Spatio-temporal modelling of remote-sensing lake surface water temperature data

Mengyi Gong[1], Claire Miller[1], Marian Scott[1]

[1] University of Glasgow, Scotland, U.K.

E-mail for correspondence: mengyi.gong@glasgow.ac.uk

Abstract: Remote-sensing technology is widely used in environmental monitoring. The coverage and resolution of satellite based data provide scientists with great opportunities to study and understand environmental change. However, the large volume and the missing observations in the remote-sensing data present challenges to statistical analysis. This paper investigates two approaches to the spatio-temporal modelling of remote-sensing lake surface water temperature data. Both methods use the state space framework, but with different parameterizations to reflect different aspects of the problem. The appropriateness of the methods for identifying spatial/temporal patterns in the data is discussed.

Keywords: state space; dimension reduction; FPC; STRE.

1 Introduction

The remote-sensing lake surface water temperature (LSWT) data are measured by the Advanced Along-Track Scanning Radiometer (AATSR), on board the European Space Agency's Envisat platform. The retrieved LSWT data can be accessed from the ARC-Lake v3.0 database (http://www.ed.ac.uk/arclake/data.html). The observation period is from June 1995 to April 2012 and the spatial resolution is $0.05° \times 0.05°$. Ecologists are interested in the spatial/temporal patterns in the data to better understand the dynamics of the environmental system, for example, as part of the GloboLakes project (www.globolakes.ac.uk).
The specific data set investigated here is the monthly LSWT of Lake Victoria. The data are stored in an array of dimension $65 \times 66 \times 203$ (longitude by latitude by time). Although the number of observations cannot be regarded as 'large' in the sense of 'big data', this 3-dimensional dataset can still be challenging for statistical analysis. Therefore, spatio-temporal

models exploiting dimension reduction are investigated. Specifically, a functional data representation is used to reduce the data dimensionality and hierarchical dynamic spatio-temporal models (DSTM) constructed under the state space framework (Cressie & Wikle, 2011) are used to describe the spatio-temporal features of the data. Due to technical and atmospheric issues, remote-sensing data can have a substantial amount of missing data (see the data panels of Figure 1). Therefore, modifications are proposed here in the estimation algorithms to account for missing data.

2 The modelling framework

Consider a hierarchical DSTM in state space form with, (1) a data model relating the observation $Z(\boldsymbol{s};t)$ to a 'true' spatio-temporal process $Y(\boldsymbol{s};t)$ and (2) a process model describing the dynamics of the latent process through lagged dependence,

$$Z(\boldsymbol{s};t) = Y(\boldsymbol{s};t) + \epsilon(\boldsymbol{s};t) = \Phi(\boldsymbol{s})\boldsymbol{\beta}_t + \zeta(\boldsymbol{s};t) + \epsilon(\boldsymbol{s};t)\ , \tag{1}$$

$$\boldsymbol{\beta}_t = \sum_q \boldsymbol{M}_q \boldsymbol{\beta}_{t-\tau_q} + \boldsymbol{u}_t\ . \tag{2}$$

Dimension reduction comes from the representation of $Y(\boldsymbol{s};t)$ as $\Phi(\boldsymbol{s})\boldsymbol{\beta}_t + \zeta(\boldsymbol{s};t)$, where $\Phi(\boldsymbol{s})$ is a basis matrix and $\zeta(\boldsymbol{s};t)$ is a non-dynamic component. If matrix $\Phi(\boldsymbol{s})$ is of a lower rank K than the dimension of the data vector $Z(\boldsymbol{\cdot};t)$, being N, then the dimension of the process model (2) would be reduced to K. Substantial computational gains may be achieved if $K \ll N$, which is often the case for remote-sensing data. The process dynamic follows a vector auto-regressive (VAR) model of order q, reflecting the temporal dependence of the spatial process $Y(\boldsymbol{s};t)$. To ensure identifiability, a parameter model putting constraints on model components may be included. Two parameterizations of this framework are proposed and investigated here. Both methods offer opportunities for dimension reduction, but with different emphases, where one focuses on the general spatio-temporal pattern and the other focuses on the spatio-temporal prediction.

