



Time-Varying Functional Principal Components for Non-Stationary EpCO_2 in Freshwater Systems

Amira ELAYOUTY¹, Marian SCOTT, and Claire MILLER

Outgassing of carbon dioxide (CO_2) from river surface waters, estimated using partial pressure of dissolved CO_2 , has recently been considered an important component of the global carbon budget. However, little is still known about the high-frequency dynamics of CO_2 emissions in small-order rivers and streams. To analyse such highly dynamic systems, we propose a time-varying functional principal components analysis (FPCA) for non-stationary functional time series (FTS). This time-varying FPCA is performed in the frequency domain to investigate how the variability and auto-covariance structures in a FTS change over time. This methodology, and the associated proposed inference, enables investigation of the changes over time in the variability structure of the diurnal profiles of the partial pressure of CO_2 and identification of the drivers of those changes. By means of a simulation study, the performance of the time-varying dynamic FPCs is investigated under different scenarios of complete and incomplete FTS. Although the time-varying dynamic FPCA has been applied here to study the daily processes of consuming and producing CO_2 in a small catchment of the river Dee in Scotland, this methodology can be applied more generally to any dynamic time series.

Supplementary materials accompanying this paper appear online.

Key Words: Functional time series; Frequency domain; Partial pressure of carbon dioxide; Smoothing; Non-stationarity.

1. INTRODUCTION

The anthropogenic increase in atmospheric carbon dioxide (CO_2) has been considered the ultimate cause for climate change, and hence, a more comprehensive understanding of the global carbon cycle, where freshwaters play an important role, is essential. The primary inputs of CO_2 into the river systems are the in situ respiration of organic material and the influx of soil CO_2 from terrestrial and wetland environments through the decomposition of

A. Elayouty (✉) · M. Scott · C. Miller, School of Mathematics and Statistics, University of Glasgow, Glasgow, UK (E-mail: amira.elayouty@glasgow.ac.uk; a.ayouti@feeps.edu.eg).

A. Elayouty, Department of Statistics, Faculty of Economics and Political Science, Cairo University, Giza, Egypt.

© 2022 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 27, Number 3, Pages 506–522
<https://doi.org/10.1007/s13253-022-00494-2>

organic material and root respiration, while the principal source of aqueous CO_2 is photosynthesis (Richey 2013). Many of the world rivers are supersaturated with CO_2 , resulting in a large outgassing that has recently been identified as a significant component of the global carbon budget (Cole et al. 2007; Butman and Raymond 2011; Wang et al. 2017).

A measure of the capacity for CO_2 exchange between freshwater and atmosphere, reflecting whether a river acts as sink or source of atmospheric CO_2 , is the excess partial pressure of carbon dioxide (EpCO_2) (Raymond et al. 1997). EpCO_2 is highly dynamic in low-order rivers; it varies spatially and seasonally due to differences in land use, water discharge, temperature, biological respiration and photosynthesis rates (Raymond et al. 1997; Wang et al. 2017). The ratio of in situ respiration to photosynthesis may also change over the course of a day, resulting in large EpCO_2 fluctuations between day and night, due to changes in temperature and solar radiation (Reiman and Xu 2019; Wang et al. 2017). Therefore, sub-daily measurements across different seasonal periods in such systems are necessary to evaluate the variability in EpCO_2 across different timescales (Dawson et al. 2009; Waldron et al. 2007; Wang et al. 2017). A wavelet analysis of 15-min resolution EpCO_2 data in a small order stream of the river Dee in Scotland showed that the major contributor to the total EpCO_2 variance is the intra-day variability, as opposed to inter-day variability, which also seems not constant over time (Elayouty et al. 2016). Establishing whether and how these daily patterns vary over time in relation to underlying hydrological and biological conditions poses statistical challenges due to the large volume of noisy data, persistent and dynamic correlation between observations, and varying relationships between the process drivers over different timescales.

By segmenting the river Dee high-resolution sequence of EpCO_2 into natural consecutive daily intervals, the data for each day can be viewed as realisations of a stochastic functional time series (FTS) $\{X_k, k \in \mathbb{Z}\}$. Each X_k in the series is a continuous smooth function $X_k(t)$, $t \in \mathcal{T}$, with k denoting the day as the discrete time parameter and t being the intra-day time parameter defined on the continuous domain \mathcal{T} . The main goal of the paper is thus to develop and evaluate a methodology that enables us to investigate whether and how the sources of variability among the daily functions of EpCO_2 change over time and possibly reveal the hydrological and/or biological drivers of these changes.

Functional data analysis (FDA) has recently proven to be an appropriate tool for the analysis of high-frequency data represented in the form of continuous functions (Ramsay and Silverman 1997; Wang et al. 2015; Ullah and Finch 2013), and functional principal component analysis (FPCA) is a key dimension reduction technique in FDA that can be used to explore the covariance structure in a set of functions. But, like traditional time series, data in FTS are dependent in time. It is also common for such environmental time series to be non-stationary over long time periods (Halliday et al. 2012; Horváth et al. 2014). Consequently, traditional FPCA may fail to explain appropriately the varying covariance structure and hence provide inefficient inference through the dimensionality reduction. To accommodate for temporal correlation and non-stationarity in multivariate time series, several PCA techniques are available in both the time domain (Ku et al. 1995; Miller and Bowman 2012; Rodrigues 2007; Vanhatalo et al. 2017; Melnikov et al. 2016) and the frequency domain (Brillinger 1981; Ombao and Ringo 2006; Pena and Yohai 2016). For the PCA of temporally correlated functional data, Jaimungal and Eddie (2007) first modelled the functions' basis

coefficients over time using vector auto-regressive (VAR) models and then applied standard PCA to the coefficients' residuals which are less correlated. This, however, can lead to the loss of valuable information about the temporal patterns present in the data. Alternatively, [Hormann et al. \(2015\)](#) extended the theory of dynamic PCA in the frequency domain founded by [Brillinger \(1981\)](#) to the context of functional data. In this approach, the FPCs are obtained based on the spectral density which contains information on the covariances at all time lags. This enables the approximation of the curves, using orthogonal PCs which involve lagged observations, based on information available at nearby observations. [Panaretos and Tavakoli \(2013\)](#) proposed a similar idea for extending dynamic PCA to FTS, in which the reconstructed curves are based on a stochastic integration with respect to some orthogonal increment functional stochastic process. Both of these recent developments did not account for non-stationarity in the covariance structure over time in a FTS. But, [Yan et al. \(2018\)](#) considered the application of traditional FPCA to longitudinal data on a series of moving time windows of the functional domain to extract window-specific FPC scores under some smoothing constraints. The FPC scores obtained within each window are used in dynamic prediction assuming independence between functions, which may not be a realistic assumption in FTS. More recently, [Dubey and Muller \(2019\)](#) extended FPCA to the case where observations are metric-valued functions such as time-varying probability distributions or time-dynamic networks. But because the proposed object FPCs rely on an auto-covariance metric measuring only the association between the functions at lag 0, they can miss the time dependence between the objects.

