The Maximum of the Periodogram of a Sequence of Functional Data

Vaidotas Characiejus[a]

Joint work with Clément Cerovecki[b] and Siegfried Hörmann[c]

Department of Economics, UC3M, Spain, September 11, 2023

[a]Department of Mathematics and Computer Science, University of Southern Denmark, Denmark

[b]Département de mathématique, Université libre de Bruxelles, Belgium

[b]Department of Mathematics, Katholieke Universiteit Leuven, Belgium

[c]Institute of Statistics, Graz University of Technology, Austria

Periodic signals

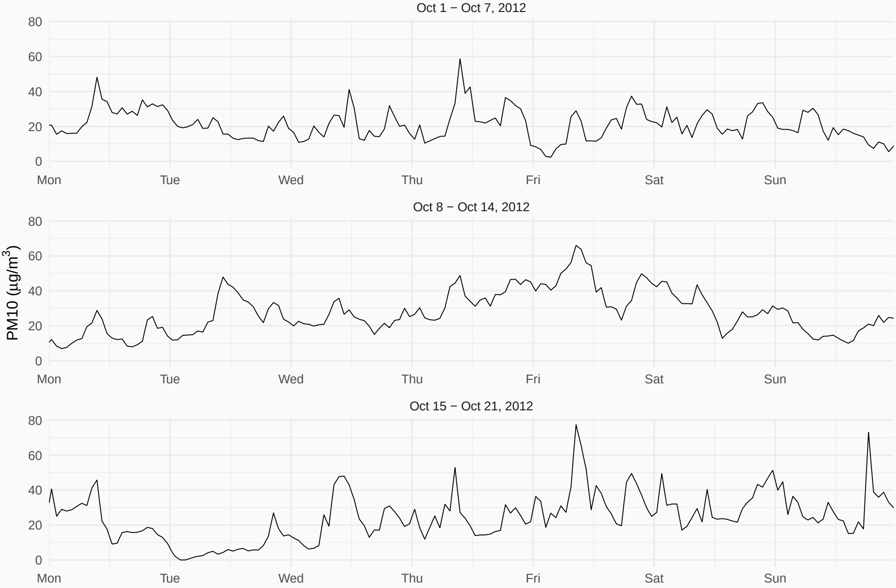
- The focus of the talk is detection, analysis and estimation of periodic signals in a sequence of functional data.
- Periodicities are one of the most important characteristics of time series.
- The interest in periodicities goes back to the origins of the field (Schuster [1898], Walker [1914], Yule [1927], Fisher [1929], etc.).

Motivation and problem

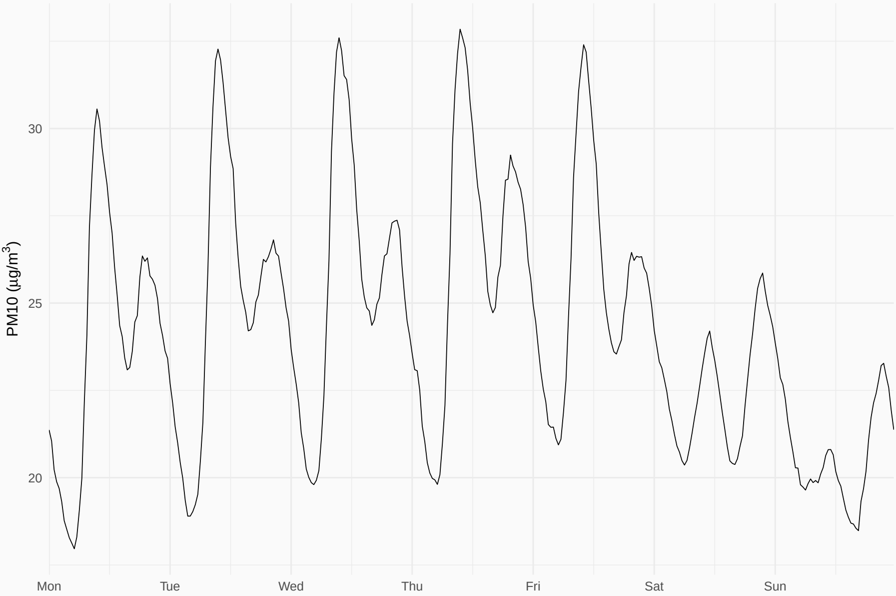
PM10 data

- Air quality data from Graz, Austria.
- The amount of particulate matter with a diameter of 10 µm or less (PM10) is measured.
- PM10 can settle in the bronchi and lungs and cause health problems.
- Starting on February 18, 2010, the amount of PM10 in µg/m^3 is recorded every 30 minutes resulting in 48 observations per day.