2.1 The FPC parameterization

The functional principal component (FPC) parameterization is based on the empirical orthogonal function (EOF) parameterization in Xu & Wikle (2007). It maps the data to the leading EOFs extracted from the data,

$$Z(\boldsymbol{s},t) = \sum_{p=1}^{a} \xi_p(\boldsymbol{s})\alpha_{pt} + \epsilon(\boldsymbol{s},t) = \Xi(\boldsymbol{s})\boldsymbol{\alpha}_t + \epsilon(\boldsymbol{s},t)\ , \tag{3}$$

and parameterizes the covariance matrix of the vector of residuals $\boldsymbol{\epsilon}_t$ as $\sigma^2\boldsymbol{I} + \sum_{p=a+1}^{P} \lambda_p \boldsymbol{\xi}_p \boldsymbol{\xi}_p^\top$, where $\boldsymbol{\xi}_p$ is the vectorized EOF $\xi_p(\boldsymbol{s})$ evaluated at all locations $\boldsymbol{s}$ and λ_p is the corresponding eigenvalue. The process model is specified as a first order VAR, which is $\boldsymbol{\alpha}_t = \boldsymbol{M}\boldsymbol{\alpha}_{t-1} + \boldsymbol{u}_t$. Exploratory analysis suggests that first order dependence is appropriate for the LSWT data after removing the trend and seasonality. The component $\zeta(\boldsymbol{s};t)$ is not considered here. In the case of the remote-sensing LSWT data, functional representation is applied and functional PCs are extracted as the analogous of the EOFs for dimension reduction. Model estimation uses the EM algorithm with the Kalman filter/smoother. Model results provide information on spatial patterns through the functional PCs and temporal evolutions through the estimated process model.

2.2 The STRE model

The spatio-temporal random effect (STRE) model of Cressie et al., (2010) can be written in the same way as formulae (1) and (2). In particular, $\Phi(\boldsymbol{s})$ is usually taken to be a spatial basis and the component $\zeta(\boldsymbol{s};t)$ represents the non-dynamic random effect unique to each spatial image $Z(\boldsymbol{s};t)$, which cannot be captured by the dynamic of $Y(\boldsymbol{s};t)$. Again, dimension reduction can be achieved through a basis representation $\Phi(\boldsymbol{s})\boldsymbol{\beta}_t$. Model estimation uses the EM algorithm, along with the fixed rank filter (FRF) and smoother (FRS) (Cressie et al., 2010; Katzfuss & Cressie, 2011). This method estimates the time-varying $\boldsymbol{\beta}_t$ using the Kalman filter/smoother and the random effect $\boldsymbol{\zeta}_t$ through a second filter based on the conditional distribution of $(\boldsymbol{\zeta}_t, \boldsymbol{\beta}_t) \,|\, \boldsymbol{Z}_{1:t}$ (FRF) and $(\boldsymbol{\zeta}_t, \boldsymbol{\beta}_t) \,|\, \boldsymbol{Z}_{1:T}$ (FRS), where $\boldsymbol{Z}_{1:t}$ represents observed data $\{\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_t\}$. In particular, it is assumed that the non-dynamic component $\boldsymbol{\zeta}_t$ only depends on the information of time point t. Temporal patterns can be extracted from the estimated process model. Spatial patterns may be modelled by assigning a correlation structure to the residual covariance matrix.

2.3 Implementation and results

The FPC parameterized model and the STRE models are then applied to a subset of the Lake Victoria LSWT data (dimension $36 \times 47 \times 202$). The subset is taken to minimize the influence of land pixels and lake border pixels, which tend to have larger uncertainties. As the data are stored on a regular grid and that shape is not critical to this analysis, a tensor spline basis is specified for the spatial basis $\Phi(\boldsymbol{s})$. A basis accounting for the shape of the lake may be used, but the implementation would require a lot more computational cost. The smoothness (degrees of freedom) of the basis is controlled directly by the number of knots of the tensor splines. This is out of the concern for the computational cost of tunning a smoothing parameter. Information criteria AIC and BIC are used to select the degrees of freedom. The variance proportion criterion is applied to select the number of FPCs in model (3). In this case, an 80% threshold gives 11 FPCs in the dynamic component of the model; a 95% threshold selects a further 11 eigenfunctions to form the residual covariance matrix. A random walk model is used for modelling the process dynamic, i.e. $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{u}_t$ for the FPC parameterized model and $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{u}_t$ for the STRE model. This is appropriate considering the feature of the LSWT data after removing a seasonal mean.

