Stochastic Environmental Research and Risk Assessment
https://doi.org/10.1007/s00477-021-02017-w

ORIGINAL PAPER

State space functional principal component analysis to identify spatiotemporal patterns in remote sensing lake water quality

Mengyi Gong[1] · Claire Miller[2] · Marian Scott[2] · Ruth O'Donnell[2] · Stefan Simis[3] · Steve Groom[3] · Andrew Tyler[4] · Peter Hunter[4] · Evangelos Spyrakos[4]

Accepted: 2 April 2021
© The Author(s) 2021

Abstract
Satellite remote sensing can provide indicative measures of environmental variables that are crucial to understanding the environment. The spatial and temporal coverage of satellite images allows scientists to investigate the changes in environmental variables in an unprecedented scale. However, identifying spatiotemporal patterns from such images is challenging due to the complexity of the data, which can be large in volume yet sparse within individual images. This paper proposes a new approach, state space functional principal components analysis (SS-FPCA), to identify the spatiotemporal patterns in processed satellite retrievals and simultaneously reduce the dimensionality of the data, through the use of functional principal components. Furthermore our approach can be used to produce interpolations over the sparse areas. An algorithm based on the alternating expectation–conditional maximisation framework is proposed to estimate the model. The uncertainty of the estimated parameters is investigated through a parametric bootstrap procedure. Lake chlorophyll-*a* data hold key information on water quality status. Such information is usually only available from limited in situ sampling locations or not at all for remote inaccessible lakes. In this paper, the SS-FPCA is used to investigate the spatiotemporal patterns in chlorophyll-*a* data of Taruo Lake on the Tibetan Plateau, observed by the European Space Agency MEdium Resolution Imaging Spectrometer.

Keywords Functional principal component analysis · State space model · AECM algorithm · Remote sensing images · Lake chlorophyll-*a*

1 Introduction

Satellite remote sensing technology provides a novel source of information for environmental monitoring, with increasingly high spatial and temporal resolution. Space borne sensors, such as the MEdium Resolution Imaging Spectrometer (MERIS), on board the European Space Agency Envisat platform, can be used to observe lake ecosystems around the world (Hout et al. 2001). The application that motivates this research is the investigation of spatiotemporal changes of lakes under environmental change through monthly lake chlorophyll-*a* (Chl) retrievals from the MERIS. Lakes are regarded as sentinels of change (Williamson et al. 2009) and 'social development and economic prosperity depend on the sustainable management of freshwater resources and ecosystems' (The United Nations 2018). Remote sensing data, such as the MERIS Chl retrievals, are a great source of information to complement in situ measurements and can provide information on water bodies which are otherwise inaccessible.

The data that motivates the model development in this paper is the Chl time series of Taruo Lake (also known as Taro Co). Taruo Lake is located on the southwestern Tibetan Plateau. It is a saline lake, has an elevation of 4566 m and covers an area of 486.6 km^2 (Alivernini et al. 2018). The major Tibetan lakes are regarded as indicators of climate change on the Tibetan Plateau (Wu et al. 2017).

✉ Mengyi Gong
m.gong1@lancaster.ac.uk

1 Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK

2 School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, UK

3 Plymouth Marine Laboratory, Plymouth PL1 3JH, UK

4 School of Biological and Environmental Science, University of Stirling, Stirling KF9 4LA, UK

Published online: 21 April 2021

Research on the changing water level, temperature and paleo-climatology has been carried out in order to better understand the regional responses to the changing climate in recent decades (Fang et al. 2016; Huang et al. 2017; Ma et al. 2014). However, the often remote locations of these lakes make it difficult to monitor them regularly. In the meantime, ground sampling to accurately capture the spatiotemporal change across the lake is not possible because of its size. In such situations, remote sensing data provide a unique source of information and have shown to be crucial to the study of lake areas and water levels (Wu et al. 2017; Fang et al. 2016). The Chl data analysed in this paper are retrieved by the Plymouth Marine Laboratory, as part of the GloboLakes project (www.globolakes.ac.uk). Section 1 of the supplement provides details of the data processing procedure. The data set provides observations of monthly Chl data from June 2002 to April 2012 (i.e. 119 months) at a spatial resolution of $0.0027^\circ \times 0.0027^\circ$, for over 1000 large lakes globally, including Taruo Lake. Examples of the retrieved Chl images from Taruo Lake are displayed in Fig. 1. Log transformation was applied here as the original data are right skewed. The percentage of pixels with observations in the lake area over the monitoring period is 74.58%.

Remote sensing data are often stored as a three-dimensional array, indexed by longitude, latitude and time. They can be viewed as a time series of spatial images, where each image is comprised of a set of pixels. The dimensionality of the data, in terms of both the number of images and the number of pixels which make up each image can be high, particularly as technology advances and the resolution of remote sensing instruments increases. However, despite the potential for high dimensional data, it is often the case that there are relatively high percentages of missing observations in both space and time due to factors such as cloud cover inhibiting sensor views, or losses in communication with instruments. This combination of high dimensional but sparse data sets, along with the spatial and temporal correlations typically displayed in data which are collected at locations and time points close to one another, presents a substantial challenge in the statistical modelling of such data.

The model proposed within this paper aims to identify and investigate the main spatiotemporal patterns in potentially high dimensional sparse data, which account for the complexities in the underlying correlation structure and reduce the dimensionality of the data to ensure computational efficiency. More specifically this paper will consider the example of modelling the spatiotemporal patterns in the remote sensing lake chlorophyll-*a* data of Taruo Lake. Throughout the paper, the phrase 'spatiotemporal patterns' refers to the statistical measures that reflect the spatial/temporal features in the data and their evolutions, e.g. an (auto)correlation function describing the dependence of data in time or space, a principal component showing the contrasts of data in different areas of the lake.

Functional data analysis (FDA) is a common approach to modelling high dimensional data. Here a functional data representation is proposed to transform the image data into bivariate functions. This helps to reduce the dimension of the remote sensing images from the number of pixels N, to the dimension of the basis K ($K \ll N$). Then a functional principal component analysis (FPCA) can be applied to these bivariate functions. A principal component analysis (PCA), or empirical orthogonal function (EOF) analysis as referred to in climate science literature, is widely used to identify important spatial patterns or driving forces of an environmental process (National Center for Atmospheric Research 2013). In the case of FPCA, the resulting functional principal components (PCs) provide not only an interpretation of the spatiotemporal covariance structure of the data, but also a way of further dimension reduction by keeping only the functional PCs that contribute most to the variation in the data in further analysis. However, substantive missing observations prohibit the use of standard methods, such as eigen-decomposition and singular value decomposition, to obtain the functional PCs. Therefore, the FPCA is reframed as a mixed effect model where the functional PCs are specified as the random effect of the mixed effect model (James et al. 2000; Zhou and Pan 2014). This method, referred to as the FPC model in this paper, can be estimated using the EM algorithm (Rice and Wu 2001), which is widely used in incomplete data estimation problems.

However, the FPC model in James et al. (2000) and Zhou and Pan (2014) assumes that the functional objects (e.g. the longitudinal curves, the spatial images) are independent from each other. This may not be appropriate for modelling remote sensing Chl data, and implementing functional PCA on correlated data could compromise the identification of the functional PCs. To overcome this problem, this paper proposes a new method, the state space functional principal component analysis (SS-FPCA), to estimate the functional PCs while taking into account the temporal correlations in the data. In a nutshell, the SS-FPCA extends the model in James et al. (2000) by adding a vector autoregressive (VAR) structure and further reframe it as a state space model. This new approach is motivated by the spatiotemporal random effect (STRE) model of Cressie et al. (2010). The STRE is a state space model of three hierarchies. It consists of

1. a data model (i.e. the observation equation in a state space model) associating the observations in specific locations at time t to the spatiotemporal process of interest;

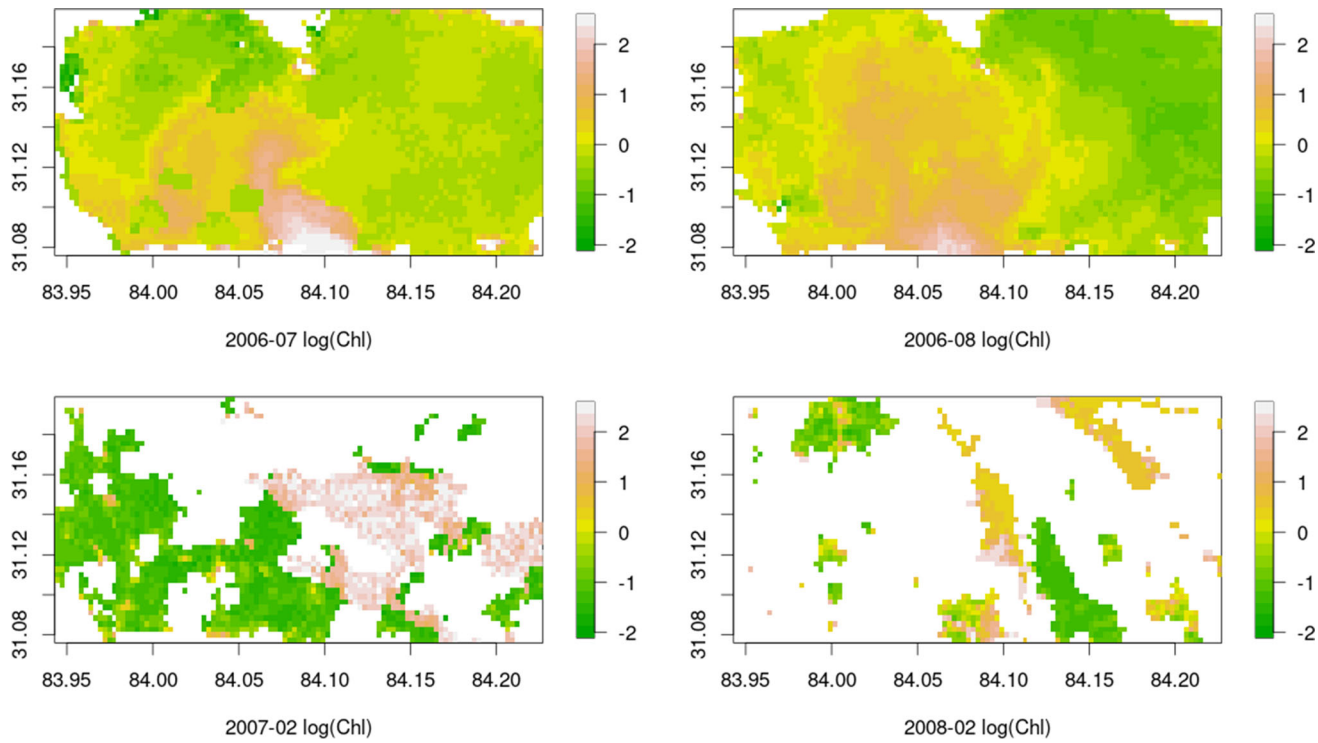

Fig. 1 Examples of the remote sensing Chl of Taruo Lake from four different months. These images are based on the 'trimmed lake', where the lake border and the narrow parts towards the edge of the lake have been removed. The detail of the trimming procedure is given in Sect. 4

2. a process model (i.e. the state transition equation in a state space model) describing the spatiotemporal dynamic of the process of interest;
3. a parameter model specifying distributional assumptions and constraints ensuring model identifiability.

