or can be estimated via the empirical covariance function of $\mathbf{v}(t_k; r)$ at each wavenumber, averaged across realizations and time.

Acknowledgements

We would like to thank Raphaël Huser for comments that have greatly improved the manuscript. We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank NCAR for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

References

- Akaike, H (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control*, **19**(6), 716–723.
- Branstator, G & Teng, H (2010), 'Two limits of initial-value decadal predictability in a CGCM', *Journal of Climate*, **23**, 6292–6311.
- Castruccio, S & Stein, ML (2013a), 'Global space–time models for climate ensembles', *Annals of Applied Statistics*, **7**(3), 1593–1611.
- Castruccio, S & Stein, ML (2013b), 'Supplement to "Global space–time models for climate ensembles"'. doi: 10.1214/13-AOAS656SUPP.
- Collins, M (2002), 'Climate predictability on interannual to decadal time scales: the initial value problem', *Climate Dynamics*, **19**, 671–692.
- Collins, M & Allen, MR (2002), 'Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability', *Journal of Climate*, **15**, 3104–3109.
- Cubasch, UD, Wuebbles, DC, Facchini, MC, Frame, D, Mahowald, N & Winther, J-G (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* in Stocker, TF, Qin, D, Plattner, G-K, Tignor, M, Allen, SK, Boschung, J, Nauels, A, Xia, Y, Bex, V & Midgley, PM (eds), *Cambridge University Press*, Cambridge, United Kingdom and New York, NY, USA.
- Davis, PJ (1979), *Circulant Matrices*, Wiley, New York.
- Gent, PR, Danabasoglu, G, Donner, LJ, Holland, MM, Hunke, EC, Jayne, SR, Lawrence, DM, Neale, RB, Rasch, PJ, Vertenstein, M, Worley, PH, Yang, Z-L & Zhang, M (2011), 'The community climate system model version 4', *Journal of Climate*, **24**, 4973–4991.
- Gneiting, T (2013), 'Strictly and non-strictly positive definite functions on spheres', *Bernoulli*, **19**, 1327–1349.
- Jones, RH (1963), 'Stochastic processes on a sphere', *The Annals of Mathematical Statistics*, **34**, 213–218.
- Jun, M (2011), 'Nonstationary cross-covariance models for multivariate processes on a globe', *Scandinavian Journal of Statistics*, **38**, 726–747.
- Jun, M, Knutti, R & Nychka, DW (2008), 'Spatial analysis to quantify numerical model bias and dependence: how many climate models are there?' *Journal of the American Statistical Association*, **103**(483), 934–947.
- Jun, M & Stein, ML (2007), 'An approach to producing space–time covariance functions on spheres', *Technometrics*, **49**, 468–479.
- Jun, M & Stein, ML (2008), 'Nonstationary covariance models for global data', *Annals of Applied Statistics*, **2**(4), 1271–1289.
- Lorenz, EN (1963), 'Deterministic nonperiodic flow', *Journal of the Atmospheric Sciences*, **20**, 130–141.
- North, GR, Wang, J & Genton, MG (2011), 'Correlation models for temperature fields', *Journal of Climate*, **24**, 5850–5862.
- Stein, ML (2007), 'Spatial variation of total column ozone on a global scale', *Annals of Applied Statistics*, **1**, 191–210.
- Taylor, KE, Stouffer, RJ & Meehl, GA (2012), 'An overview of CMIP5 and the experiment design', *Bulletin of the American Meteorological Society*, **93**, 485–498.
- Van Vuuren, DP, Edmonds, J, Kainuma, M, Riahi, K, Thomson, A, Hibbard, K, Hurtt, GC, Kram, T, Krey, V, Lamarque, J-F, Masui, T, Meinshausen, M, Nakicenovic, N, Smith, SJ & Rose, SK (2011), 'The representative concentration pathways: an overview', *Climatic Change*, **109**, 5–31.