The R code for implementing the two methods has been developed as part of this work. To accommodate missing data, the approach similar to the Kalman filter for sparse data (Shumway & Stoffer, 2006) is adopted. The implementation of FRF and FRS follows the procedure described in Katzfuss & Cressie (2011). In extreme circumstances where there are only a few observations available for a spatial image, a filtering threshold may also be applied to avoid over-interpolation.

The fitted LSWT images constructed using the smoothed $\boldsymbol{\alpha}_t$ or $\boldsymbol{\beta}_t$ and MLEs from the converged EM algorithms of the two models are shown in Figure 1. Both methods provide a good fit to the data. The residual sum of squares (RSS) from the FPC parameterized model is 0.1021; that of the STRE model is 0.0810. The fitted images from the FPC parameterized model are smoother, as the model is designed to capture the general patterns. The results from the STRE model show more detail, as the model is designed for interpolation and prediction. The contrast in the fitted images of the FPC parameterized model appears to be larger than that of the STRE model. In terms of very sparse images, imputation with smaller contrast maybe preferred to avoid over-interpolating the unobserved areas.

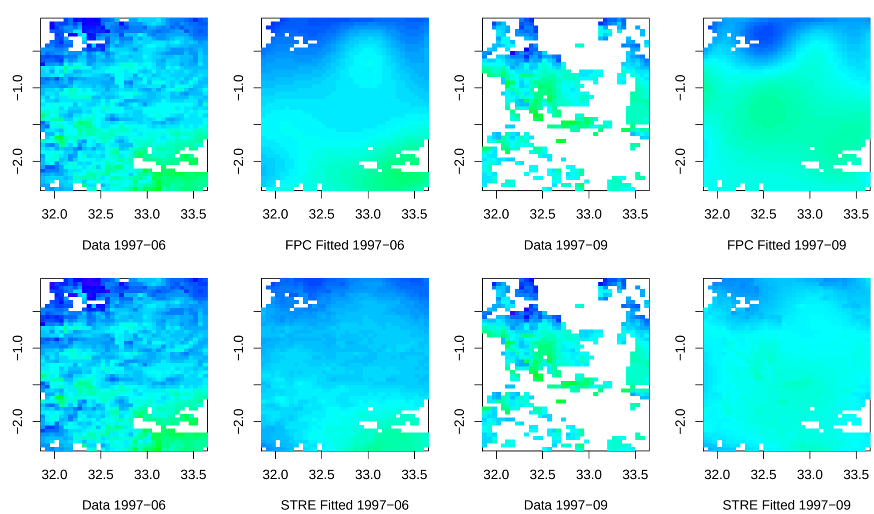

FIGURE 1. Examples of the Lake Victoria LSWT data and their fitted versions using the FPC parameterized model (upper) and the STRE model (lower).

An investigation can also be carried out on the smoothed process state to understand the patterns of the process dynamics. For the FPC parameterized model, the smoothed states reflect the temporal evolution of the corresponding FPCs. An example of the smoothed α_{1t} and α_{2t} time series from model (3) is given in Figure 2. In this case, α_{1t} seems to be showing certain seasonal fluctuations not covered by the seasonal mean; whereas α_{2t} displays mainly random fluctuation with a few spikes. The dynamics of $\boldsymbol{\beta}_t$ in the STRE model (2) might be less straightforward to interpret, as they are spatial basis coefficients that do not always have a clear meaning in practice. Nonetheless, the time series of $\boldsymbol{\beta}_t$ may still be useful to aid with the understanding of spatio-temporal patterns in the data.