There are numerous ways of specifying the temporal structure in the STRE model, e.g. VAR process, PDEs describing a physical process (Xu and Wikle 2007). Whereas the spatial covariance structure in the STRE is typically modelled using a spatial covariance function, the SS-FPCA model uses a functional PCA to describe the spatial covariance structure.

The STRE model can be implemented using the empirical hierarchical modelling (EHM) based on the maximum likelihood method (Katzfuss and Cressie 2011) or the Bayesian hierarchical modelling (BHM) exploiting a fully Bayesian approach (Katzfuss and Cressie 2012). This paper follows the EHM approach and develops 2-cycle alternating expectation–conditional maximisation (AECM) algorithm (Meng and Van Dyk 1997) to estimate the SS-FPCA model. The AECM algorithm is an extension of the EM algorithm. Its flexibility enables the computation to be carried out with fewer numerical optimization procedures than a standard EM algorithm would require.

To further investigate the performance of the estimation algorithm and the model's ability in extracting spatial and temporal patterns, a simulation study is carried out on a 1-dimensional space. The simulated data, though not in the same dimension as the remote-sensing Chl data, are appropriate in assessing the model because the model specification and estimation algorithm do not change with the dimension. To obtain the standard errors of the estimated model parameters, this paper adopted a spatiotemporal bootstrap procedure proposed by Fassò and Cameletti (2009). This may be computationally intensive for large models, but it is straightforward to implement.

The rest of the paper consists of four sections. Section 2 provides the preliminaries of the SS-FPCA model, followed by the formal introduction of the SS-FPCA. Section 2.1 proposes the 2-cycle AECM algorithm, along with the bootstrap method of parameter standard errors. The simulation study is presented in Sect. 3 and the application of the model to the remotely sensed Chl is given in Sect. 4. Section 5 concludes the paper with a discussion of some potential future extensions.

2 The state space functional PCA (SS-FPCA) model

The SS-FPCA model is a 3-level hierarchical model built under the state space model framework. It consists of an observation equation describing the temporal mean and the spatial covariance structure through functional PCA, a state

transition equation categorizing the temporal dynamic of the process and the corresponding distributional assumptions ensuring the identifiability of the model. The SS-FPCA model can be written using the following two equations,

$$\begin{aligned} Z_{t,s} &= Y_t(\boldsymbol{s}) + \epsilon_t(\boldsymbol{s}) \\ &= \mu_t(\boldsymbol{s}) + \Phi(\boldsymbol{s})^\top \boldsymbol{\beta}_t + \sum_{p=1}^{P} \Phi(\boldsymbol{s})^\top \boldsymbol{\theta}_p \alpha_{pt} + \epsilon_t(\boldsymbol{s}) \qquad (1) \\ &= \mu_t(\boldsymbol{s}) + \Phi(\boldsymbol{s})^\top \boldsymbol{\beta}_t + \Phi(\boldsymbol{s})^\top \boldsymbol{\Theta}\boldsymbol{\alpha}_t + \epsilon_t(\boldsymbol{s}) \end{aligned}$$

$$\boldsymbol{\beta}_t = \boldsymbol{M}\boldsymbol{\beta}_{t-1} + \boldsymbol{u}_t, \qquad (2)$$

where $Z_{t,s}$ is the observed process at location $\boldsymbol{s}$, $Y_t(\boldsymbol{s})$ is the underlying spatial process and $\epsilon_t(\boldsymbol{s})$ is the measurement error. The level 1 data model (1) consists of the components of the underlying process. Specifically, $\mu_t(\boldsymbol{s})$ is an optional fixed mean component. The term $\Phi(\boldsymbol{s})^\top \boldsymbol{\beta}_t$ is the dynamic mean function constructed using spatial basis functions $\Phi(\boldsymbol{s}) = (\phi_1(\boldsymbol{s}, \ldots, \phi_K(\boldsymbol{s}))^\top$ and time varying basis coefficient vector $\boldsymbol{\beta}_t$ (referred to as the dynamic component), modelling the temporal evolution of the spatial process. $\boldsymbol{\beta}_t$ are assumed to be temporally dependent to capture the temporal dependence in the data, and they do not need to be stationary. The term $\Phi(\boldsymbol{s})^\top \boldsymbol{\Theta}\boldsymbol{\alpha}_t$ is an infinite Karhunen–Loève expansion truncated at the Pth order[1] using P eigenfunctions $\Phi(\boldsymbol{s})^\top \boldsymbol{\theta}_p$, $p = 1, \ldots, P$, and the corresponding random coefficient vectors $\boldsymbol{\alpha}_t$ (referred to as the FPCA component), describing the non-dynamic spatial variation in the data. The order P is typically chosen so that the truncated expansion accounts for an appropriate amount of variation in the data. More information on the selection of order P is given in Sect. 2.3.1. The random coefficients $\boldsymbol{\alpha}_t$, though indexed by t, are assumed to be temporally independent. The level 2 process model (2) specifies the temporal dynamic in the data. Here the dynamic is characterized by a 1st order vector autoregressive model, VAR(1), with coefficient matrix $\boldsymbol{M}$. This is motivated by the exploratory analysis of the remote sensing Chl data. As the deterministic part of the temporal structure (e.g. seasonality) is accounted for by the mean component $\mu_t(\boldsymbol{s})$, a first order dependence is appropriate for this application. In particular, the autocorrelation functions of the Chl time series from individual pixels consisting of Taruo Lake were computed in the exploratory analysis, and it appears that majority of the time series follow an AR(1) process. Higher order dependence is plausible. The space-time AR(p) model in Lagos-Álvarez et al. (2019) and Padilla et al. (2020) provides a potential route. Although it may complicate the model estimation, partly due to the difficulty in deciding on the optimal order, and partly due to the increased computational cost as high dimensional optimization procedures will be inevitable.

The following distributional assumptions are made. The measurement errors $\epsilon_t(\boldsymbol{s})$ are assumed to be independently and identically distributed (i.i.d) as $\mathcal{N}(0, \sigma^2)$. The residuals of the process model $\boldsymbol{u}_t$ are assumed to be normally distributed as $\mathcal{N}(\boldsymbol{0}, \boldsymbol{H})$, where $\boldsymbol{H}$ is symmetric and positive definite. The basis matrix $\boldsymbol{\Phi} = (\phi_1(\boldsymbol{s})^\top, \ldots, \phi_K(\boldsymbol{s})^\top)^\top$ and the coefficient matrix $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_P)$ are assumed to be orthonormal, so that their product would be orthogonal eigenfunctions. The random coefficient vector $\boldsymbol{\alpha}_t$ is required to satisfy the assumptions of PC scores as in James et al. (2000), with covariance matrix $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_P\}$, and λ_p, $p = 1, \ldots, P$, arranged in decreasing order. In other words,

$$\begin{aligned} &\boldsymbol{\Phi}^\top \boldsymbol{\Phi} = \mathbf{I}, && \boldsymbol{\Theta}^\top \boldsymbol{\Theta} = \mathbf{I} \\ &\boldsymbol{\alpha}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}), && \boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_P\} \\ &\boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \mathbf{I}), && \boldsymbol{u}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}). \end{aligned}$$

Some remarks on the SS-FPCA model are made here.

1. The SS-FPCA model has a structure that resembles the STRE model in Cressie et al. (2010) and Katzfuss and Cressie (2011). However, the two models are built for different purses. Hence the specification of the spatial and temporal components differs. Whilst the STRE model aims at spatiotemporal prediction, the SS-FPCA model focuses more on identifying the spatiotemporal patterns in the form of functional PCs. Instead of using a spatial correlation function to describe the spatial variation as in the STRE, the SS-FPCA interprets the spatial variation through the functional PCs.
2. The VAR(1) coefficient matrix $\boldsymbol{M}$ may be parameterised according to the properties of the data. For example, $\boldsymbol{M} = \mathrm{diag}\{m_1, \ldots, m_K\}$ assumes that each element in $\boldsymbol{\beta}_t$ evolves separately and $\boldsymbol{M} = \mathbf{I}$ corresponds to a local level model for the system transition equation (2). According to Shumway and Stoffer (2006), estimation of an unconstrained $\boldsymbol{M}$ is straightforward. The estimation of a diagonal $\boldsymbol{M}$ can also be made using only analytic solutions, following the method in Xu and Wikle (2007). The estimation of a coefficient matrix $\boldsymbol{M}$ of more complicated structure would require more effort. However, it could be beneficial if the data suggest such a dynamic structure.
3. It is possible to parameterise the $\boldsymbol{H}$ matrix using, e.g. a covariogram model and a conditional autoregressive model (Cressie and Wikle 2011; Xu and Wikle 2007),

[1] The Karhunen–Loève expansion is an infinite expansion of a stochastic process, whose covariance function can be defined through a Hilber–Schmit kernel. The connection between the K–L expansion and functional PCA is explained in James et al. (2000) and Zhou and Pan (2014).

to reflect the spatiotemporal dynamic of the real process. Since the state transition equation (2) is to do with the spatial basis coefficients, the structure of its residuals is not straightforward to see. Therefore, in this paper, $\boldsymbol{H}$ is left unstructured to avoid many impractical assumptions.

4. To ensure identifiability, it is assumed that $\{\boldsymbol{\beta}_t\}_{t=1}^T$ and $\{\boldsymbol{\alpha}_t\}_{t=1}^T$ are independent; $\{\boldsymbol{\beta}_t\}_{t=1}^T$ is independent of $\{\boldsymbol{\epsilon}_t\}_{t=1}^T$; $\{\boldsymbol{\alpha}_t\}_{t=1}^T$ is independent of $\{\boldsymbol{u}_t\}_{t=1}^T$ and $\{\boldsymbol{\epsilon}_t\}_{t=1}^T$. In addition, it is assumed that the estimation of the dynamic coefficient $\boldsymbol{\beta}_t$ at time point t relies on information from all the observed data $\{\boldsymbol{Z}_t\}_{t=1}^T$. On the other hand, the estimation of the PC scores $\boldsymbol{\alpha}_t$ at time point t requires only the information from $\boldsymbol{Z}_t$ as in a FPCA.

2.1 Model estimation

In theory, both the EHM (likelihood method) and the BHM (Bayesian method) are applicable to the estimation of the SS-FPCA model (in a similar way to the STRE model). However, when it comes to the large volume of the remote sensing data, the EHM becomes a natural choice as a fully Bayesian approach would be computationally intensive. Implementation using the BHM could encounter difficulties, e.g. sampling from high-dimensional distributions and monitoring convergence of a large number of parameters. On the contrary, the EHM implementation requires only analytical solutions or low-dimensional numerical optimizations. In this section, a 2-cycle AECM algorithm is developed to estimate the SS-FPCA model.

2.2 The alternating expectation–conditional maximisation (AECM) method

The AECM algorithm was first proposed by Meng and Van Dyk (1997) and was developed based on various extensions of the classic EM algorithm. It makes use of data augmentation and model reduction to create an efficient algorithm for models with complex structures. Particularly, data augmentation refers to the 'methods for constructing iterative optimization or sampling algorithms via the introduction of unobserved data or latent variables' (van Dyk and Meng 2001) and model reduction refers to 'using a set of conditional distributions in a computation method designed to learn about the corresponding joint distribution' (van Dyk and Meng 2010). Both techniques, when appropriately applied, lead to an improved algorithm.