In this paper, we develop a nonparametric time–frequency-domain FPCA procedure (and associated inference) that investigates and statistically assesses the changes over time in the variability structure of a non-stationary FTS, taking into account the potentially valuable information contained in the past and future observed functions. This version of FPCA allows the principal modes of variability to vary over time simultaneously with changes in the covariance structure in the FTS. This, in turn, will help optimal reconstruction of the curves from the low-dimensional time-varying FPCs that better explain the temporally varying covariance structure in the series. Based on these time-varying FPCs, a nonparametric inference procedure is proposed to test for the significance of changes in the auto-covariance structure. Although this time-varying functional principal components analysis (FPCA) has been motivated here by a highly dynamic series of daily aqueous CO₂ profiles to study its dynamics and modes of variability, it can be applied to any dynamic FTS.

The rest of the paper is organised as follows. Section 2 gives the motivation for the proposed methodology and the transformation of 15-min raw time series of EpCO₂ into a FTS of daily EpCO₂ profiles. Section 3 describes an empirical estimation procedure of a smooth auto-covariance structure. Then, Sect. 4 presents the methodological and practical implementation of the proposed time-varying FPCs to the daily profiles of EpCO₂. Section 5 evaluates the changes in these components over time, where a bootstrap-based inference procedure is proposed for the detection of changes in the auto-covariance structure. In Sect. 6, the performance of the time-varying FPCs is assessed numerically using an extensive simulation study under different scenarios of complete and incomplete FTS. Finally, the results are discussed and concluding remarks are drawn in Sect. 7. Online supplementary

material includes technical definitions for FTS and frequency-domain methods in Appendix A, with additional results and figures for the proposed methodology in Appendix B.

2. EPCO₂ DATA FROM DISCRETE TO FUNCTIONAL TIME SERIES

As primary analysis of the high-resolution EpCO₂ data highlighted varying smooth daily patterns of the process over time, the drivers of such variability in EpCO₂ over the span of a day and how these diel profiles vary across seasons and different hydrological conditions are of specific interest to geo-chemists. To investigate this, the discrete (15-min) EpCO₂ data can be converted into smooth daily profiles and then explored using FTS and FPCA approaches, which will be introduced here through the application. [Elayouty et al. \(2016\)](#) found that EpCO₂ exhibits a seasonal pattern, reaching low levels in summer due to autotrophic processes and lack of carbon and organic inputs associated with lower discharge. To avoid such seasonality effects on the mean level from dominating the variability in the data, a smooth global mean was first estimated and subtracted from the data before further analysis.

Given the smoothness of the EpCO₂ process, the 15-min de-trended data over the span of each day are used to derive daily (smooth) functions using a basis function expansion. This constitutes a FTS so that each functional observation $X_k(t)$ on day k can be expressed by:

$$X_k(t) = \sum_{j=1}^p Z_{kj} \psi_j(t) = \boldsymbol{\psi}(t)^\top \mathbf{Z}_k, \quad (1)$$

where $\boldsymbol{\psi}(t)$ is the vector of basis functions $(\psi_1(t), \dots, \psi_p(t))^\top$ and \mathbf{Z}_k is the vector of corresponding coefficients $(Z_{k1}, \dots, Z_{kp})^\top$ to be estimated for the k th function. The most popular choices of basis systems to smooth environmental data include B-splines and Fourier series. Here, a cubic B-splines basis is used, which is characterised by its flexibility that does not force a strongly cyclic variation. A large dimension, $p = 15$, is specified for the basis to capture the function characteristics with a second-order derivative roughness penalty. The smoothing parameter controlling the curvature relative to the model fit is set equal 1, based on the choice of the restricted maximum likelihood method that tends to be more robust to data dependence. A sample of the resulting 1095 daily EpCO₂ curves is displayed in Fig. 1.

3. ESTIMATION OF SMOOTH COVARIANCE STRUCTURE FOR FTS

Let $X_k(t)$, $k = 1, \dots, N$, $t \in \mathcal{T}$ be the sequence of serially correlated random continuous daily curves of de-trended EpCO₂ drawn from the FTS process $(X_k : k \in \mathbb{Z})$. To study how the directions of variability in this FTS vary over time, we introduce a smoothly time-varying version of the frequency-domain FPCs by [Hormann et al. \(2015\)](#). The proposed technique relies on performing frequency-domain FPCA at each time k by using a time-varying (local) spectral density operator, whose construction is based on a whole family of

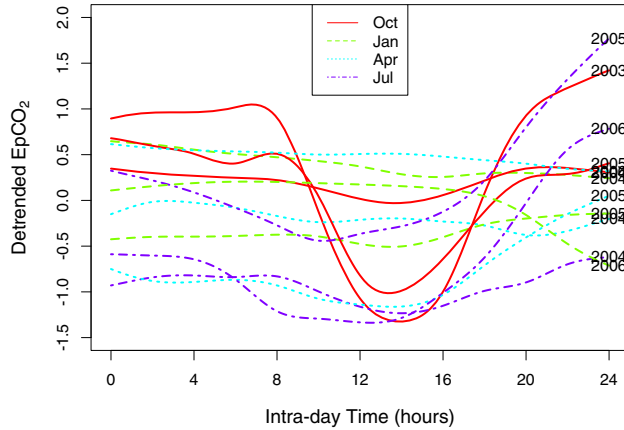


Figure 1. Sample of daily (de-trended) EpCO₂ smooth curves in the different years.

lag- h smoothed covariance operators of X_k . A covariance operator is a functional analogue of scalar TS covariance. Following the basis expansion in (1), the local lag- h covariance matrix that corresponds to the lag- h covariance operator at day k is given by $V_{k,h}^X = V_{k,h}^Z \Psi$, where $V_{k,h}^Z = \mathbb{E}[\mathbf{Z}_k \mathbf{Z}_{k+h}^\top]$ is the lag- h auto-covariance matrix of the coefficients (\mathbf{Z}_k) at time k and $\Psi := (\langle \psi_i, \psi_j \rangle = \int_{\mathcal{T}} \psi_i(t) \psi_j(t) dt : 1 \leq i, j \leq p)$ is a $p \times p$ symmetric matrix of the inner product of basis functions evaluated sometimes by numerical integration based on the chosen basis system; see Appendix A in supplementary material. This implies that the main interest lies in estimating the lag- h covariances of the basis expansion coefficients, $V_{k,h}^Z, \forall k, h \in \mathbb{Z}$.

