Functional principal component analysis for non-stationary dynamic time series

Amira Elayouty[1], Marian Scott[2], Claire Miller[2], Susan Waldron[3]

[1] Faculty of Economics and Political Science, Cairo University, Egypt
[2] School of Mathematics and Statistics, University of Glasgow, UK
[3] School of Geographical and Earth Sciences, University of Glasgow, UK

E-mail for correspondence: a.ayouti@feps.edu.eg

Abstract: Motivated by a highly dynamic hydrological high-frequency time series, we propose time-varying Functional Principal Component Analysis (FPCA) as a novel approach for the analysis of non-stationary Functional Time Series (FTS) in the frequency domain. Traditional FPCA does not take into account (i) the temporal dependence between the functional observations and (ii) the changes in the covariance/variability structure over time, which could result in inadequate dimension reduction. The novel time-varying FPCA proposed adapts to the changes in the auto-covariance structure and varies smoothly over frequency and time to allow investigation of whether and how the variability structure in an FTS changes over time. Based on the (smooth) time-varying dynamic FPCs, a bootstrap inference procedure is proposed to detect significant changes in the covariance structure over time. Although this time-varying dynamic FPCA can be applied to any dynamic FTS, it has been applied here to study the daily processes of partial pressure of CO_2 in a small river catchment in Scotland.

Keywords: Functional Time Series; Frequency Domain; Smoothing; Principal Components; Non-stationarity; Functional spectral density

1 Introduction

Recent advances in sensor technology allow environmental monitoring programs to record measurements at high-temporal resolutions over long time periods, for processes which are in reality continuous in time. These High-Frequency Data (HFD) pose several challenges in terms of statistical modeling and analysis due to the complexity of such large volumes of data stemming from the persistent and dynamic dependence structure over the

---

different timescales (Elayouty et al. 2016). Functional Time Series (FTS) analysis and its recent developments (Hörmann et al. 2015) provide an appropriate framework for analyzing such HFD, taking into consideration these technical challenges.

This paper introduces a novel approach that identifies and accounts for volatility in FTS. This approach involves the development of frequency domain Functional Principal Components (FPCs) that vary smoothly over time, taking into account both temporal correlation and non-stationarity in the series. Using bootstrap procedures, these time-varying FPCs are employed to statistically assess changes over time in the covariance structure and variability modes of the underlying FTS process.

This work is motivated by highly dynamic HFD of excess partial pressure of carbon dioxide ($EpCO_2$) measured every 15 minutes over 3 years at a small catchment of the River Dee, Scotland. The long sequence of 15-minute measurements is segmented into daily intervals, which are then smoothed using B-splines to form a sequence of daily $EpCO_2$ functions. The data thus form a FTS and are viewed as realizations of a functional stochastic process $\{X_k(t) : k \in \mathbb{Z}, t \in \mathcal{T}\}$ valued in the Hilbert space $L^2(\mathcal{T})$, with k denoting the day as a discrete time parameter and t being the intra-day time defined continuously on $\mathcal{T}$.

2 Methodology

The proposed methodology relies on evaluating the Spectral Density (SD) of the FTS process $\{X_k\}$ at each time point k and obtaining the dynamic FPCs (Hörmann et al. 2015) via the eigen-decomposition of the SD at each time point k, assuming that the process varies smoothly over time. Because the SD contains information on the whole family of lag-h covariances, the novel FPCs accommodate the varying serial correlation. Due to the limited number of replicates at each time point k, the local lag-h covariances and spectral densities are computed by smoothing the sample lag-h covariances over time using a weight kernel $w_s(.)$ with smoothing parameter s,

$$\hat{V}_{k,h} = \frac{1}{\sum_{k' \in \mathbb{Z}} w_s(|k-k'|)} \sum_{k \in \mathbb{Z}} w_s(|k-k'|) X_{k'} \otimes X_{k'+h}. \tag{1}$$

$w_s(.)$ is a monotonically decreasing weight function of the distance $|k-k'|$ regardless of the lag h, ensuring that the highest weights are assigned to the pairs $(X_{k'}, X_{k'+h})$ near the target point k. The neighborhood contributing to the covariance estimation is determined by the choice of the kernel and smoothing parameter. The choice of weight kernel is based on the nature of the variable of interest; a common choice is the Gaussian density. The smoothing parameter is chosen so that the process within each neighborhood is stationary without over-fitting the original process.