An AECM algorithm typically consists of C $(C \geq 1)$ cycles within each iteration. Each cycle corresponds to one type of data augmentation and is paired with S_c $(S_c \geq 1)$ conditional maximisation (CM)[2] steps. The subscript c of S indicates that the number of CM-steps is allowed to vary with cycles (Meng and Van Dyk 1997), giving more flexibility to the design of the algorithm. Omitting the iteration index (it), the target function in the E-step of the $(c+1)$th cycle of the AECM algorithm can be written as (Meng and Van Dyk 1997)

$$\mathcal{Q}^{[c+1]}\left(\Psi;\Psi^{[c]}\right) = \mathbf{E}\left[\mathcal{L}\left(\Psi;\boldsymbol{Z}_{aug}^{[c+1]}\right)\middle|\boldsymbol{Z}_{obs},\Psi^{[c]}\right],$$

where Ψ is the complete parameter set, $\Psi^{[c]}$ is the parameter set corresponding to the data augmentation of cycle c, $\boldsymbol{Z}_{obs}$ is the observed data set, $\boldsymbol{Z}_{aug}^{[c+1]}$ is the augmented data set in cycle $c+1$ and $\mathcal{L}(\cdot)$ represents the log-likelihood function. Then the sth CM-step in cycle $c+1$ finds $\Psi^{[c+\frac{s}{S_{c+1}}]}$ such that

$$\begin{aligned}&\mathcal{Q}^{[c+1]}\left(\Psi^{[c+\frac{s}{S_{c+1}}]};\Psi^{[c]}\right) \geq \mathcal{Q}^{[c+1]}\left(\Psi;\Psi^{[c]}\right)\\ &\quad \forall\,\Psi \in \mathcal{W}_s^{[c+1]} \equiv \left\{\Psi \in \mathcal{W} : g_s^{[c+1]}(\Psi)\right.\\ &\qquad \left. = g_s^{[c+1]}\left(\Psi^{[c+\frac{s}{S_{c+1}}]}\right)\right\},\end{aligned}$$

where $\mathcal{W}_s^{[c+1]}$ is the parameter space and $g_s^{[c+1]}(\Psi)$ is the corresponding constraint function. Due to its flexibility, the AECM algorithm has seen wide applications. Examples include the estimation of mixture models (McLachlan et al. 2003; McNicholas and Murphy 2008), fitting mixed models with non-exponential family distributions (Ho and Lin 2010).

2.3 The 2-cycle AECM for the SS-FPCA model

The AECM algorithm for the SS-FPCA model consists of two cycles. Specifically, the first cycle estimates the parameters $\boldsymbol{H}$, $\boldsymbol{M}$ and random coefficient $\boldsymbol{\beta}_t$ associated with the dynamic component; the second cycle estimates the parameters and random effects associated with the FPCA component, $\boldsymbol{\Theta}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\alpha}_t$, and the error variance σ^2. Using this partition of parameter space and the corresponding data augmentation, all parameters have closed form expressions for their MLEs. A flow chart of this algorithm is given in Fig. 2.

To introduce the algorithm in full, the following notations are used. The subscripts *obs, mis* and *aug* indicate the observed, missing and augmented data respectively. The superscript [1], [2] are the cycle indices and superscript (it) is the iteration index. The subscript $1{:}t$ refers to the time

[2] In a generalised EM algorithm, instead of maximise the likelihood over all parameters simultaneously, 'conditional maximisation' is sometimes used where the computation of the maximum likelihood estimation of some of the parameters is conditioned on the current optimal values of the other parameters. Details on generalised EM algorithm can be found in McLachlan and Krishnan (1997).

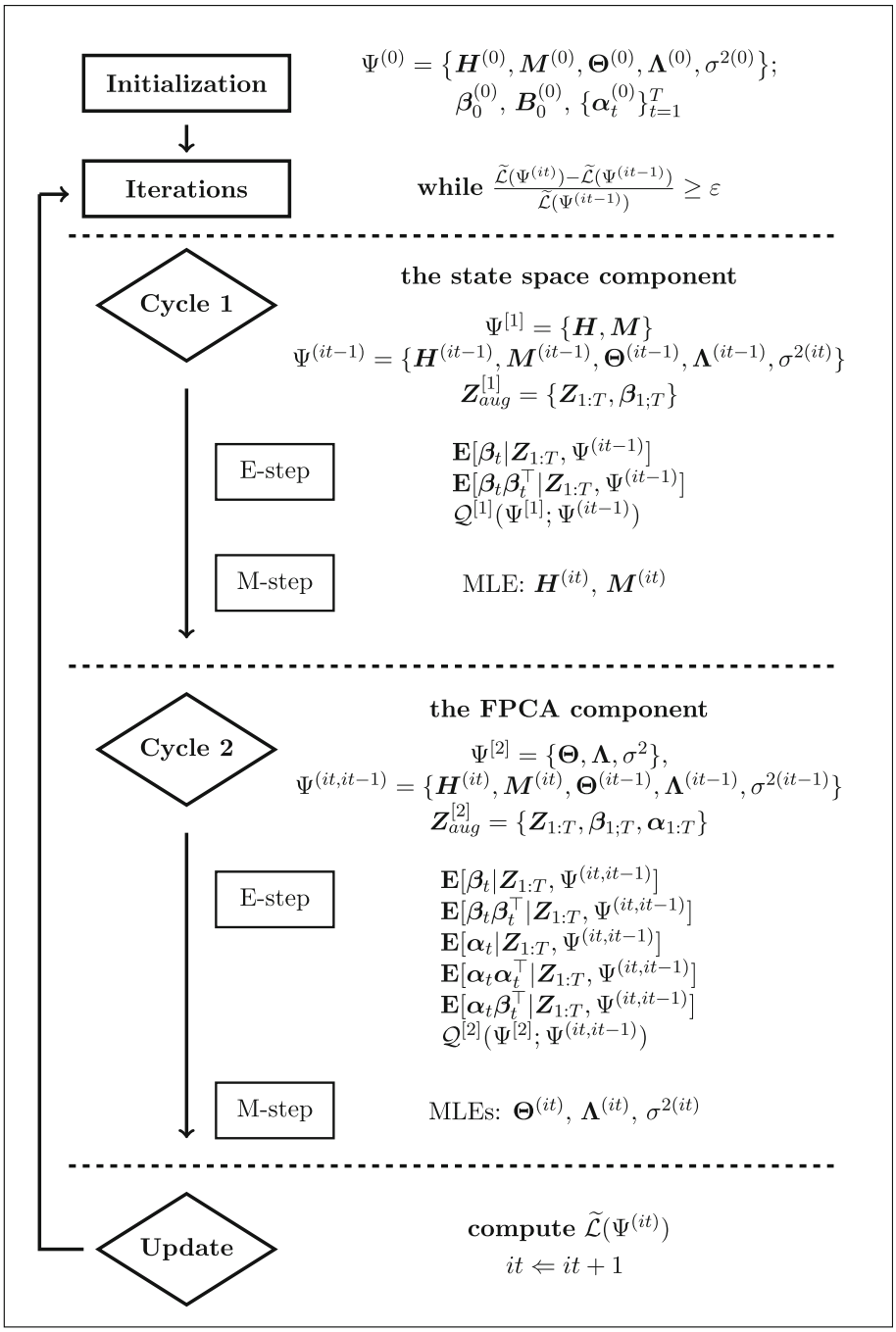

Fig. 2 The flow chart of the 2-cycle AECM algorithm for the SS-FPCA model

series from time point 1 to t, e.g. $\boldsymbol{Z}_{1:T} = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_T\} = \{\boldsymbol{Z}_t\}_{t=1}^T$, where $\boldsymbol{Z}_t = (Z_{t,s_1}, \ldots, Z_{t,s_n})^\top$ is the vector of all observations at time t. To emphasize the impact of missing observations, the basis matrix corresponding to the observed locations at time t is denoted as $\boldsymbol{\Phi}_t$, which is matrix $\boldsymbol{\Phi}$ with rows corresponding to missing locations removed.

Specifically, the observed data of cycle 1 are $\boldsymbol{Z}_{obs} = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_T\}$; the missing data are $\boldsymbol{Z}_{mis} = \{\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_T\}$. The random component $\boldsymbol{\alpha}_t$ is treated as part of the residuals in this cycle. The complete data distribution is

$$f(\boldsymbol{Z}_{1:T}, \boldsymbol{\beta}_{0:T}) = \prod_{t=1}^{T} f(\boldsymbol{Z}_t|\boldsymbol{\beta}_t;\, \Psi) f(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1};\, \Psi) f(\boldsymbol{\beta}_0;\, \Psi). \tag{3}$$

This gives the likelihood function and hence the target function $\mathcal{Q}^{[1]}\left(\Psi^{[1]};\, \Psi^{(it-1)}\right)$ for maximisation as $\mathbf{E}\left[-2\mathcal{L}\left(\Psi^{[1]}; \boldsymbol{Z}^{[1]}, \widetilde{\Psi}^{[1]}\right)\middle|\boldsymbol{Z}_{1:T}, \Psi^{(it-1)}\right]$. In the (it)th iteration, the E-step estimates the sequence of $\{\boldsymbol{\beta}_t\}_{t=1}^{T}$ using the Kalman filter/smoother. Then the CM-steps of cycle 1 are run to obtain the estimation of parameters in $\Psi^{[1]} = \{\boldsymbol{H}, \boldsymbol{M}\}$, which maximizes $\mathcal{Q}^{[1]}\left(\Psi^{[1]};\, \Psi^{(it-1)}\right)$. They both have closed form solutions for their MLEs,

$$\boldsymbol{M}^{(it)} = \boldsymbol{V}_{10}\boldsymbol{V}_{00}^{-1} \tag{4}$$

$$\boldsymbol{H}^{(it)} = \frac{1}{T}\left(\boldsymbol{V}_{11} - \boldsymbol{V}_{10}\left(\boldsymbol{M}^{(it)}\right)^{\top} - \boldsymbol{M}^{(it)}\boldsymbol{V}_{10}^{\top} + \boldsymbol{M}^{(it)}\boldsymbol{V}_{00}\left(\boldsymbol{M}^{(it)}\right)^{\top}\right), \tag{5}$$

where the matrices are $\boldsymbol{V}_{11} = \sum_{t=1}^{T}\left(\boldsymbol{B}_{t|T} + \boldsymbol{\beta}_{t|T}\boldsymbol{\beta}_{t|T}^{\top}\right)$, $\boldsymbol{V}_{00} = \sum_{t=1}^{T}\left(\boldsymbol{B}_{t-1|T} + \boldsymbol{\beta}_{t-1|T}\boldsymbol{\beta}_{t-1|T}^{\top}\right)$ and $\boldsymbol{V}_{10} = \sum_{t=1}^{T}\left(\boldsymbol{B}_{t,t-1|T} + \boldsymbol{\beta}_{t|T}\boldsymbol{\beta}_{t-1|T}^{\top}\right)$, following Shumway and Stoffer (2006). Note that $\boldsymbol{\beta}_{t|T}$ and $\boldsymbol{B}_{t|T}$ are the Kalman smoothed version of $\boldsymbol{\beta}_t$ and its corresponding variance matrix $\boldsymbol{B}_t$. In addition, for a diagonal structured VAR coefficient matrix $\boldsymbol{M} = \mathrm{diag}\{m_1, \ldots, m_K\}$, the MLE can be derived as $(m_1^{(it)}, \ldots, m_K^{(it)})^{\top} = \boldsymbol{L}^{-1}\boldsymbol{b}$, where $\boldsymbol{L} = (l_{ij}) = (\mathrm{tr}\{\boldsymbol{H}^{-1}\frac{\partial \boldsymbol{M}}{\partial m_j}\boldsymbol{V}_{00}\frac{\partial \boldsymbol{M}}{\partial m_i}^{\top}\})$, $\boldsymbol{b} = (b_i) = (\mathrm{tr}\{\boldsymbol{H}^{-1}\frac{\partial \boldsymbol{M}}{\partial m_i}\boldsymbol{V}_{10}^{\top}\})$, following the approach in Xu and Wikle (2007).