To estimate these lag- h covariance matrices locally at each time point (day) k , we suggest using the Nadaraya–Watson covariance estimator. Due to the limited number of replicates at each time point (one functional observation per day), the sequence of lag- h covariances is smoothed over time using a weighting kernel and the sample smooth lag- h auto-covariance matrices of the coefficients \mathbf{Z}_k at each day k are computed as follows:

$$\begin{aligned} \hat{V}_{k,h}^Z &= \frac{1}{\sum_{k'=1}^{N-h} w_{k,k'}} \sum_{k'=1}^{N-h} w_{k,k'} \mathbf{Z}_{k'} \mathbf{Z}_{k'+h}^\top, & h \geq 0 \\ \hat{V}_{k,h}^Z &= (\hat{V}_{k,-h}^Z)^\top, & h < 0, \end{aligned} \tag{2}$$

where $w_{k,k'}$ is the weight assigned to the pair of observations k' and $k'+h$, based on a chosen kernel $w(\cdot)$. The weighting kernel is a monotonically decreasing function of the distance $|k - k'|$ regardless of lag h , ensuring that higher weights are assigned to observations k' near the target point k to contribute most in the estimate. Since the auto-correlation function between daily curves was found to decay exponentially with h , we have chosen to use here the Gaussian density kernel:

$$w(k - k', s) = \exp\left(-\frac{1}{2} \left(\frac{k - k'}{s}\right)^2\right), \tag{3}$$

where s is the parameter controlling the kernel's width and the smoothness of the estimated covariance. For the choice of the smoothing parameter s , a sensitivity analysis was performed and $s = 20$ seems to offer a good approximation for the original data. The sensitivity analysis indicated that using a smaller s tends to over-fit the curves, while a larger s overlooks the local variability in the curves; see Figure S.1. Note that as the smoothing parameter s tends to infinity, $\hat{V}_{k,h}^X$ tends towards the overall lag- h covariance of the data \hat{V}_h^X .

4. TIME-VARYING FREQUENCY-DOMAIN FPCA

Based on the local lag- h covariances ($V_{k,h}^X : k, h \in \mathbb{Z}$), we define the matrix corresponding to the time-varying spectral density operator of the FTS X_k at time k and frequency θ by:

$$F_{k,\theta}^X = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} V_{k,h}^X \exp(-ih\theta) = \frac{1}{2\pi} \left\{ \sum_{h \in \mathbb{Z}} V_{k,h}^Z \exp(-ih\theta) \right\} \Psi^\top = F_{k,\theta}^Z \Psi^\top, \quad (4)$$

where i denotes the imaginary unit. After the estimation of the locally auto-covariances $\hat{V}_{k,h}^Z$ as described above, $F_{k,\theta}^X$ is estimated empirically at each time point k using the lag window estimator ω as follows:

$$\hat{F}_{k,\theta}^X = \frac{1}{2\pi} \sum_{|h| \leq D} \omega\left(\frac{h}{D}\right) \hat{V}_{k,h}^Z \exp(-ih\theta) \Psi^\top, \quad k = 1, \dots, N - D; \theta \in [-\pi, \pi]. \quad (5)$$

Here, we used the Bartlett weight kernel $\omega(h/D) = 1 - |h/D|$ with bandwidth $D = \lfloor \sqrt{N} \rfloor = 33$, which controls the extent of data observed through the window for spectral density estimation. For more details about the choices of ω , refer Politis (2011). For the choice of D , a more sophisticated calibration can be used for better results; however, in such hydrological systems correlation between observations is not expected to go beyond a month.

An eigen-analysis is then performed on the local spectral densities $\hat{F}_{k,\theta}^X$ to produce smoothly time-varying frequency-domain FPCs at each time point k . The eigen-decomposition yields the eigenvalues $\hat{\lambda}_{k,m}(\theta)$ with eigenvectors $\hat{\phi}_{k,m}(\theta)$, $\forall m = 1, \dots, p$, such that $\hat{\lambda}_{k,m}(\theta)$ is the m th largest eigenvalue of $\hat{F}_{k,\theta}^X$. However, the spectral density matrix $\hat{F}_{k,\theta}^X$ is found not necessarily Hermitian, since the basis system chosen here to approximate the functional data is not orthogonal due to penalisation, which may result in non-real eigenvalues. To ensure obtaining real eigenvalues, we perform the eigen-analysis on $\Psi^{1/2} \hat{F}_{k,\theta}^Z \Psi^{1/2\top}$ instead of $\hat{F}_{k,\theta}^X$ given by (5), where $\Psi^{1/2}$ is the Cholesky decomposition of Ψ . This yields the eigenvalues $\hat{\lambda}_{k,m}(\theta)$ and the eigenvectors $\hat{\phi}_{k,m}^*(\theta) = \Psi^{1/2} \hat{\phi}_{k,m}(\theta)$. The frequency-domain eigenfunctions $\hat{\phi}_{k,m}(\theta)$ of $\hat{F}_{k,\theta}^X$ can then be obtained by $\psi(t)^\top \Psi^{-1/2} \hat{\phi}_{k,m}^*(\theta)$, from which the time-varying FPCs $\phi_{kml}(t)$ are estimated via its inverse Fourier transform as follows:

$$\hat{\phi}_{kml}(t) = \frac{\psi(t)^\top}{2\pi} \int_{-\pi}^{\pi} \hat{\phi}_{k,m}(\theta) \exp(i\theta t) d\theta = \psi(t)^\top \hat{\mathbf{b}}_{kml}, \quad (6)$$

for a set of lags $l = -L, \dots, L$ at each time $k = 1, \dots, N - D$ ($D > L$). By definition, $\sum_{l \in \mathbb{Z}} \|\hat{\phi}_{kml}(t)\|^2 = 1$ at each time point k . To guarantee a fair and good approximation across all time points, the maximum lag L is chosen such that the average $\frac{1}{N-D} \sum_{l=-L}^L \|\hat{\phi}_{kml}(t)\|^2 \geq 1 - \epsilon$, for some small ϵ . Here, L is set equal to 30 for which $\epsilon < 0.01$. Note that the eigenvectors $\hat{\phi}_{k,m}(\theta)$ do not have analytic form, and so, the time-domain functional filters $\hat{\phi}_{kml}(t)$ are obtained via numerical integration; see [Hormann et al. \(2015\)](#).

Similar to the frequency-domain FPCs by [Hormann et al. \(2015\)](#), the score of the m th FPC at time k is obtained by filtering the original series using the estimated FPCs as follows:

$$\hat{Y}_{m,k} = \sum_{l=-L}^L \int_{t \in \mathcal{T}} X_{k-l}(t) \hat{\phi}_{kml}(t) dt = \sum_{l=-L}^L \mathbf{z}_{k-l}^\top \Psi \hat{\mathbf{b}}_{kml}. \quad (7)$$

To compute the scores for $1 \leq k \leq L$, we set $X_{-L+1} = \dots = X_0 = \mathbb{E}[X_k]$; which may create bias at the start of the observation period. Using the first q time-varying FPCs, $q \leq p$, and corresponding scores, we can thus approximate the original daily curves $X_k(t)$ of EpCO₂ by:

$$\hat{X}_k^q(t) = \sum_{m=1}^q \sum_{l=-L}^L \hat{Y}_{m,k+l} \hat{\phi}_{kml}(t), \quad k = 1, \dots, N - D. \quad (8)$$