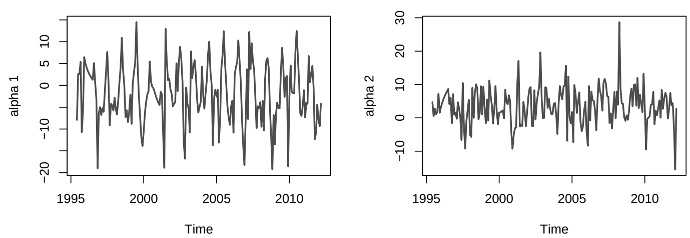

FIGURE 2. Examples of the dynamics of the FPCs from model (3).

Finally, the model residuals are investigated. The images in Figure 3 present the pixel-wise RSS from the two models, reflecting the regional fit of the models as opposed to the overall fit. The pixels towards the northwest corner appear to have larger RSS values. However, there is not any big discrepancies between the RSS of different pixels, suggesting that the two methods are appropriate for providing missing data imputations despite the varying data availability in different areas of the lake. Katzfuss & Cressie (2011) also derived the formula of the mean squared prediction errors (MSPE) for the STRE model (1), which are defined as the diagonal elements of $\mathbf{E}\left[(\boldsymbol{Y}_t - \boldsymbol{Y}_{t|T})(\boldsymbol{Y}_t - \boldsymbol{Y}_{t|T})^\top\right]$, where $\boldsymbol{Y}_t$ is the vectorized $Y(\boldsymbol{s};t)$ and $\boldsymbol{Y}_{t|T}$ is the FRS version of $\boldsymbol{Y}_t$. In this case, the spatial pattern of the MSPEs corresponds to that of the missing percentages, but again, the values are at a similar scale.

3 Discussion

The FPC parametrized model and the STRE model provide two efficient approaches to the spatio-temporal modelling of the sparse high-dimensional remote-sensing data. Missing data imputation can be carried out while the spatial and temporal patterns are extracted. One criticism on the EOF (i.e. FPC in this case) based method is that the leading principal components may not be adequate in explaining the dominant system dynamics, despite

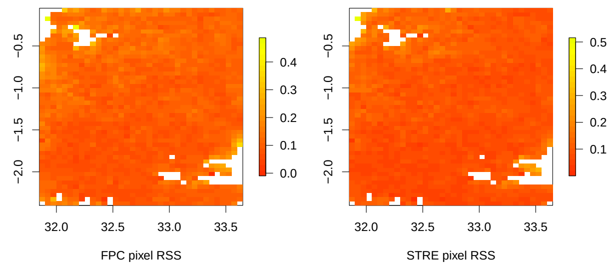

FIGURE 3. The pixel-wise RSS from the Lake Victoria LSWT data fitted using the FPC parameterized model (left) and the STRE model (right).

their power in describing the variation in the data. For the STRE model, the random component $\boldsymbol{\zeta}_t$, while accounting for the individual effect unique to each spatial image, cannot provide a conclusive summary of the spatial variation. Unless the residual spatial structure is modelled, which could be computationally expensive, it is difficult to use the STRE model to identify the spatial variation patterns. Potential developments may be to parameterize the random effect $\boldsymbol{\zeta}_t$ to reflect certain spatial patterns in a flexible manner. This will be investigated in the future to improve the modelling of the remote-sensing environmental data.

Acknowledgments: The authors are grateful to the ARC-Lake project for access to the data. The authors were partly funded for this work through the NERC GloboLakes project (NE/J022810/1).

References

Cressie, N., Shi, T., and Kang, E.L. (2010). Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics*, **19**, 723 – 745.

Cressie, N., and Wikle, C. (2011). *Statistics for Spatio-temporal Data*. New-Jersey: John Wiley & Sons.

Katzfuss, M., and Cressie, N. (2011). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, **32**, 430 – 446.

Shumway, R. H., and Stoffer, D. S. (2006). *Time Series Analysis and Its Application with R Examples*, New York: Springer.

Xu, K., and Wikle, C. (2007). Estimation of parameterized spatio-temporal dynamic models. *Journal of Statistical Planning and Inference*, **137**, 567 – 588.