After computing $\hat{V}_{k,h}$, the local SD is estimated at each time point k by:

$$\hat{F}_{k,\theta} = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \hat{V}_{k,h} \exp(-\mathrm{i}h\theta), \quad \theta \in [-\pi, \pi], \tag{2}$$

and the local eigenvalues $\hat{\lambda}_{k,m}(\theta)$ and eigenvectors $\hat{\varphi}_{k,m}(\theta)$ of $\hat{F}_{k,\theta}$ are calculated. The local functional filters $\{\hat{\phi}_{kml}(t) : l \in \mathbb{Z}\}$ are estimated, via the inverse Fourier transform of $\hat{\varphi}_{k,m}(\theta)$, and subsequently used to filter the original FTS across a number of lags and leads l to obtain the m^{th} local dynamic FPC scores at k as: $\hat{Y}^{(k)}_{m,k} = \sum_{l=-L}^{L} \int_{t \in \mathcal{T}} X_{k-l}(t) \hat{\phi}_{kml}(t) dt$.
The original curves can thus be approximately reconstructed, based on these scores, using q-term (smooth) time-varying dynamic FPCs, $q < \infty$, as follows:

$$\hat{X}_k(t) \approx \sum_{m=1}^{q} \sum_{l=-L}^{L} \hat{Y}^{(k)}_{m,k+l} \hat{\phi}_{kml}(t), \quad \forall k. \tag{3}$$

Based on the frequency domain local eigenvalues corresponding to the leading smooth dynamic FPCs, we propose a covariance stationarity test of FTS. The null hypothesis of the test, that the SD of the FTS does not vary throughout time, is investigated by evaluating whether the changes over time in the eigenvalues of the process SD are consistent with sampling variation. Following Miller and Bowman (2012), we propose a test statistic based on comparing the m^{th} eigenvalue of the local SD $\hat{\lambda}_{k,m}(\theta)$ and the corresponding eigenvalue of the global SD obtained for the full FTS $\hat{\lambda}_m(\theta)$, both averaged over all frequencies θ, at each time point. Using the parametric bootstrap, the null distribution of the test statistic is constructed which is then used to produce a point-wise reference band highlighting where in time there are significant deviations from the null hypothesis.

3 Simulation Study

To assess the performance of our proposed time-varying dynamic FPCs versus the stationary dynamic FPCs proposed by Hörmann et al. (2015) in approximating the original process, a simulation study for a variety of non-stationary data-generating processes was conducted. The simulation study was designed to mimic the $EpCO_2$ data presented in this paper.
Firstly, a FTS $\{X_k\}$ of 400 observations is generated from a functional autoregressive of order 1, FAR(1). This simulation, in practice, is performed in a finite dimension p, using the basis expansion representation of the functions. The coefficients for the p basis functions $\mathbf{z}_k = (z_{k1}, \ldots, z_{kp})^\top$ associated with the functions $X_k(t) : k = 1, ..., 400$ are simulated according to the vector auto-regressive of order 1: $\mathbf{z}_{k+1} = \mathbf{R}\mathbf{z}_k + \boldsymbol{\epsilon}^*_{k+1}$; where $\mathbf{R}$ is the matrix of auto-regressive parameters whose norm defines the level of time dependence in the data and $\boldsymbol{\epsilon}^*_{k+1}$ are i.i.d normally distributed noise with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. This simulation is performed