The asymptotic properties of the (Nadaraya–Watson) smooth covariance estimator are studied, and convergence is guaranteed at a suitable rate as the degree of smoothing tends to zero ([Petersen et al. 2019](#)). These results can be approximately extended to the local spectral density estimator given the consistency of $\hat{\mathcal{F}}_\theta^X$; see [Hormann et al. \(2015\)](#). The complexity of the eigen-decomposition of this spectral density as a function of its individual components prevents a simple derivation of expressions for the bias and variance of eigenvalues and eigenfunctions. However, by considering the case where the true spectral density at each time k has a unique eigen-decomposition then the convergence of the estimated local spectral density at least ensures the convergence of the estimated local eigenvalues and eigenfunctions to their true values, given the right amount of smoothness. Based on this statement and the asymptotic properties of dynamic FPCs discussed in [Hormann et al. \(2015\)](#), when smoothed appropriately, $\mathbb{E}\|X_k\|^2 \approx \sum_{m=1}^\infty \int_{-\pi}^\pi \lambda_{k,m}(\theta) d\theta$ for all k belonging to the set $1, \dots, N - D$. Thus, the proportion of variance explained by the first q time-varying FPCs $\sum_k \sum_{m \leq q} \int_{-\pi}^\pi \lambda_{k,m}(\theta) d\theta / \sum_k \mathbb{E}\|X_k\|^2$ can be estimated as 1 minus the normalised mean square error (NMSE) given by:

$$\text{NMSE}(q) = \sum_{k=1}^{N-D} \|X_k(t) - \hat{X}_k^q(t)\|^2 / \sum_{k=1}^{N-D} \|X_k(t)\|^2. \quad (9)$$

This NMSE is also considered a measure for the information loss resulting from reducing the data dimensionality to dimension q .

Figure 2 shows that, using 1 component, the time-varying version promotes better approximation in retrieving the original mean pattern of the EpCO₂ data as well as the within-day

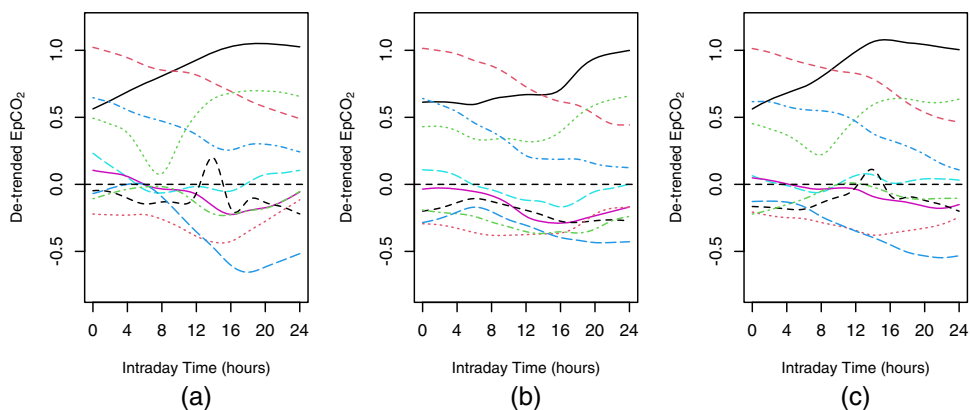


Figure 2. **a** 10 successive daily curves of de-trended EpCO_2 and the corresponding reconstructions based on **b** the first dynamic FPC and **c** the first time-varying FPC with $s = 20$.

variability. That is, this one time-varying FPC takes over the roles of the subsequent ones and account for 94% of the variability at $s = 20$ versus 74% for the first dynamic FPC. More components can be used in the approximation but one is enough here for demonstration purposes. These results imply that the time-varying dynamic FPCA simplifies the complexity in data variability using a smaller number of components, compared to dynamic FPCA that does not account for the varying time dependence structure. Comparing the scores of the first dynamic and time-varying FPCs, we found that they are similar over time except in periods where the time-varying FPC seems to account for more variability in the series; see Figure S.2. Those periods characterised by higher EpCO_2 volatility are mostly observed between April and October, where the diel pattern is more pronounced due to higher temperatures, longer daylight lengths and increased autotrophic processes in summer relative to winter. Those differences in EpCO_2 diel patterns are not driven solely by seasonality but also regulated by the system hydrology (Reiman and Xu 2019).

5. EVALUATION OF THE TEMPORAL VARIATIONS IN EPCO_2

It appears from the analysis above that the dependence structure in the FTS of EpCO_2 changes throughout time. For a better understanding of the drivers of those EpCO_2 diel variations, a K-means clustering of the diel patterns based on the coefficients of the three central functional filters of the first time-varying FPC is performed and results are discussed in Sect. 5.1. Next, a method of inference along with a graphical output based on the frequency-domain local eigenvalues of particular FPCs is presented in Sect. 5.2 to empirically detect changes in the auto-covariance structure of the EpCO_2 FTS.

5.1. K-MEANS CLUSTERING

The results of K-means clustering of the central functional filters' coefficients of the first frequency-domain FPC highlight the presence of three main clusters governing the

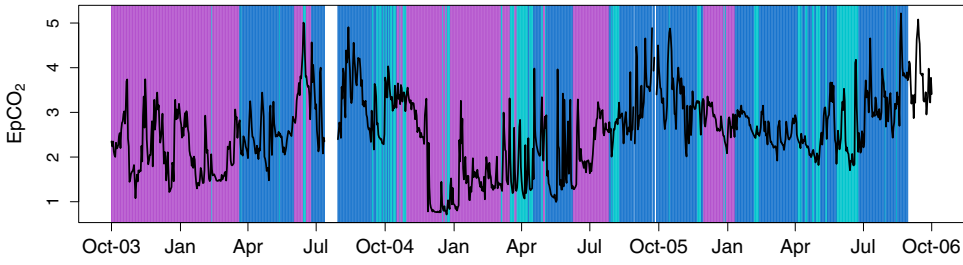


Figure 3. Class membership of EpCO_2 daily curves obtained using K-means clustering based on the corresponding coefficients of the three central functional filters of the first FPC. The solid curve is the corresponding daily averaged EpCO_2 .

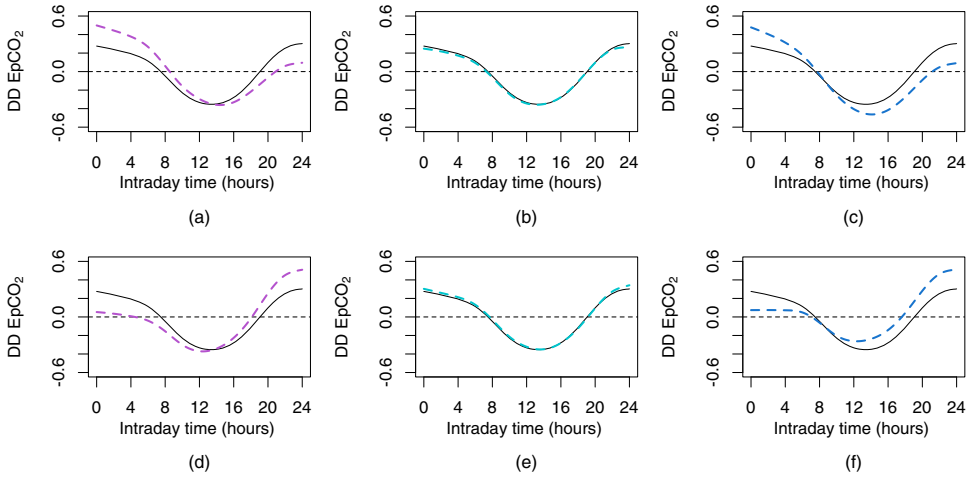


Figure 4. The functional mean $\mu(t)$ of the de-trended EpCO_2 daily curves (solid line) and $\mu(t) + \sum_{l=-1}^1 \delta_l \bar{\phi}_{1l}(t)$ with $(\delta_{-1}, \delta_0, \delta_1) = (-1, 1, 1)$ (top) and $(1, -1, -1)$ (bottom) for the three clusters.