The algorithm then moves on to cycle 2, where the observed data are $\boldsymbol{Z}_{obs} = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_T\}$ and the missing data are $\boldsymbol{Z}_{mis} = \{\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_T\,;\, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T\}$. Assuming independence between $\{\boldsymbol{\alpha}_t\}_{t=1}^{T}$ and $\{\boldsymbol{\beta}_t\}_{t=1}^{T}$, the complete data distribution in this cycle is

$$f(\boldsymbol{Z}_{1:T}, \boldsymbol{\beta}_{0:T}, \boldsymbol{\alpha}_{1:T}; \Psi) = \prod_{t=1}^{T} f(\boldsymbol{Z}_t|\boldsymbol{\beta}_t, \boldsymbol{\alpha}_t; \Psi) f(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}; \Psi) f(\boldsymbol{\alpha}_t; \Psi) f(\boldsymbol{\beta}_0; \Psi) \tag{6}$$

This gives the likelihood function and hence the target function $\mathcal{Q}^{[2]}\left(\Psi^{[2]}; \Psi^{(it,it-1)}\right)$ for maximisation. Note that the superscript of the parameter set is $(it, it-1)$, emphasizing that part of its elements have already been updated in cycle 1. The E-step of cycle 2 estimates the sequence of $\{\boldsymbol{\beta}_t\}_{t=1}^{T}$ with the up-to-date $\boldsymbol{H}^{(it)}$ and $\boldsymbol{M}^{(it)}$ and the sequence of $\{\boldsymbol{\alpha}_t\}_{t=1}^{T}$ using the estimation algorithm of the FPC model. The CM-step computes the MLEs of the elements in parameter set $\Psi^{[2]} = \{\boldsymbol{\Theta}, \boldsymbol{\Lambda}, \sigma^2\}$. In particular, the coefficient matrix $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_P)$ is estimated column by column and the covariance matrix $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_P\}$ is estimated element by element.

$$\begin{aligned}\sigma^{2(it)} = \frac{1}{N}\sum_{t=1}^{T}\Big[&\mathrm{tr}\Big\{\boldsymbol{\Phi}_t\boldsymbol{B}_{t|T}\boldsymbol{\Phi}_t^{\top} + \left(\boldsymbol{Z}_t - \boldsymbol{\Phi}_t\boldsymbol{\beta}_{t|T}\right)\left(\boldsymbol{Z}_t - \boldsymbol{\Phi}_t\boldsymbol{\beta}_{t|T}\right)^{\top}\Big\} \\ &+ \mathrm{tr}\left\{\boldsymbol{\Phi}_t\boldsymbol{\Theta}\widehat{\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^{\top}}\boldsymbol{\Theta}^{\top}\boldsymbol{\Phi}_t^{\top}\right\} \\ &- 2\,\mathrm{tr}\left\{\boldsymbol{\Phi}_t\boldsymbol{\Theta}\hat{\boldsymbol{\alpha}}_t\boldsymbol{Z}_t^{\top} - \boldsymbol{\Phi}_t\boldsymbol{\Theta}\widehat{\boldsymbol{\alpha}_t\boldsymbol{\beta}_t^{\top}}\boldsymbol{\Phi}_t^{\top}\right\}\Big]\end{aligned} \tag{7}$$

where $N = \sum_{t=1}^{T} n_t$ is the sum of the number of observations at each time point t, and for $p = 1, \ldots, P$,

$$\lambda_p^{(it)} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^{\top}}_{(p,p)}, \tag{8}$$

$$\boldsymbol{\theta}_p^{(it)} = \left[\sum_{t=1}^{T}\widehat{\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^{\top}}_{(p,p)}\boldsymbol{\Phi}_t^{\top}\boldsymbol{\Phi}_t\right]^{-1}\sum_{t=1}^{T}\boldsymbol{\Phi}_t^{\top}\left[\hat{\boldsymbol{\alpha}}_{t(p)}\boldsymbol{Z}_t - \boldsymbol{\Phi}_t\left(\widehat{\boldsymbol{\alpha}_t\boldsymbol{\beta}_t^{\top}}_{(p,\cdot)}\right)^{\top} - \sum_{j\neq p}\widehat{\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^{\top}}_{(p,j)}\boldsymbol{\Phi}_t\hat{\boldsymbol{\theta}}_j\right], \tag{9}$$

where $\hat{\boldsymbol{\alpha}}_{t(p)}$ is the pth element of vector $\mathbf{E}\left[\boldsymbol{\alpha}_t\middle|\boldsymbol{Z}_{1:T}, \Psi^{(it,it-1)}\right]$, $\widehat{\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^{\top}}_{(p,j)}$ is the (p, j)th element of $\mathbf{E}\left[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^{\top}\middle|\boldsymbol{Z}_{1:T}, \Psi^{(it,it-1)}\right]$ and $\widehat{\boldsymbol{\alpha}_t\boldsymbol{\beta}_t^{\top}}_{(p,\cdot)}$ represents the pth row of $\mathbf{E}\left[\boldsymbol{\alpha}_t\boldsymbol{\beta}_t^{\top}\middle|\boldsymbol{Z}_{1:T}, \Psi^{(it,it-1)}\right]$. Note that $\boldsymbol{\theta}_p^{(it)}$ is updated sequentially with $\hat{\boldsymbol{\theta}}_j = \boldsymbol{\theta}_j^{(it)}$ for $j < p$ and $\hat{\boldsymbol{\theta}}_j = \boldsymbol{\theta}_j^{(it-1)}$ for $j > p$. Full details of the algorithm are provided in section 2 of the supplement.

After running through cycle 1 and cycle 2, the parameter set is updated to $\Psi^{(it)}$, completing one iteration of the AECM algorithm. The iteration stops when the relative change of the log-likelihood is smaller than a threshold. The convergence of some crucial parameters can be used to evaluate the convergence of the algorithm.

Last but not least, the design of the 2-cycle AECM algorithm, including partition of the parameter space and the data augmentation, also considers the asymptotic properties of the algorithm. It is widely acknowledged that a generalized EM algorithm will converge to a stationary point (even if not the maximum) (McLachlan and Krishnan 1997). To ensure the convergence of the conditional M-steps in an ECM algorithm, an additional condition called the 'space-filling' condition is required. It was introduced in Meng and Rubin (1993) and its extension to the AECM algorithm was made in Meng and Van Dyk (1997). The algorithm

described above uses a partition of the parameter space that satisfies the 'space-filling' condition. Note that such partition is not unique. The advantage of this particular design is that all parameters have analytical solutions of their MLEs, i.e. no (high-dimensional) numerical optimization is required.

2.3.1 Initialization

The initialization of the algorithm follows the common approaches used in the initialization of the state space model and the FPC model.

1. The initial state $\boldsymbol{\beta}_0$ follows a normal distribution, $\boldsymbol{\beta}_0 \sim \mathcal{N}(\boldsymbol{0},\ \tau^2\boldsymbol{I})$, where τ^2 is set to a relatively large number to reflect the lack of knowledge of the initial situation, e.g. $\tau^2 = 100$. The initial value of $\boldsymbol{\beta}$ is computed through fitting the linear regression model $\boldsymbol{Z} = \boldsymbol{\Phi\beta}$ using vectorized data, $\boldsymbol{Z} = \mathbf{vec}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_T)$, and is denoted as $\boldsymbol{\beta}^{(0)}$.
2. The initial value of $\boldsymbol{M}$ is taken to be $\boldsymbol{M} = \mathbf{I}$. The initial value of the covariance matrix of the state transition equation $\boldsymbol{H}$ is initialized as $\sigma_h^2\boldsymbol{I}$, where $\sigma_h^2 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Var}[Z_t(x_i, y_i)]$. Some other values of σ_h^2 may be used depending on the features of the data.
3. The sum of the residuals and the FPCA component is then calculated as $\hat{\boldsymbol{r}}_t = \boldsymbol{\Phi}_t\boldsymbol{\Theta\alpha}_t + \boldsymbol{\epsilon}_t = \boldsymbol{Z}_t - \boldsymbol{\Phi}_t\boldsymbol{\beta}^{(0)}$. Rewriting $\boldsymbol{\Phi}_t\boldsymbol{\Theta\alpha}_t$ as $\boldsymbol{\Phi}_t\boldsymbol{\eta}_t$ and fitting the model $\hat{\boldsymbol{r}}_t = \boldsymbol{\Phi}_t\boldsymbol{\eta}_t + \boldsymbol{\epsilon}_t$ gives the least square estimate $\hat{\boldsymbol{\eta}}_t = (\boldsymbol{\Phi}_t^\top\boldsymbol{\Phi}_t)^{-1}\boldsymbol{\Phi}_t^\top\hat{\boldsymbol{r}}_t$. Apply the eigenvalue decomposition $\mathbf{Cov}[\hat{\boldsymbol{\eta}}_t] = \boldsymbol{U\Sigma}_\alpha\boldsymbol{U}^\top$. The initial value of $\boldsymbol{\Theta}$ can be obtained as $\boldsymbol{\Theta}^{(0)} = \boldsymbol{U}$.
4. Finally, set the initial value of $\boldsymbol{\Lambda}$ to be $\boldsymbol{\Sigma}_\alpha$ and the initial value of σ^2 to be $\frac{1}{N}\sum_{t=1}^{T}\hat{\boldsymbol{r}}_t^\top\hat{\boldsymbol{r}}_t$.

There are two more parameters to select before implementing the AECM algorithm, the spatial basis dimension K and the number of the functional PCs P. The following two stage method is proposed here to avoid selection using cross validation (which is computationally costly).

1. First choose the basis dimension K. The selection uses the information criteria, such as AIC and BIC, and can be based on (1) the functional data representation or (2) the SS-FPCA model. The advantage of using the functional data representation is computational efficiency. The advantage of using the SS-FPCA is that it involves a comprehensive consideration of the dynamic and the FPCA components. However, the computation time would be much longer and therefore might not be practical in some situations.
2. Then the selection of the expansion order P follows. There are also two possible approaches, (1) fitting a series of models with increasing expansion orders and terminating at the expansion order where the variances of the remaining PCs are negligible as in Zhou and Pan (2014), (2) fitting a high rank or full rank model, then choosing the expansion order based on the magnitudes of the variances of the PCs. It is also possible to use AIC, BIC to select the expansion order. This might serve the interpolation purpose better; whereas the variance proportion criterion may suit the interpretation purpose better. Note that applying the information criteria can be computationally more intense as models with higher P are required in order to make the selection. A looser convergence criterion may be used during the selection to reduce computational cost.

The choice of K and P may also be made based on the background of the application, if relevant information is available, such as the level of smoothness of the process by nature.