based on the estimates of $\mathbf{R}$ and $\mathbf{\Sigma}$ obtained for the $EpCO_2$ data, assuming that the covariance structure does not vary with time.
To construct a non-stationary FTS with a covariance structure that changes over time, an ordinary FPCA is performed on the FTS simulated above from the FAR(1) to obtain the eigenvalues $(\lambda_1, \ldots, \lambda_p)$ and corresponding eigenfunctions $(E_1(t), \ldots, E_p(t))$. These eigenvalues are then used to produce a p-dimensional vector of eigenvalues $\boldsymbol{\lambda}_b$ such that both absolute and relative variance of the FPCs vary smoothly over a grid of B time blocks. By naturally extending the work of Mardia et al. (1979) to a functional context, a functional process $\{X_k\}$ can be constructed by using:

$$X_k(t) = \bar{X}(t) + \sum_{m=1}^{p} S_{mk} E_m(t), \quad t \in \mathcal{T}, k = 1, \ldots, N \tag{4}$$

where $\bar{X}(t)$ is the functional mean, S_{mk} is the score of the m^{th} FPC for the k^{th} observation and $E_m(t)$ is the m^{th} eigenfunction. Based on this result, a sequence of locally stationary functions is generated by simulating at $b = 1, \ldots, B$, blocks of N/B p-dimensional vectors of PC scores from a VAR(1) with a pre-specified level of dependence ρ and normally distributed noise with mean $\mathbf{0}$ and variance-covariance matrix $\text{diag}(\boldsymbol{\lambda}_b)$. This provides $N = 400$ functions, where N/B functions share the same covariance structure.
The above simulation procedure is repeated 200 times for different choices of ρ, reflecting weak to strong levels of dependence in the data. For each simulated non-stationary FTS $\{X_k\}$, we compute the stationary dynamic FPCs and the novel time-varying dynamic FPCs using the values $s = 10, 20, 40$ and 100 for the smoothing parameter and estimate the corresponding scores, as per the methodology described in Section 2. These quantities are then used to recover the approximating series $\{\hat{X}_k(t)\}$ using $q = 1, 2$ and 3 components. The performances of these approximations are then measured in terms of the normalized mean squared errors, NMSE, computed as:

$$\sum_k \|X_k(t) - \hat{X}_k(t)\|^2. \tag{5}$$

Due to space limitations, we only present here the simulation results for $\rho = 0.1$ and 0.9. It is evident from Fig 1 that the (smooth) time-varying dynamic FPCs outperform the stationary dynamic FPCs in terms of NMSE. The performance of both methods improves as the dependence level in the data increases. This is a result of both methodologies accounting for the correlation structure in the data. However, as we may expect, the differences become more striking as the dependence level in the data decreases and the series becomes more dynamic. It is also noticed that the differences become negligible as the number of components q used in the reconstruction increases and that using a smaller value for the smoothing parameter s provides better approximations which deteriorate as q increases. This justifies the trade-off between the smoothness of the FPCs over time and over-fitting.