variability in the EpCO_2 diel patterns and the differences in the clustering structure between the 3 years; see Fig. 3. To study the differences in the covariance structure between clusters, the average impact of consecutive scores $(\hat{Y}_{1,k-1}, \hat{Y}_{1,k}, \hat{Y}_{1,k+1})$ on the k th daily curve of EpCO_2 in each of the clusters is investigated in Fig. 4. The purple cluster appears to underlie the less pronounced diel patterns that mostly prevail in autumn and winter where a change in the average EpCO_2 at day $k - 1$ is likely to have a delayed impact on the night/dark levels of day k . The blue cluster governs periods with persisting sharp diel patterns across adjacent days. These diurnal patterns are more common in spring and summer where a change in EpCO_2 at day $k - 1$ results in a rapid change over the daylight hours of the following day. The turquoise cluster includes the periods with more volatile and less correlated diel patterns, where the diel pattern is more likely to be driven by the mean level of EpCO_2 on the day.

We can thus conclude that the EpCO_2 diel pattern is very likely driven by the autotrophic processes that increase and become more persistent in summer due to higher temperatures, longer daylight lengths and increased solar radiation. However, in a small-order river, the hydrological activity and discharge events represent another important factors that affect

this diurnal activity in different ways, which coincides with the conclusion of Wang et al. (2017). This is that, discharge events on one hand can transport soil CO₂ to the river which may contribute to rising biological activity and hence the riverine CO₂, while on the other hand, EpCO₂ may decline due to the dilution effect of constant or large discharge events (Reiman and Xu 2019). The above results indicate that the system is coherent over time except some periods (in Turquoise), which can be referred to as transitional periods. This degree of coherence between the diel patterns of EpCO₂ also varies over time. For instance, the system is less variable and changes are slower in the winter, possibly due to the low levels of nutrients and temperature, whereas the biological activity accompanied with soil run-off is likely to be accelerated in summer as a result of the relatively stronger correlation between days in summer. These results vary to some extent between the years based on the underlying mixing conditions of water temperature and hydrological regimes. To verify changes over time in the system correlation structure, we propose the following inference procedure.

5.2. INFERENCE METHOD

To detect changes in the auto-covariance structure of a FTS, we propose utilising a similar approach to that suggested for multivariate time series by Miller and Bowman (2012), but rather based on the spectral density eigenvalues corresponding to the leading time-varying FPCs. The general null hypothesis of interest here, $H_0: (X_k : k \in \mathbb{Z})$, is a weakly stationary FTS, versus the alternative hypothesis, $H_A: (X_k : k \in \mathbb{Z})$, is a locally stationary FTS. By removing the trend from the data as described in Sect. 2, the null hypothesis becomes equivalent to testing that the FTS has a covariance structure and hence a spectral density that do not vary throughout time (i.e. $V_{k,h}^X = V_h^X \forall k, h$ and $F_{k,\theta}^X = F_\theta^X \forall k, \theta$). This hypothesis can be investigated empirically by evaluating whether or not the observed changes over time in dynamic FPCs are consistent with sampling variation. We, therefore, formulated the test statistic for evaluating whether the whole spectra of a particular frequency-domain FPC m changes over time (i.e. $H_0: \Lambda_{m,k} = \Lambda_m \forall k$ Vs $H_A: \Lambda_{m,k} \neq \Lambda_m$ for at least one k) as follows:

$$T_m = \sum_{k=1}^{N-D} |\hat{\Lambda}_{mk} - \hat{\Lambda}_m|, \quad \theta \in [-\pi, \pi], \tag{10}$$

where $\hat{\Lambda}_{mk} = \int_{-\pi}^{\pi} \hat{\lambda}_{mk}(\theta) d\theta$ and $\hat{\Lambda}_m = \int_{-\pi}^{\pi} \hat{\lambda}_m(\theta) d\theta$. Note that $\hat{\lambda}_m(\theta)$ is the m th eigenvalue of the stationary spectral density \hat{F}_θ^X of the fully observed time series at frequency θ .

Kokoszka and Jouzdani (2020) have derived asymptotic results for the frequency-domain eigenvalues of a stationary FTS, such that $\sqrt{N/D}|\hat{\Lambda}_m - \Lambda_m| \xrightarrow{d} N(0, \sigma_m^2)$. However, it is difficult to modify these results to suit the local nature of the covariance structure and spectral density presented here and therefore a resampling-based inference approach is used. The reference distribution is constructed in a way that illustrates the expected behaviour of a particular eigenvalue over time if the null hypothesis, that the auto-covariance structure does not vary over time, is true. For FTS like EpCO₂, subsequent daily curves are auto-correlated, and hence, the reference distribution is constructed using parametric bootstrapping. This is

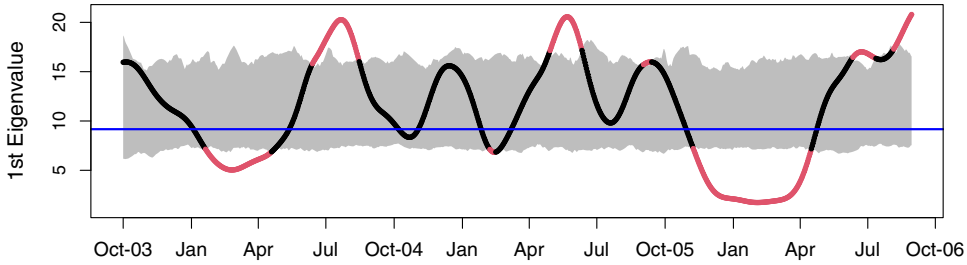


Figure 5. Plot of the first time-varying dynamic FPC eigenvalues obtained at each time point averaged over all frequencies $\theta \in (-\pi, \pi)$ along with the corresponding 95% variability band. The blue horizontal line represents the corresponding eigenvalue of the first stationary dynamic FPC.

done by simulating FTS from a stationary model fitted to the observed data that incorporates the correlation via a known covariance structure that does not change over time.