2.4 Estimate the standard errors of model parameters

A potential drawback of the EM-type algorithm is that there is no straightforward solution to the standard errors of the estimated parameters. To carry out statistical inference, resampling methods, such as bootstrap, are often used. Implementations can be found in Rice and Wu (2001), Zhou and Pan (2014) and Fassò and Cameletti (2009). This approach, though computationally intensive for a complex model, is relatively straightforward to apply. Alternatively, the inverse of the Fisher information matrix $\mathcal{I}(\Psi)^{-1}$ can be used to estimate the standard errors of the parameters Ψ (McLachlan and Krishnan 1997). However, approximation of the Fisher information matrix of the (observed) incomplete data is required (Louis 1982; McLachlan and Krishnan 1997) and it is not easy for dependent data. The relatively large number of unknown parameters can also make the inversion of the Fisher information matrix problematic. Cressie et al. (2010) and Katzfuss and Cressie (2011) derived the mean squared prediction errors (MSPE) for their spatio-temporal predictors. In situations where the predicted values are of interest, the MSPE can provide a computationally efficient way to quantify the uncertainty. Unfortunately, this is not easy for the SS-FPCA model. Due to the unknown parameters in the observation equation (1), approximations are required to compute the MSPE. For a predictor of the true process $\widehat{\boldsymbol{Y}}_t(\boldsymbol{\Theta}) = \mathbf{E}[\boldsymbol{\Phi}_t\boldsymbol{\beta}_t + \boldsymbol{\Phi}_t\boldsymbol{\Theta\alpha}_t \,|\, \boldsymbol{Z}_{1:T}]$ with known parameter $\boldsymbol{\Theta}$, the MSPE has a simple expression,

$$m(\boldsymbol{\Theta}) = \mathbf{E}\left[\left(\widehat{\boldsymbol{Y}}_t(\boldsymbol{\Theta}) - \boldsymbol{Y}_t\right)^2\right] = \boldsymbol{\Phi}\boldsymbol{B}_{t|T}\boldsymbol{\Phi}^\top + \boldsymbol{\Phi}\boldsymbol{\Theta}\boldsymbol{A}_{t|T}\boldsymbol{\Theta}^\top\boldsymbol{\Phi}^\top,$$

where $\boldsymbol{A}_{t|T} = \mathbf{Var}[\boldsymbol{\alpha}_t \,|\, \boldsymbol{Z}_{1:T}] = \widehat{\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\top} - \hat{\boldsymbol{\alpha}}_t\hat{\boldsymbol{\alpha}}_t^\top$. The problem is, $\boldsymbol{\Theta}$ is unknown in practice and plugging the estimated $\widehat{\boldsymbol{\Theta}}$ in $m(\boldsymbol{\Theta})$ underestimates the MSPE. To solve this problem, Zimmerman and Cressie (1992) proposed an adjustment method, where the MSPE is approximated by

$$\text{MSPE} \approx m(\boldsymbol{\Theta}) + \mathbf{trace}\{C(\boldsymbol{\Theta})D(\boldsymbol{\Theta})\}, \tag{10}$$

where $C(\boldsymbol{\Theta}) = \mathbf{Var}[\partial\widehat{\boldsymbol{Y}}_t(\boldsymbol{\Theta})/\partial\boldsymbol{\Theta}]$ and $D(\boldsymbol{\Theta}) = \mathbf{E}[(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^\top]$. For the SS-FPCA model, however, $D(\boldsymbol{\Theta})$ requires another approximation using the inverse Fisher information matrix and its performance is yet to be investigated. Due to this concern, this paper will not consider the MSPE, but the spatiotemporal bootstrap to quantify the uncertainty.

There are various methods to bootstrap the temporally or spatially dependent data, as documented in Lahiri (2003). Resampling method based on the innovation sequence is frequently mentioned in literature for state space models (Stoffer and Wall 1991; Shumway and Stoffer 2006; Katzfuss and Cressie 2011). For a spatiotemporal process modelled as a state space model, however, it is important to ensure that both the spatial and temporal components are sampled appropriately. This paper follows the approach of Fassò and Cameletti (2009) and implements a parametric bootstrap procedure tailored to the SS-FPCA model. Specifically, the spatial and temporal components are generated separately, based on their own covariance structures, giving the bootstrap sample

$$\boldsymbol{\beta}_t^* = \boldsymbol{M}\boldsymbol{\beta}_{t-1}^* + \boldsymbol{u}_t^* \tag{11}$$

$$\boldsymbol{Z}_t^* = \boldsymbol{\Phi}\boldsymbol{\beta}_t^* + \boldsymbol{\xi}_t^* + \boldsymbol{\epsilon}_t^*, \tag{12}$$

for $t = 1, \ldots, T$. The distributions associated with the residual components $\boldsymbol{u}_t^*$, $\boldsymbol{\epsilon}_t^*$ and the spatial random effects $\boldsymbol{\xi}_t^*$ are $\boldsymbol{u}_t^* \sim \mathcal{N}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}}_u)$, $\boldsymbol{\xi}_t^* \sim \mathcal{N}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}}_\xi)$ and $\boldsymbol{\epsilon}_t^* \sim \mathcal{N}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}}_\epsilon)$, where $\widehat{\boldsymbol{\Sigma}}_u = \widehat{\boldsymbol{H}}$, $\widehat{\boldsymbol{\Sigma}}_\xi = \boldsymbol{\Phi}\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Theta}}^\top\boldsymbol{\Phi}^\top$ and $\widehat{\boldsymbol{\Sigma}}_\epsilon = \hat{\sigma}^2\mathbf{I}$, from the estimated model using the original data. To carry out the inference, generate a large number of bootstrap samples using the above distributions and then estimate the standard errors of the estimated parameters from the estimated parameters of the bootstrap samples. The corresponding confidence intervals can be constructed using percentiles or as normal intervals (i.e. based on the bootstrap variance estimates). A percentile interval of the interpolatons can be constructed directly from the bootstrap samples. The approximated MSPE in formula (10) may be computed by plugging in the bootstrap variances as the elements in $D(\boldsymbol{\Theta})$. Although it still requires much computational effort to calculate the elements in $C(\boldsymbol{\Theta})$.

3 Simulation

Before the SS-FPCA model is applied to the remote sensing Chl data, a simulation study is carried out to investigate if the model, estimated using the proposed 2-cycle AECM algorithm, can identify the temporal and spatial structure in the data. For computational efficiency and better visualisation of the results, the simulation was conducted on a 1-dimensional space. This, though different from the data under study, would not result in loss of generality as the model assumptions and the estimation method remain the same. A similar study on the STRE model using 1-dimensional spatial data was conducted in Katzfuss and Cressie (2011).

3.1 Simulation design

A few key aspects of the simulation design are listed below.

1. The dimension of the simulated data is 50×100, where $n = 50$ is the number of observations (indexed by s) at each time point in the 1-dimensional space $\mathcal{D} = [1.1521, 1.5661]$ and $T = 100$ is the total number of time points (indexed by t). The function argument s represents the spatial location in the 1-dimensional space. This means, the data are $Z_{t,s}$, $s \in \mathcal{D}$, for $s = 1, \ldots, 50$ and $t = 1, \ldots, 100$.
2. A univariate cubic B-spline basis with 3 equally spaced interior knots is used. This gives the basis dimension of $K = 3 + 3 + 1 = 7$. The basis coefficient vector series $\boldsymbol{\beta}_t = (\beta_{1t} \ldots \beta_{Kt})^\top$, $t = 1, \ldots, T$, are generated using $K = 7$ random walk processes $\{u_{kt}\}_{t=1}^T$, $k = 1, \ldots, 7$, each with distribution $u_{kt} \sim \mathcal{N}(0, h_k)$, and a random zero mean starting point. This means $\boldsymbol{M} = \mathbf{I}$ for the state transition equation. Specifically,

   $$\{h_1, \ldots, h_7\} = \{0.33,\ 0.25,\ 0.42,\ 0.25,\ 0.27,\ 0.62,\ 0.28\}.$$

   This gives the dynamic component (using matrix notation) $\boldsymbol{Z}_t^{(d)} = \boldsymbol{\Phi}\boldsymbol{\beta}_t$.
3. Apply functional PCA to a subset of the remote sensing Chl data of Lake Victoria,[3] to extract the eigenfunctions and eigenvalues. A valid covariance matrix $\boldsymbol{\Sigma}_{chl}$ can be constructed using the leading two eigenfunctions, $\xi_1(s)$, $\xi_2(s)$, and the corresponding eigenvalues, $\lambda_1 = 9.64$, $\lambda_2 = 1.80$. The FPCA component are then generated as $\boldsymbol{Z}_t^{(s)} = \boldsymbol{\Sigma}_{chl}^{\frac{1}{2}}\boldsymbol{Y}_t$, where $\boldsymbol{Y}_t =$

[3] Lake Victoria is a large lake in Africa. The Lake Victoria Chl data are part of the MERIS Chlorophyll-*a* product processed by Plymouth Marine Laboratory. There are near complete observations in some areas of the lake, and hence are used here to generate artificial data that resemble the real data.

$(Y_{t,1}, \ldots, Y_{t,50})^\top$ is a sequence of random realizations from the $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ distribution. Finally, multiply the FPCA component $\boldsymbol{Z}_t^{(s)}$ with a factor κ ($\kappa \geq 1$) to control the strength of the spatial signal. In this simulation study, $\kappa = 1.25$ and $\kappa = 1.5$ are considered.

4. The error component $\boldsymbol{\epsilon}_t$ is generated from the normal distribution $\mathcal{N}(\boldsymbol{0}, \sigma^2 \mathbf{I})$. In this simulation study $\sigma^2 = 0.01$ and $\sigma^2 = 0.25$ are considered.
5. The dynamic, FPCA and error components are then combined to obtain the simulated spatio-temporal data, as $\boldsymbol{Z}_t = \boldsymbol{Z}_t^{(d)} + \kappa \boldsymbol{Z}_t^{(s)} + \boldsymbol{\epsilon}_t$.
6. To create the missing data, first generate a series of 100 missing proportions p_t, $t = 1, \ldots, 100$, from the uniform $\mathcal{U}(0, 1)$ distribution. Then generate 50 binomial random variables from distribution $\mathcal{B}(1, p_t)$ for each t. Regard the observations at the locations corresponding to 0 as missing. This gives a much higher missing proportion than the Taruo Lake Chl time series, but such level of missing observations is not uncommon in the retrieved data of other lakes from MERIS.

The spatial missing patterns are not considered in this simulation in order not to over-complicate the simulation design. In the application for this paper, missing pixels/images in the remotely sensed Chl data are identified to be due to exogenous causes. Hence, such patterns should not overly impact parameter estimation. In a simulation study on the FPC model applied to remote sensing lake surface water temperature data in Gong (2017), spatial patterns in missing data were investigated, along with different levels of missing proportions. The results suggested that the spatial missing pattern itself did not appear to have a strong impact on the fitted model. The interaction of high missing proportions and missing patterns tend to be more influential.

An additional factor of interest is the initialization method of the 2-cycle AECM. Two initialization methods are considered, the standard method as described in Sect. 2.3.1 and a separate method which uses $\boldsymbol{Z}_t^{(d)} + \boldsymbol{\epsilon}_t$ to initialize the dynamic component and $\boldsymbol{Z}_t^{(s)} + \boldsymbol{\epsilon}_t$ to initialize the FPCA component. The separate method, though only plausible in simulation, is supposed to provide initial values with higher precision. It is interesting to see if there would be any distinctive difference between the results using two different initialization methods.