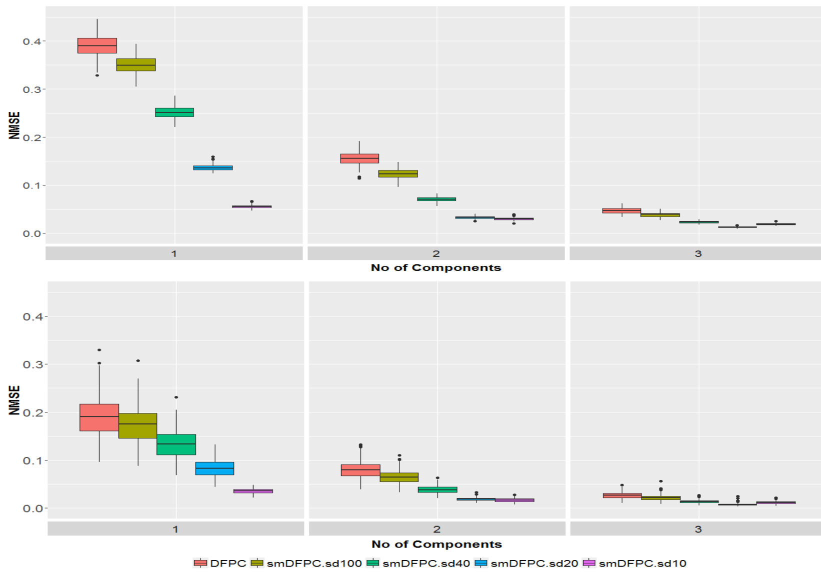

FIGURE 1. Box-plots of the NMSE between the simulated curves and their reconstructed versions using $q = 1, 2$ and 3 (from left to right) dynamic (red) and smooth time-varying dynamic FPCs with $s = 100$ (olive), 40 (green), 20 (blue) and 10 (purple), computed for 200 non-stationary simulation runs with $\rho = 0.1$ (top) and 0.9 (bottom).

4 Results and Discussion

The simulation results indicated that the novel non-stationary FPCs outperform their stationary counterparts in approximating the original process curves and simplifying the complexity of data in almost all settings and that improvements are more obvious as temporal dependence between curves weakens and the system becomes more dynamic.

The time-varying dynamic FPCs are used to investigate and assess the dynamics and variability structure in the daily smooth profiles of $EpCO_2$ over time. The novel FPCs better approximate the original pattern as well as the within-day variability (Fig 2); the first time-varying dynamic FPC captured 94% of the variability. The proposed stationarity test identified significant changes in the covariance structure over frequency and time. The $EpCO_2$ system appears to involve high correlation throughout summer and winter, when the $EpCO_2$ daily pattern is ruled by biological activity. In transitional periods between summer and winter, the system is mostly determined by hydrological activity and therefore exhibits more variability responding to hydrological events like heavy storms or rainfall.

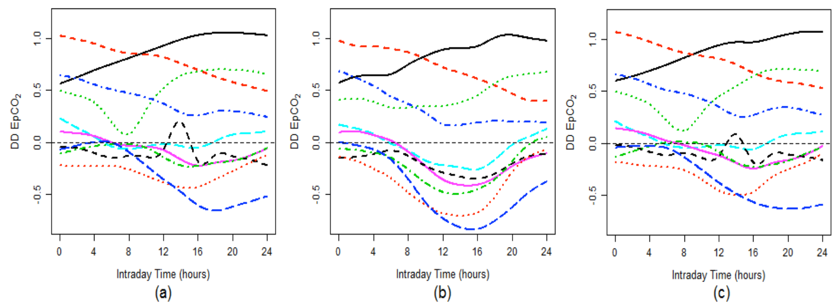

FIGURE 2. (a) 10 successive daily curves of de-trended $EpCO_2$ and the corresponding reconstructions based on (b) the first dynamic FPC and (c) the first time-varying dynamic FPC using $s = 20$ (chosen based on a sensitivity analysis)

5 Conclusion

Time-varying (smooth) FPCA proved to be an appropriate tool for reducing dimensionality, extracting the most important characteristics and simplifying the complexity in the variability structure of high-dimensional non-stationary time series. Traditional methods of FPCA may ignore the time dependence and non-stationarity in FTS and hence the variability modes identified may be biased.

Acknowledgments: A.E. is grateful to the University of Glasgow sensor studentship for funding. Authors are grateful to Susan Waldron for providing $EpCO_2$ data profiles generated through NERC Fellowship, NER/J/S/2001/00793.

References

Elayouty, A., Scott, M., Miller, C., Waldron, S. and 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. *Journal of Environmental and Ecological Statistics*, **23(1)**, 65 – 87.

Hörmann, S., Kidzinski, L. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society, Series B*, **77(2)**, 319 – 348.

Mardia, K. and Kent, J.T. and Bibby, J. (1979). *Multivariate Analysis.* New York: Academic Press.

Miller, C. and Bowman, A. (2012). Smooth principal components for investigating changes in covariance over time. *Journal of the Royal Statistical Society, Series C*, **61(5)**, 693 – 714.