Exploratory analysis of the EpCO_2 daily curves indicated that a functional AR(1) provides a reasonable choice for modelling the correlation in these data. Following from this, the distribution of the test statistic under the null hypothesis, that the underlying process follows a stationary functional AR(1), is constructed by simulating 200 FTSs from the functional AR(1) fitted to the EpCO_2 data. Practically, this simulation is performed in a finite dimension p using the basis functions $\{\psi_j\}_{j=1}^p$ such that the basis coefficients, $\mathbf{Z}_k = (\langle X_k, \psi_1 \rangle, \dots, \langle X_k, \psi_p \rangle)$, are generated from a VAR(1) of the form $\mathbf{Z}_k = \mathfrak{R}\mathbf{Z}_{k-1} + \tilde{\boldsymbol{\epsilon}}_k$, fitted to the basis coefficients of the de-trended EpCO_2 daily functions, where \mathfrak{R} is the matrix of auto-regressive parameters estimated using least squares and $\tilde{\boldsymbol{\epsilon}}_k \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ are independent and identically distributed noise with a covariance structure that does not change over time. For each simulated FTS, having the same number of curves as the original data, the smooth covariance structure and spectral density are estimated at each time point using a Gaussian weighting kernel with $s = 20$. The sequences of local eigenvalues associated with the leading time-varying FPCs can then be used to evaluate whether the power spectra of those PCs, and hence the variability structure in the data, vary over time. For illustration, only the first component responsible for maximum variability is considered here. The first largest eigenvalues of local spectral densities are estimated and summed across all $\theta \in (-\pi, \pi)$ at each time point. Figure 5 displays the sequence of the first general eigenvalue $\hat{\Lambda}_{1k}$ along with a corresponding point-wise 95% variability band, which indicates that the power spectra of the first frequency-domain PC vary over time with higher values observed in summer and considerably low values during winter. To protect ourselves more effectively against falsely rejecting the null hypothesis of stationarity as a result of multiple testing, we compute the overall test statistic, given by (10), based on the eigenvalue of the 1st component. This test produces a p-value less than 0.001 highlighting a significant change in the variability structure throughout time in the sense that most drastic intra-day fluctuations are observed in summer. Although the above inference procedure is presented in terms of the EpCO_2 application, it can be widely applied to other applications. But note that the construction of the reference distribution will change based on the choice of the appropriate model for the correlation structure in the FTS.

6. SIMULATION STUDIES

To evaluate the performance of our proposed time-varying dynamic FPCs, we have adopted two simulation settings with increasing complexity. The *first setting* involves densely sampled FTS for a variety of non-stationary data-generating conditions with different intensity of time dependence within the series. The *second setting* involves sparsely sampled FTS with different levels of sparsity. The two settings show that the proposed FPCs do not require the original FTS to be stationary or complete without missing data. For each simulated FTS, the first q dynamic FPCs and their time-varying analogue are estimated and used to recover the simulated FTS. The performance of these approximations is compared using the NMSE in (9). The smaller the NMSE, the superior the approximation.

Simulation of non-stationary (complete) FTS We start by generating a FTS $\{X_k\}$ from a stationary functional AR(1) with a correlation structure that mimic that of the EpCO₂ data application as described in Sect. 5.2. Next, we estimate the traditional time-domain eigenvalues $\lambda = (\lambda_1, \dots, \lambda_p)$ and eigenfunctions $E_1(t), \dots, E_p(t)$ of this simulated FTS, ignoring the time dependence in the series. By extending the findings of [Mardia et al. \(1979\)](#) to a functional context, we can thus construct a functional process \mathcal{X}_k by:

$$\mathcal{X}_k(t) \approx \bar{\mathcal{X}}(t) + \sum_{m=1}^p S_{mk} E_m(t), \quad t \in \mathcal{T}, \quad (11)$$

where $\bar{\mathcal{X}}(t)$ is the functional mean of $\mathcal{X}_1(t), \dots, \mathcal{X}_N(t)$, S_{mk} is the score of the m th PC for the k th observation and $E_m(t)$ is an m th eigenfunction for the data. Based on this result, we can simulate locally stationary FTS with a power spectrum that varies smoothly over a grid of, say $B = 20$, time blocks by multiplying λ by $b/10$ and getting λ_b for $b = 1, \dots, B$. At each block b , the scores for each PC are generated from a VAR(1) process with an auto-regression matrix \mathfrak{R}^* and MVN($\mathbf{0}, \Sigma_b$) noise, such that $\Sigma_b = \text{diag}(\lambda_b)$. The matrix \mathfrak{R}^* is set equal to $\kappa G/\|G\|$, where κ denotes a pre-specified dependence coefficient and G is a $p \times p$ matrix with elements $(G_{m,m'} : m \geq 1, m' \leq p)$ that are mutually independent $N(0, \exp\{-(m + m')\})$. By multiplying these simulated scores by the eigenfunctions $E_1(t), \dots, E_p(t)$, we obtain a block stationary process \mathcal{X}_k of 400 observations.

The above simulation is repeated 200 times for $\kappa = 0.1, 0.3, 0.6, 0.9$, reflecting weak to strong dependence structure in the data. For each choice of κ , the simulated FTSs are approximated by the first $q = 1, 2, 3$ stationary and time-varying dynamic FPCs at different values of the weighting kernel smoothing parameter s . We considered the values $s = 20, 40, 100$ for low to high levels of smoothness of the estimated covariance structure. Figure 6 and Figure S.3 and Table S.1 in supplementary material show the mean and standard deviation of NMSE computed for each scenario. Results indicate that in almost all scenarios the time-varying FPCs outperform the stationary ones with lower mean NMSE between the simulated series and their reconstructed versions. The NMSE of both versions of FPCs decreases and approximation improves as the dependence in the series increases. Those differences between the two reconstructions become negligible as the number of PCs and serial dependence increase. It is also noticed that although the variance of NMSE increases with serial dependence, it is systematically lower for the time-varying version.

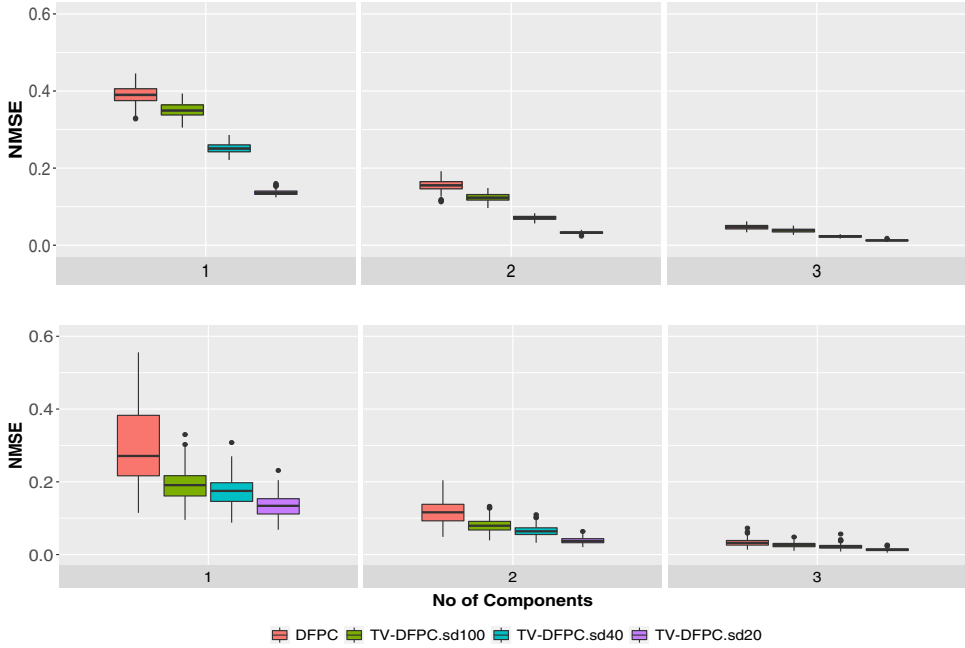


Figure 6. Box plots of NMSE between the simulated curves and their recovered versions using 1, 2 and 3 (from left to right) dynamic (red) and time-varying (smooth) dynamic FPCs with a smoothing parameter $s = 100$ (olive), 40 (blue) and 20 (purple) based on the results from 200 non-stationary simulation runs with $\kappa = 0.1$ (top) & $\kappa = 0.9$ (bottom). Results for $\kappa = 0.3$ & $\kappa = 0.6$ are in supplementary material.