The initialization method and the strength of the spatial signals are matched to create three combinations, weak + standard, weak + separate and strong + standard. For each combination, four different situations based on noise levels (small or large) and missing conditions (complete or missing) are created, giving 12 scenarios in total (see Fig. 3).

3.2 Simulation results

500 replicates were run for each scenario. The computation times varies, depending on the number of iterations involved (from less than 10 to 500). On average, one iteration took 0.25–0.5 s on a standard Windows desktop computer.

The fitted models are capable of recovering the patterns of the dynamic and the FPCA components (up to a sign change of the eigenfunctions). The top three panels in Fig. 4 represent the estimated first eigenfunction $\xi_1(s)$ from scenarios S4, S8 and S12, with the true eigenfunction plotted as the red curve. S4, S8 and S12 represent three simulation scenarios labeled as weak + standard, weak + separate, strong + standard, each paired with large noise and missing observations. Clearly, the pattern in $\xi_1(s)$ is very well identified. All 500 replicates produced curves bearing the feature of the true eigenfunction. The situation with the second eigenfunction is slightly worse, with occasional miss of the target, especially when the spatial signal is weak (see bottom left panel of Fig. 4). However, considering that the magnitude of the variance of the first PC is more than five times of the variance of the second PC, it is not surprising that the pattern in $\xi_2(s)$ is harder to capture.

The patterns in the time series of the coefficient vector were also captured by the smoothed series $\{\boldsymbol{\beta}_{t|T}\}_{t=1}^T$. Figure 5 gives an example of the smoothed series of each component of $\boldsymbol{\beta}_t$, β_{kt}, $k = 1, \ldots, 7$, taken from scenario S1 (the weak + standard, paired with small noise and complete data scenario). It is straightforward to see that the smoothed series (dark grey curves) track the true simulated series (red curves) in the majority of the cases. The figure also shows a relatively large difference in the variations, with the 3rd component having the largest variation and the 7th component varying the least. This result could be attributed to the feature of the data, where the variation

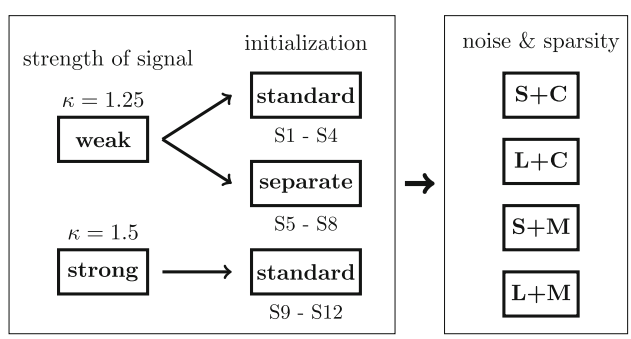

Fig. 3 A diagram showing the settings of 12 simulation scenarios. The abbreviations used in the diagram are *S* small noise, *L* large noise, *C* complete data and *M* missing data

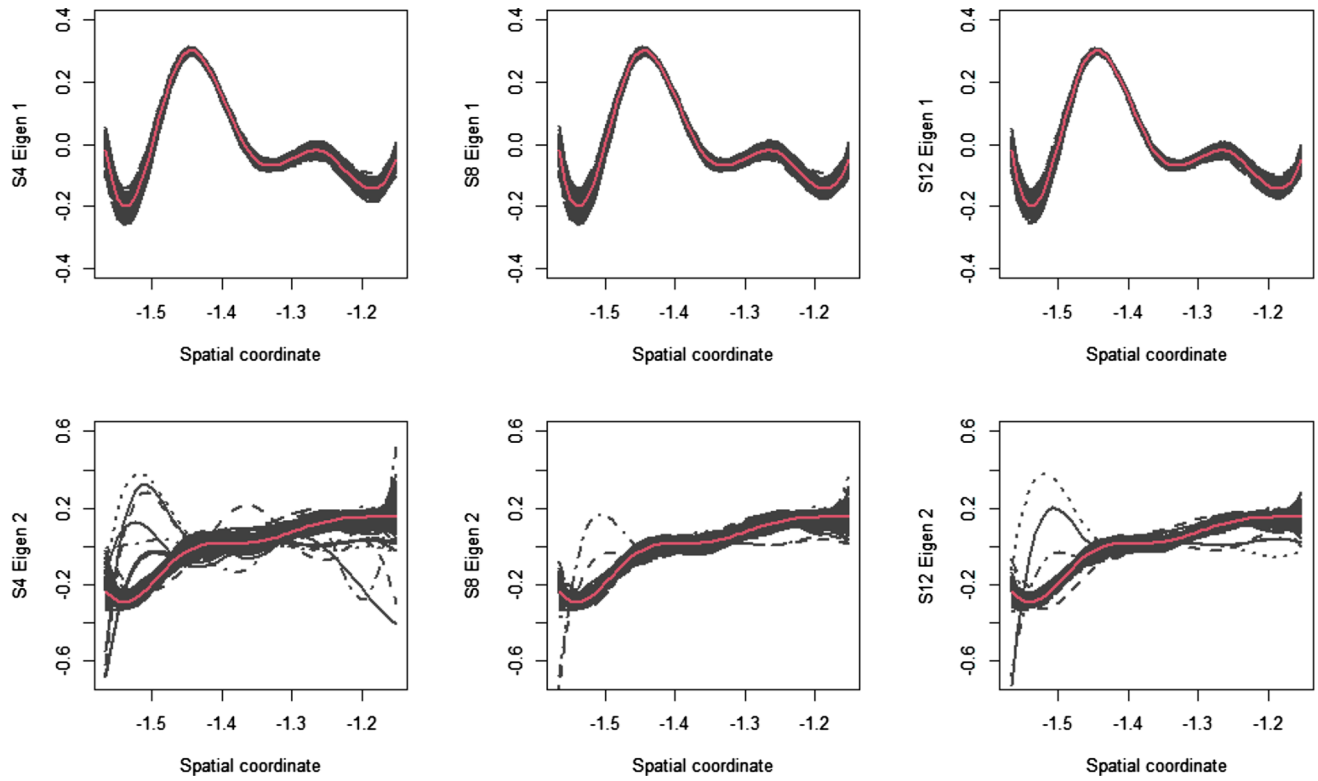

Fig. 4 (Top) The estimated (black curves) and the true (red curve) values of eigenfunction of PC1, from scenario S4, S8 and S12. (Bottom) The estimated (black curves) and the true (red curve) values of eigenfunction of PC2, from scenarios S4, S8 and S12

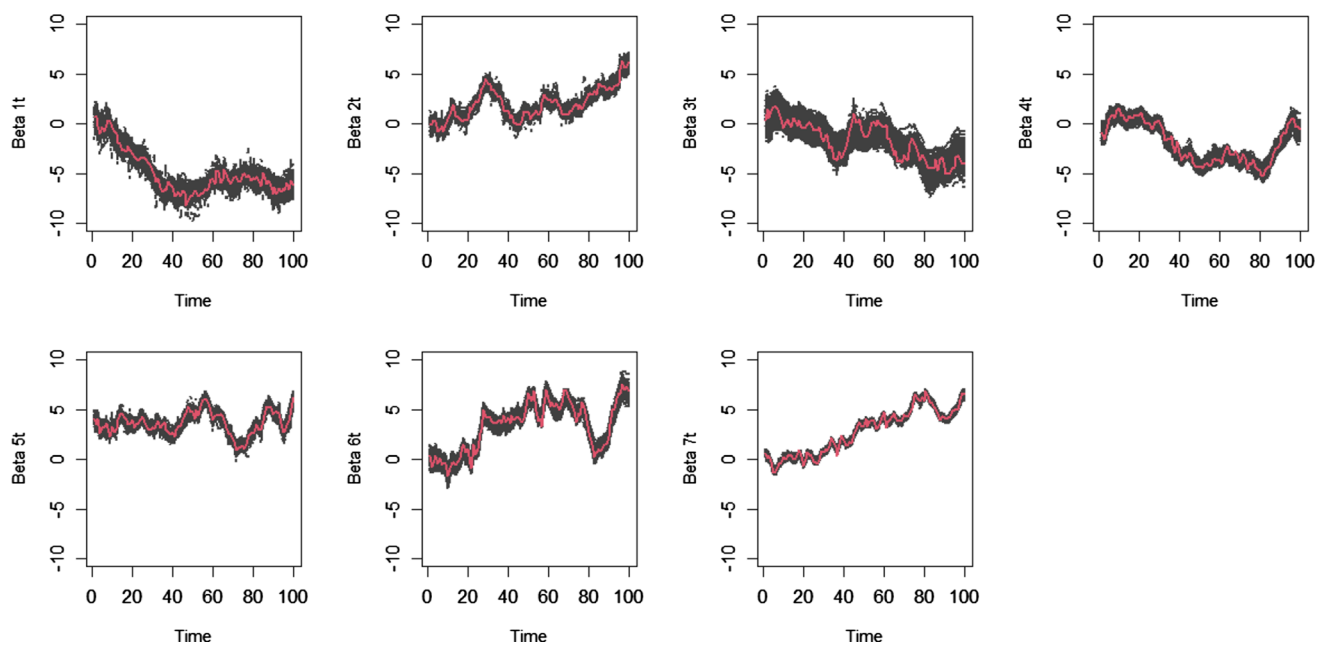

Fig. 5 The Kalman smoothed $\{\boldsymbol{\beta}_{t|T}\}_{t=1}^{T}$ from scenario S1. From left to right, top to bottom are the smoothed $\beta_{kt|T}$, $k = 1, \ldots, 7$ curves (black) and the true curves (red)

is larger in the range of support of the third basis function $\phi_3(s)$ in $\Phi(s) = (\phi_1(s), \ldots, \phi_7(s))^\top$.

The estimation of the variances of three model components appears to be more difficult. There is an underestimation of λ_1, λ_2 and an overestimation of σ^2. The

estimation of h_k, $k = 1, \ldots, 7$, also tends to be biased in some situations, but the pattern in the scales is captured by the estimated values. The increase in the strength of the spatial signal and the more precise separate initialization method did not have a big influence on the estimation. However, the separate initialization method appears to have the potential of avoiding extreme results in the estimation of h_k. In addition, introducing sparsity to the data did not make a big difference in the residual sum of squares (RSS), and the RSS values are consistent with the true variance of the model residuals. Figures and tables showing the estimated variance parameters and the RSS of the interpolations for all 12 simulation scenarios can be found in section 3 of the supplement.

In general, this simulation study showed that the SS-FPCA model was capable of identifying the spatio-temporal patterns in the data. Particularly, the estimated model captured the temporal evolution of the basis coefficients and identified the eigenfunction in majority of the replicates. The SS-FPCA model appeared to lack precision in estimating the variance components. Strengthening spatial signal and changing initialization method did not improve the results in this case. This can be attributed to the spatial confounding between various model components, which is common to many spatial and spatio-temporal models. Discussion on this topic can be found in Reich and Hodges (2008), Hodges and Reich (2010) and Hughes and Haran (2013). Solutions to the problem of the SS-FPCA model can be quite challenging. Examining the scales and features of the variation of the data may be of help. Estimating and then fixing some parameters before launching the AECM algorithm may also be sensible, such as in the estimation of the spatiotemporal fixed rank kriging model in Zammit-Mangion and Cressie (2020). In addition, running the algorithm multiple times using different initial values can be a choice. Although the simulation study and the test carried out in Gong (2017) suggested that the algorithm appears to be quite robust to initialization of the algorithm.