We have also used the above simulation to empirically compare the time-varying eigenvalues' to the true underlying frequency-domain eigenvalues of the generating process, compared to the time-invariant ones. This is done by calculating the mean squared errors, $\frac{1}{N-D} \sum_{k=1}^{N-D} \|\Lambda_{km} - \hat{\Lambda}_{km}\|^2$, between the m th true eigenvalue and its corresponding estimate from the eigen-analysis of the empirically estimated spectral densities summed over all $\theta \in [-\pi, \pi]$ and aggregated across the whole grid of time points k . By expressing each eigenfunction $E_m(t)$ in terms of the same basis functions used to simulate the original FTS as $\psi(t)^\top \mathbf{c}_m$, the true frequency-domain eigenvalues of \mathcal{X}_k for each block b can be obtained via the eigen-decomposition of the spectral density matrix:

$$F_b^{\mathcal{X}}(\theta) = \frac{1}{2\pi} \Psi^{1/2} \mathbf{C} (\mathbf{I}_{15} - \Re^* e^{-i\theta})^{-1} \Sigma_b (\mathbf{I}_{15} - \Re^* e^{-i\theta})^{-1} \mathbf{C}^\top \Psi^{1/2\top}, \quad (12)$$

where \mathbf{C} is a $p \times p$ matrix whose m th column is the vector \mathbf{c}_m and \mathbf{I}_{15} is an identity matrix of order 15. Results in Figure S.4 of supplementary material indicate that, for non-stationary FTS, the difference between the estimated time-varying frequency-domain eigenvalues and the corresponding true values is smaller than its counterpart for the stationary eigenvalues. Results also show that a value of $s = 40$ provides a better accuracy for the estimates, which highlights the importance of the smoothing parameter choice to the estimation accuracy.

Simulation of non-stationary (sparse) FTS Here, we generate a sequence of $N = 400$ functional observations from $X_{k+1}(t) = \Upsilon_k(X_k(t)) + \varepsilon_{k+1}(t)$, where $\Upsilon_k(\cdot)$ is an auto-

regressive operator with a Hilbert Schmidt norm $\|\Upsilon_k(\cdot)\|_{\mathcal{S}}$ that varies smoothly over time between 0.1 and 0.9. To obtain the realisations of the curves at a finite set of time points and produce sparsity in the data, this simulation is performed in a discrete setting by assuming that the operator Υ_k has a Gaussian kernel $\Upsilon_k(t, s) = \|\Upsilon_k\|_{\mathcal{S}} \exp\left(-\frac{t^2 + s^2}{2}\right)$. To ensure the smoothness of those densely simulated functions, the innovations ε_k are generated from a smooth Gaussian process with a Matèrn covariance function. This simulation is repeated 200 times. Then, for each FTS, a random collection of curves constituting $\nu = 10\%$ and 35% of the total number of curves in the series are made incomplete. For each randomly chosen incomplete curve, a block of $\tau = 35\%$, 60% and 100% of the raw data, representing the realisations of the curve at a set of finite time points, are made missing. Thus, for each FTS, different levels of sparsity using a mixture of ν and τ are applied. The estimated stationary and time-varying dynamic FPCs are then used to reconstruct the original series, and the NMSE is calculated. As frequency-domain FPCA requires a complete FTS, a naive imputation of the missing values is used as initial values. Results in Figures S.5–S.7 indicate that, in all missing data patterns, the missing data were better recovered using the time-varying FPCs with $s = 20$ relative to the stationary FPCs.

7. DISCUSSION AND CONCLUSIONS

FDA has proven to be an appropriate tool for extracting the most important characteristics of high-dimensional data and reducing the data dimensionality. A common situation in environmental applications is when functional data are observed sequentially in time and typically exhibit auto-correlation and/or non-stationarity over time. Traditional FPCA ignores information carried by nearby observations, which may lead to inappropriate dimension reduction. In this paper, we propose a time-varying version of frequency-domain FPCA that accounts for both time dependence and covariance non-stationarity in FTS. The motivating application is the analysis of daily curves of excess partial pressure of CO_2 derived from a 15-min sensor records over 3 years in a small catchment of the river Dee in Scotland.

The proposed time-varying FPCA relies on the estimation of a local spectral density estimated based on a smooth covariance structure. The method presented here uses kernel-based smoothing methods to estimate the lag- h covariances at each time point. Although a Gaussian weighting kernel has been used here, alternative weight functions can be used to adapt the method of smoothing to the nature of the studied time series. Such smoothness of the covariance structure ensures identifying time-varying principal modes of variability for non-stationary FTS that are not cross-correlated at short lags. The time-varying FPCs have shown their ability to approximate the original curves of the highly dynamic system of EpCO_2 using a smaller number of leading PCs, by taking into account the time-varying covariance structure. According to the simulation study, the time-varying FPCs outperform their stationary counterparts in constructing the original curves with a better accuracy, with lower NMSE, in the absence and presence of missing data in non-stationary FTS. The simulation study indicates that differences become more substantial as temporal dependence between curves decreases and system becomes more dynamic. We have also noticed that, given a reasonable degree of smoothness, the estimated eigenvalues of the local spectral

densities provide good estimates for the corresponding true ones. These results are mainly empirical, and future work should involve further investigation of the theoretical properties of those estimators which are discussed in Sect. 4 through simulation but not rigorously proven.

We have also proposed a bootstrap inference procedure to detect changes in the covariance structure of a FTS based on the time-varying frequency-domain FPCs. Although this procedure has been illustrated in details using the EpCO₂ data, it can be applied to another FTS given a suitable null stationary model to simulate the reference distribution. The proposed test has detected significant changes in the covariance and variability structure of the series of EpCO₂ daily curves over time. The test identified time periods where there are shifts in the system dynamics, thus demonstrating the need of non-stationary functional PCA to explain the changing variability structure in such complex systems. The time-varying FPCs highlighted that changes in the EpCO₂ diel patterns are not driven solely by seasonality but also by changes in the system hydrology depending on the strength and extent of correlation in the system across the days. Evaluating these temporal changes in the diel patterns of EpCO₂ is important for a better understanding of the carbon cycle and for providing reliable estimates of the carbon budgets in freshwater systems.

Possible extensions include multivariate time-varying dynamic functional principal component analysis, following the new idea by [Happ and Greven \(2018\)](#). Such an approach would focus on multivariate functional data, where correlation does not only exist between the curves of one FTS but also across different FTS, not necessarily defined on the same domain or the same number of dimensions. This, in turn, could help to better understand the dynamics in EpCO₂ in relation to variability in other series like the flow velocity in the river ([Long et al. 2015](#)). Another challenge that is currently being under consideration is to reduce the computational cost of applying the time-varying FPCs to a longer FTS.

Environmental data often exhibit dependence and non-stationarity over both time and space. A paper by [King et al. \(2018\)](#) proposed a new approach to study the variability of air pollution over time and across space using FDA techniques. In this method, the yearly profiles at a certain location represented a FTS and modelled as a function of: (1) an annual mean level, (2) a linear combination of smooth annual trends with site-specific coefficients that vary over time and (3) an annual specific residual effect. The second component is based on the eigenfunctions of the pooled covariance, obtained by ignoring the dependence across space and years, with coefficients that vary over time and space according to a space–time covariance model. This raises interest in developing FPCs that appropriately reduce the data dimensionality while accounting for dependence and non-stationarity in time and space.

SUPPLEMENTARY MATERIAL

The supplementary material available online contains an Appendix for some technical definitions and details of FTS and frequency-domain methods, in addition to more results and figures of the simulation studies and application to EpCO₂ data. The R code necessary for the computation of the time-varying frequency-domain FPCs is also available. The folder

freedom by [Hormann and Kidzinski \(2017\)](#) is a dependency package for this work. New and modified functions are available in the folder `smfreedom`.

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/>.

[Received June 2020. Revised November 2021. Accepted February 2022. Published Online March 2022.]

REFERENCES

- Brillinger DR (1981) Time series: data analysis and theory, 1st edn. Holden-Day, San Francisco
- Butman D, Raymond PA (2011) Significant efflux of carbon dioxide from streams and rivers in the United States. *Nat Geosci* 4:839–842
- Cole JJ, Prairie NF, Caraco NF, McDowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA (2007) Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* 10:172–185
- Dawson J, Soulsby C, Hrachowitz M, Telzlauff D (2009) Seasonality of EpCO_2 at different scales along an integrated river continuum within the Dee Basin, NE Scotland. *Hydrol Process* 23(20):2929–2942
- Dubey P, Muller H (2019) Functional models for time-varying random objects. arXiv preprint [arXiv:1907.10829](https://arxiv.org/abs/1907.10829)
- Elayouty A, Scott M, Miller C, Waldron S, Franco-Villoria M (2016) Challenges in modeling detailed and complex environmental data sets: a case study modeling the excess partial pressure of fluvial CO_2 . *J Environ Ecol Stat* 23(1):65–87
- Halliday SJ, Wade AJ, Skeffington RA, Neal C, Reynolds B, Rowland P, Neal M, Norris D (2012) An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales. *Sci Total Environ* 434:186–200
- Happ C, Greven S (2018) Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J Am Stat Assoc* 113(522):649–659
- Hormann S, Kidzinski L, Hallin M (2015) Dynamic functional principal components. *J R Stat Soc Ser B* 77(2):319–348
- Hormann S, Kidzinski L (2017) Package `freedom` (Frequency Domain Based Analysis: Dynamic PCA) The R Project for Statistical Computing
- Horváth L, Kokoszka P, Rice G (2014) Testing stationarity of functional time series. *J Econom* 179(1):66–82
- Jaimungal S, Eddie K (2007) Consistent functional PCA for financial time series. In: FEA 07 proceedings of the fourth IASTED international conference on financial engineering and applications, pp 103–108
- King MC, Staicu A-M, Davis JM, Reich BJ, Eder B (2018) A functional data analysis of spatiotemporal trends and variation in fine particulate matter. *Atmos Environ* 184:233–243
- Kokoszka P, Jouzdani NM (2020) Frequency domain theory for functional time series: Variance decomposition and an invariance principle. *Bernoulli* 26(3):2383–2399
- Ku W, Storer RH, Georgakis C (1995) Disturbance detection and isolation by dynamic principal component analysis. *Chemom Intell Lab Syst* 30:179–196
- Long H, Vihermaa L, Waldron S, Hoey T, Quemin S, Newton J (2015) Hydraulics are a first-order control on CO_2 efflux from fluvial systems. *J Geophys Res Biogeosci* 120:1912–1922

- Mardia K, Kent JT, Bibby J (1979) *Multivariate analysis*. Academic Press, New York
- Melnikov O, Raun LH, Ensor KB (2016) Dynamic principal component analysis: identifying the relationship between multiple air pollutants arXiv preprint [arXiv:1608.03022](https://arxiv.org/abs/1608.03022)
- Miller C, Bowman A (2012) Smooth principal components for investigating changes in covariance over time. *J R Stat Soc Ser C* 61(5):693–714
- Ombao M, Ringo H (2006) Time-dependent frequency domain principal component analysis of multichannel non-stationary signals. *Comput Stat Data Anal* 50:2339–2360
- Panaretos VM, Tavakoli S (2013) Cramer–Karhunen–Loeve representation and harmonic principal component analysis of functional time series. *Stoch Process Appl* 123:2779–2807
- Pena D, Yohai VJ (2016) Generalized dynamic principal components. *J Am Stat Assoc* 111(515):1121–1131
- Petersen A, Deoni S, Müller H (2019) Frechet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. *Ann Appl Stat* 13(1):393–419
- Politis DN (2011) Higher order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econom Theory* 27:703–744
- Ramsay JO, Silverman BW (1997) *Functional data analysis*, 1st edn. Springer, New York
- Raymond PA, Caraco NF, Cole JJ (1997) Carbon dioxide concentration and atmospheric flux in the Hudson River. *Estuaries* 20:381–390
- Reiman JH, Xu YJ (2019) Diel variability of $p\text{CO}_2$ and CO_2 outgassing from the lower Mississippi River: implications riverine CO_2 outgassing estimation. *Water* 11(43):1–15
- Richey JE (2013) Pathways of atmospheric CO_2 through fluvial systems. In: *Scientific Committee on Problems of the Environment (SCOPE)/United Nations Environmental Programme (UNEP)- The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World*, Island Press, Washington, DC, USA
- Rodrigues PC (2007) Principal component analysis of dependent data. 15th European Young Statisticians Meeting 1–5
- Ullah S, Finch CF (2013) Applications of functional data analysis: a systematic review. *BMC Med Res Methodol* 13(43):1–12
- Vanhatalo E, Kulahci M, Bergquist B (2017) On the structure of dynamic principal component analysis used in statistical process monitoring. *Chemom Intell Lab Syst* 167:1–11
- Waldron S, Scott M, Soulsby C (2007) Stable isotope analysis reveals lower-order river dissolved inorganic carbon pools are highly dynamic. *Environ Sci Technol* 41(17):6156–6162
- Wang J, Chiou J, Muller H (2015) *Functional data analysis*. *Annu Rev Stat Appl* 3:257–295
- Wang X, He Y, Yuan X, Chen H, Peng C, Zhu Q, Yue J, Ren H, Deng W, Liu H (2017) $p\text{CO}_2$ and CO_2 fluxes of the metropolitan river network in relation to the urbanisation of Chongqing, China. *J Geophys Res Biogeosci* 122:470–486
- Yan F, Lin X, Li R, Huang X (2018) Functional principal components analysis on moving time windows of longitudinal data: dynamic prediction of times to event. *J R Stat Soc Ser C*