4 Application

The monthly Chl (mg/m^3) time series from June 2002 to April 2012 for Taruo Lake were investigated here. A total of 19 images at the beginning and the end of the observing period were excluded from the modelling due to limited data availability. This gives 100 monthly images in the data set, with 15.03% data missing. The data were first centered by subtracting a spatial mean at each time point to remove the seasonal pattern. As satellite retrievals of the lake border pixels are often associated with higher uncertainty (Rodgers 1990), the lake was trimmed to retain a rectangular grid of size 103×45, which covers the main body of the lake. The resulting images were modelled using SS-FPCA. Specifically, a VAR(1) process was used in the state transition equation and the coefficient matrix $\boldsymbol{M}$ was set to be diagonal. This specification is reasonable to account for temporal correlation in the centered Chl image time series.

The orthogonal spatial basis $\boldsymbol{\Phi}$ was created based on a tensor spline basis (Wood 2017) with equally spaced knots. For a lake of more irregular shape, bivariate basis on triangular mesh might be required (Guillas and Lai 2010; Ettinger et al. 2012). The basis dimension K and the number of PCs P were chosen using the two stage method described in Sect. 2.3.1. The appropriate basis according to the AIC was of dimension $K = 8 \times 6$ (4 knots along longitude and 2 knots along latitude). The variance proportion criterion of $\geq 80\%$ suggested $P = 6$. The standard procedure in Sect. 2.3.1 was used to initialize the model parameters. In addition, a Kalman filtering threshold of 90% was applied to avoid the overfitting of a few very sparse images. That is, any images with less than 10% of observations available were not filtered, but smoothed based on the information from their neighbouring images. The AECM algorithm converged after 6 iterations, under the criterion of the change of the log-likelihood being $\leq 0.05\%$.

A total of 6 functional PCs estimated from the model, of which three have greater than 10% contributions to the variation explained by the FPCA component. Their corresponding eigenvalues are $\hat{\lambda}_1 = 356.18$, $\hat{\lambda}_2 = 254.21$ and $\hat{\lambda}_3 = 170.95$. Figure 6 shows the plot of the eigenfunctions and the corresponding scores of the leading three functional PCs. The 1st PC explains 35.4% of the variations in the FPCA component. It displays the contrast between the west and the east parts of the lake (top left of Fig. 6), specifically, a contour of high values is evident around 84.05–84.10 longitude and 31.08–31.10 latitude, which could indicate the potential for an algal bloom and is perhaps the most distinctive spatial pattern in the Chl anomalies (recall that the data were centered) of Taruo Lake. The PC scores and the comparison to original data indicate that this feature mainly occurs in July, August between 2006 and 2008 (see Fig. 1 for original data) and February, March in 2008. The proportion of variation explained by the 2nd PC is 25.2%. It has a distinctive northeast corner (top middle of Fig. 6) of low values in contrast to the rest of the lake. The dominant feature in the scores of the 2nd PC indicates an image in February 2007 when the remote sensing values are very different from the other images (bottom left of Fig. 1), which highlights the potential of the method to automatically detect images that require closer inspection for e.g. unusual event, retrieval

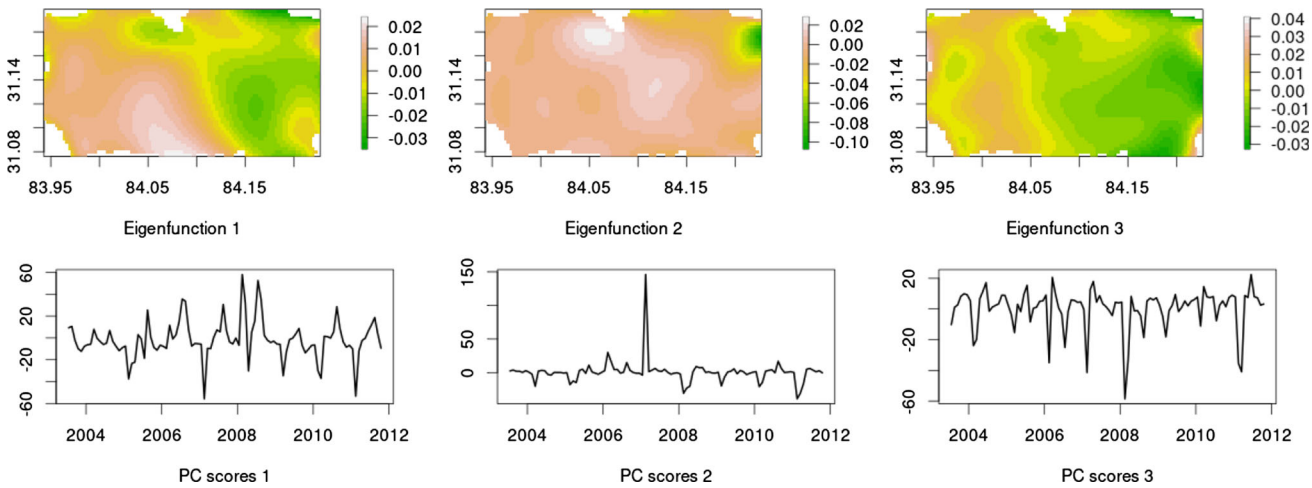

Fig. 6 The plots of the leading 3 eigenfunctions (top) and the corresponding scores (bottom). The horizontal and vertical axes of the eigenimages represent longitude and latitude respectively. The horizontal axis of the score plots represents the index of time points

algorithm problem. The 3rd PC explains 17.0% of the variation. It shows another contrasting pattern between the east and west (top right of Fig. 6). Traces of this pattern can still be seen in original data, but it is less distinctive than the pattern displayed in the 1st PC. The 3rd PC appears to have low scores for images in February for 2006, 2007, 2008 and 2011, when there are very few data initially (bottom right of Fig. 1).

The estimated VAR(1) coefficient matrix $\widehat{\boldsymbol{M}}$ has the norm of the determinant less than 1, suggesting that the temporal evolution of the spatial process is stationary. The smoothed dynamic component $\boldsymbol{\beta}_{t|T}$ provides additional information on the system dynamics. However, as they are spatial basis coefficients, it would be more helpful to look at the evolution of the product $\boldsymbol{\Phi}\boldsymbol{\beta}_{t|T}$, rather than the coefficient time series alone (see section 4 of supplement for an example).

Interpolations can be produced using the estimated SS-FPCA model. Figure 7 presents the observed (left column) and the interpolated (right column) log transformed chlorophyll-*a* data from July and February 2006, which corresponds to two of the images presented in Fig. 1. In summary, the mean squared errors (MSE) of the interpolations across space and time is 0.0889, as compared to 0.1054 from the interpolations using the FPC model with the same expansion order $P = 6$ (see section 4 of the supplement for more information of the FPC model applied to the Taruo Lake Chl data). This indicates another benefit of accounting for the temporal correlations in the data. The residual sum of squares (RSS) over time appeared to be associated with data availability. The values vary from close to 0 at well observed time points to around 0.7 at time points with very sparse observations. The RSS over space are of the similar scale in majority of the pixels, apart from a few pixels around an island in the lake, where large residuals are found. This is consistent with the higher uncertainties in the satellite retrievals in these pixels. Related figures are given in section 4 of the supplement. There appear to be a few values outside the range of the observed data in the interpolations of the very sparse images, such as the bottom left image in Fig. 7. Similar issues (very high or low values) were found in the interpolations produced using the FPC model. It is hard to tell whether these values over or under estimate the real situation, as observations in those pixels are missing. However, it can be helpful to carry out some investigations, especially if the interpolations are to be used in further analysis.

Finally, the standard errors of the estimated parameters were produced using the spatiotemporal bootstrap described in Sect. 2.4. 200 bootstrap samples were generated using the estimated parameters above. The resulting confidence intervals of the VAR(1) coefficients $\hat{m}_1, \ldots, \hat{m}_{48}$, the variance of the PCs $\hat{\lambda}_1, \ldots, \hat{\lambda}_6$, and the PC coefficient vectors $\hat{\boldsymbol{\theta}}_p$, $p = 1, \ldots, 6$, are presented in section 4 of the supplement.

5 Discussion

The SS-FPCA model proposed in this paper provides a way of investigating the spatiotemporal patterns in a time series of spatial images with missing observations. Particularly, the estimated coefficients of the dynamic component, the functional principal components and the PC scores can be used to identify the dominant spatial/temporal patterns in the data. This information can be crucial to the investigation of the important driving forces of an environmental process in space and time, which is of great interest to

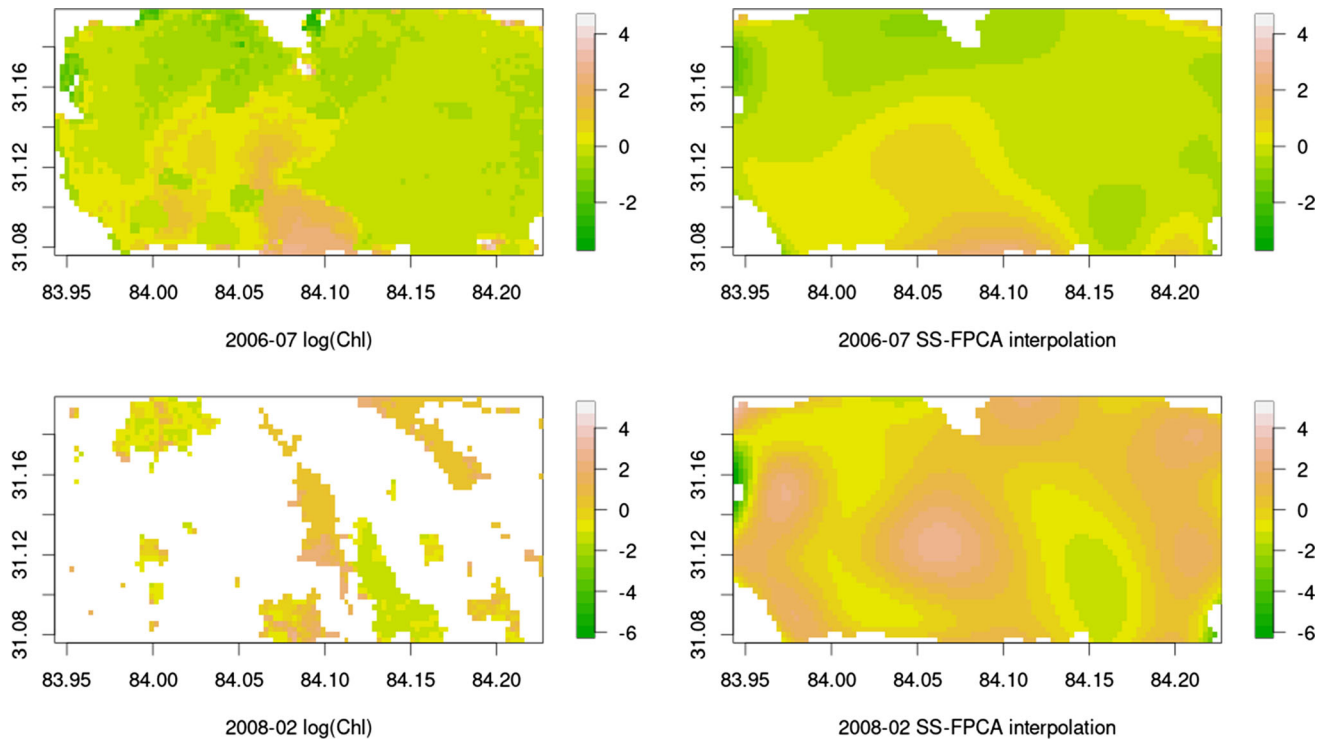

Fig. 7 Observed (left) and interpolated (right) log transformed chlorophyll-*a* data from July and February 2006. The horizontal and vertical axes are longitude and latitude respectively

environmental science research. In the application to the Taruo Lake chlorophyll-*a* data, the SS-FPCA model enabled the identification of changes in dominant spatial variation patterns over time and the automatic detection of more unusual features, which may signal an issue in lake water quality over corresponding time period. Such information from automatic processing of a large amount of Earth observation data is essential to the management of lakes where in situ measurements are hard to obtain.

The implementation of the SS-FPCA follows the empirical hierarchical modelling approach described in Cressie and Wikle (2011). A 2-cycle AECM algorithm was developed such that analytical solutions are available for the MLEs of all model parameters. The simulation study in Sect. 3 suggested that the estimations from the 2-cycle AECM algorithm were robust under different simulation scenarios, e.g. noise levels, sparsity and initial values. It also showed that the SS-FPCA model was able to capture the spatiotemporal patterns in the data. However, the estimation of the variances of some model components may be compromised due to the confounding between different model components, a common problem in spatiotemporal modelling. A potential consequence of this can be a low coverage probability of the bootstrap confidence interval, as the estimated parameters are used to generate the data in a parametric bootstrapping procedure. Further investigations are required to improve the model estimation method in the future.

An alternative method to investigate spatiotemporal data in remote sensing is the DINEOF method based on empirical orthogonal functions (Alvera-Azárate et al. 2005). It is widely used in remote sensing and is known to be computationally efficient. Compared to DINEOF, the SS-FPCA model may still be considered as computationally expensive and the AECM algorithm may suffer from slow convergence if the likelihood function is flat or if the initial values are badly selected. However, DINEOF does not account for the temporal correlations and it is not designed for identifying spatiotemporal patterns in the data. Hence, it is not an ideal tool for investigating the evolution of an environmental variable in space and time. Nevertheless, it is important to improve the computation efficiency of the SS-FPCA model, as batch processing of hundreds of data sets is a common task in remote sensing. After timing the code developed for the AECM algorithm, it is found that the majority of the computation time is consumed by the Kalman filter, where high-dimensional matrix inversions are sometimes involved. One potential solution may be to adopt the sparse matrix techniques.

Extensions to the SS-FPCA model can be made by modifying the specifications of various model components, e.g. the design of system dynamic and the dependence of random components. In this paper, the system dynamic is

supposed to follow a VAR(1) process. Alternatively, PDEs and physical/chemical models may be used if there are evidences supporting such a specification (Cressie and Wikle 2011). The covariance matrix $\boldsymbol{H}$ of the state transition equation may also be parameterized to reflect a specific spatial/temporal dependence. Parameterisations summarised in Xu and Wikle (2007) provide some options. Another extension could be to use different bases for the dynamic and the FPCA component, or to go a step further and use a multi-resolution basis (Katzfuss 2017). This would offer more flexibility in describing the spatial/temporal variations. For example, the basis for the dynamic component may be designed to capture the large scale temporal variation; whereas that of the FPCA component is intended to explain the smaller scale spatial variation. This would be a helpful extension whenever the behavior of the processes on different spatial scales are of interest.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00477-021-02017-w.

Acknowledgements The research presented in this manuscript is associated with the NERC GloboLakes project (2012–2018). M. Gong was funded by College of Science & Engineering, University of Glasgow for this work during her PhD (2013–2017). C. Miller, E. M. Scott, R. O'Donnell, S. Simis, S. Groom, A. Tyler, P. Hunter and E. Spyrakos were partly funded for this work through the NERC GloboLakes project (NE/J024279/1). The manuscript is original. No part of the manuscript has been published before in any language, nor is it under consideration for publication at another journal. The authors declare that there is no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter discussed in this manuscript.

Data availability The satellite-derived chlorophyll-*a* data used in this paper were produced within the NERC GloboLakes project and can be made available on request from calimnos-support@pml.ac.uk. A follow-on dataset is operationally available in the form of 10-day aggregated Trophic State Index and Turbidity, freely available from the operational Copernicus Land Monitoring Service: https://land.copernicus.eu/global/products/lwq. The lake surface water temperature data are publicly available from the ARC Lake project: http://www.laketemp.net/home_ARCLake/data_access.php. The R code to implement the statistical method developed in this paper is available on request from m.gong1@lancaster.ac.uk

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alvera-Azárate A, Barth A, Rixen M, Beckers JM (2005) Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. Ocean Model 9(4):325–346

Alivernini M, Lai Z, Frenzel P, Furstenberg S, Wang J, Yun Guo, Peng P, Haberzettl T, Borner N, Mischke S (2018) Late quaternary lake level changes of Taro Co and neighbouring lakes, southwestern Tibetan Plateau, based on OSL dating and ostracod analysis. Glob Planet Change 166:1–18

Cressie N, Wikle C (2011) Statistics for spatio-temporal data. Wiley, Toronto

Cressie N, Shi T, Kang EL (2010) Fixed rank filtering for spatio-temporal data. J Comput Graph Stat 19(3):724–745

Ettinger B, Guillas S, Lai M (2012) Bivariate splines for ozone concentration forecasting. Environmentrics 23(4):317–328

Fang Y, Cheng W, Zhang Y, Li J, Wang N, Zhao S, Zhou C, Chen X, Bao A (2016) Changes in inland lakes on the Tibetan Plateau over the past 40 years. J Geogr Sci 26(4):415–438

Fassò A, Cameletti M (2009) The EM algorithm in a distributed computing environment for modelling environmental space-time data. Environ Model Softw 24:1027–1035

Gong M (2017) Statistical methods for sparse image time series of remote-sensing lake environmental measurements. PhD thesis, University of Glasgow. http://theses.gla.ac.uk/8608/

Guillas S, Lai M (2010) Bivariate splines for spatial functional regression models. J Nonparametr Stat 22(4):477–497

Ho JH, Lin TI (2010) Robust linear mixed model using the skew *t* distribution with application to schizophrenia data. Biom J 52(4):449–469

Hodges JS, Reich BJ (2010) Adding spatially-correlated errors can mess up the fixed effect you love. Am Stat 64(4):325–334

Hout JP, Tait H, Rast M, Delwart S, Bèzy J, Lervini G (2001) The optical imaging instruments and their applications: AATSR and MERIS. EESA Bull 106:56–66

Huang L, Wang J, Zhu L, Ju J, Daut G (2017) The warming of large lakes on the Tibetan Plateau: evidence from a lake model simulation of Nam Co, China, during 1979–2012. J Geophys Res Atmos 122:13095–13107

Hughes J, Haran M (2013) Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. J R Stat Soc Ser B 75(1):139–159

James GM, Hastie TJ, Sugar CA (2000) Principal component models for sparse functional data. Biometrika 87(3):587–602

Katzfuss M (2017) A multi-resolution approximation for massive spatial datasets. J Am Stat Assoc 112(517):201–214

Katzfuss M, Cressie N (2011) Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. J Time Ser Anal 32(4):430–446

Katzfuss M, Cressie N (2012) Bayesian hierarchical spatio-temporal smoothing for very large datasets. Environmetrics 23:94–107

Lagos-Álvarez B, Padilla L, Mateu J, Ferreira G (2019) A Kalman filter method for estimation and prediction of space-time data with an autoregressive structure. J Stat Plan Inference 203:117–130

Lahiri SN (2003) Resampling methods for dependent data. Springer, New York

Louis TA (1982) Finding the observed information matrix when using the EM algorithm. J R Stat Soc Ser B 44(2):226–233

Ma Q, Zhu L, Lu X, Guo Y, Ju J, Wang J, Wang Y, Tang L (2014) Pollen-inferred Holocene vegetation and climate histories in Taro Co, southwestern Tibetan Plateau. Chin Sci Bull 59(31):4101–4114

McLachlan GJ, Krishnan T (1997) The EM algorithm and extensions. Wiley, Toronto

McLachlan GJ, Peel D, Bean RW (2003) Modelling high-dimensional data by mixtures of factor analyzers. Comput Stat Data Anal 41:379–388

McNicholas PD, Murphy TB (2008) Parsimonious Gaussian mixture models. Stat Comput 18:285–296

Meng X, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80(2):267–278

Meng X, Van Dyk D (1997) The EM algorithm—an old folk-song sung to a fast new tune. J R Stat Soc Ser B 59(3):511–567

National Center for Atmospheric Research Staff (eds) (2013) The climate data guide: empirical Orthogonal Function (EOF) analysis and rotated EOF analysis. http://climatedataguide.ucar.edu/climate-data-tools-and-analysis/empirical-orthogonal-function-eof-analysis-and-rotated-eof-analysis. Accessed April 2021

Padilla L, Lagos-Álvarez B, Mateu J, Porcu E (2020) Space-time autoregressive estimation and prediction with missing data based on Kalman filtering. Environmetrics 31(7):1–16

Reich BJ, Hodges JS (2008) Identification of the variance components in the general two-variance linear model. J Stat Plan Inference 138:1592–1604

Rice JA, Wu CO (2001) Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics 57(1):253–259

Rodgers CD (1990) Characterization and error analysis of profiles retrieved from remote sounding measurements. J Geophys Res 95(D5):5587–5595

Shumway RH, Stoffer DS (2006) Time series analysis and its application with R examples. Springer, New York

Stoffer DS, Wall KD (1991) Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. J Am Stat Assoc 86(416):1025–1033

The United Nations (2018) Sustainable Development Goal 6 Synthesis Report 2018 on Water and Sanitation. United Nations Publications, New York. http://www.unwater.org/publication_categories/sdg-6-synthesis-report-2018-on-water-and-sanitation/. Accessed April 2021

van Dyk DA, Meng X (2001) The art of data augmentation. J Comput Graph Stat 10(1):1–50

van Dyk D, Meng X (2010) Cross-fertilizing strategies for better EM mountain climbing and DA field exploration: a graphical guide book. Stat Sci 25(4):429–449

Williamson CE, Saros JE, Vicent WF, Smol JP (2009) Lakes and reservoirs as sentinels, integrators, and regulators of climate change. Limnol Oceanogr 54(6):2273–2282

Wood SN (2017) Generalised additive models: an introduction with R, 2nd edn. Chapman & Hall, Boca Raton

Wu Y, Zhang X, Zheng H, Li J, Wang Z (2017) Investigating changes in lake systems in the south-central Tibetan Plateau with multi-source remote sensing. J Geogr Sci 27(3):337–347

Xu K, Wikle C (2007) Estimation of parameterized spatio-temporal dynamic models. J Stat Plan Inference 137:567–588

Zammit-Mangion A, Cressie N (2020) Fixed rank kriging: the R package. https://cran.r-project.org/web/packages/FRK/vignettes/FRK_intro.pdf. Accessed April 2021

Zimmerman DL, Cressie N (1992) Mean squared prediction error in the spatial linear model with estimated covariance parameters. Ann Inst Stat Math 44(1):27–43

Zhou L, Pan H (2014) Principal component analysis of two-dimensional functional data. J Comput Graph Stat 23(3):779–801

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.