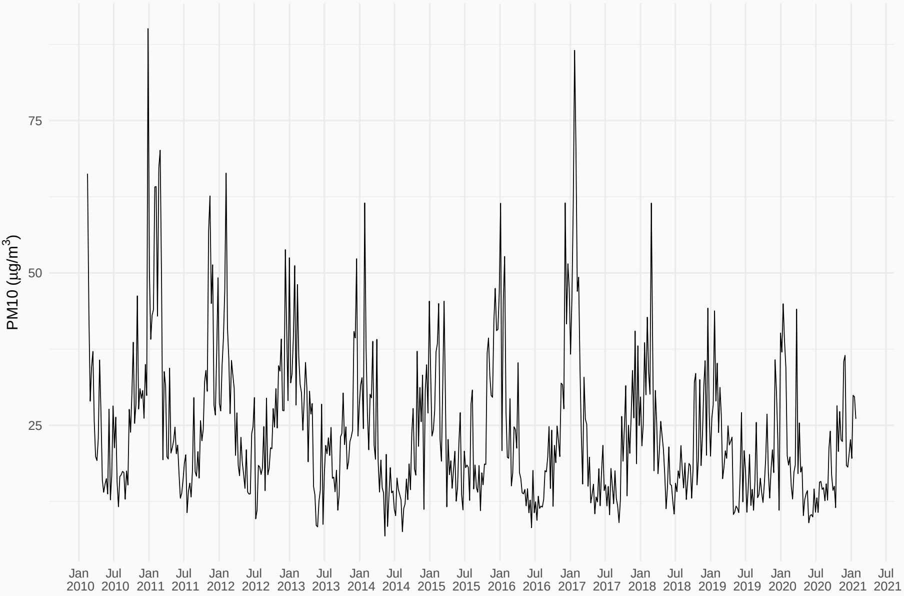
Raw data


Oct 1 − Oct 7, 2012
Oct 8 − Oct 14, 2012
Oct 15 − Oct 21, 2012
PM10 (μg/m^3)
80
60
40
20
0
Mon
Tue
Wed
Thu
Fri
Sat
Sun

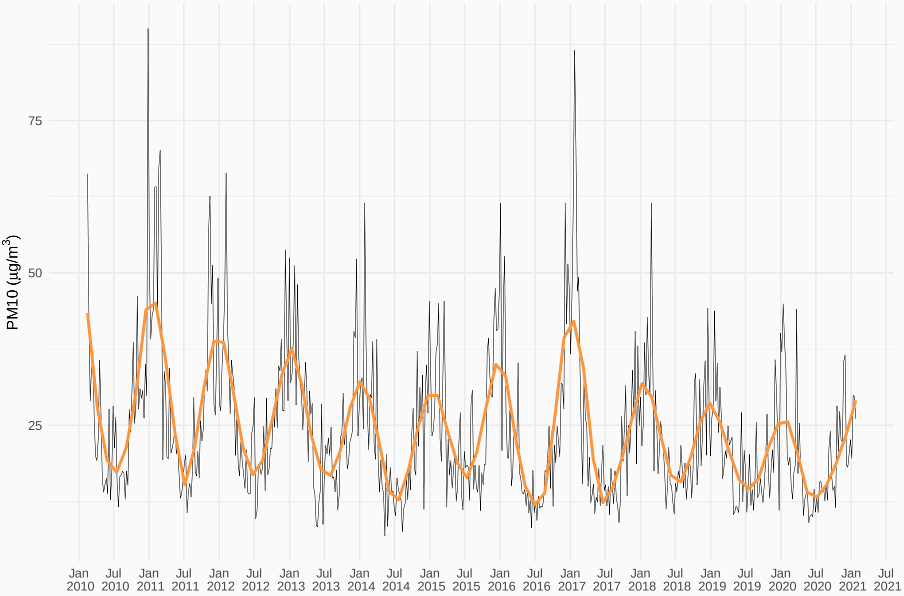
Weekly mean curve



Weekly averages



Weekly averages



Functional time series

- We investigate the PM10 data as a functional time series, i.e., as a sequence of daily curves.
- A functional time series is a sequence $\{X_t\}_{t \in \mathbb{Z}}$ such that each X_t is a curve $\{X_t(u)\}_{u \in [0,1]}$.
- We separate a continuous time process $\{\xi(u)\}_{u \in \mathbb{R}}$ using natural consecutive intervals, i.e.

$$X_t(u) = \xi(t + u)$$

  for $u \in [0, 1]$ and $t \in \mathbb{Z}$.
- Such segmentation accounts for a periodic structure in the underlying continuous time process.
- There might still remain some periodic signal with respect to the discrete time parameter $t \in \mathbb{Z}$.

Model

$\{X_t\}_{t\in\mathbb{Z}}$ is a time series with values in a real separable Hilbert space $\mathbb{H}$ (e.g. $\mathbb{R}^d$ with $d \geq 1$, $L^2[0,1]$, etc.) defined by

$$X_t = \mu + s_t + Y_t$$

for each $t \in \mathbb{Z}$, where

- $\mu \in \mathbb{H}$;
- $\{s_t\}_{t\in\mathbb{Z}} \subset \mathbb{H}$ is a deterministic sequence such that

$$s_t = s_{t+T} \quad \text{and} \quad \sum_{t=1}^{T} s_t = 0$$

  for all $t \in \mathbb{Z}$ with some $T \geq 2$;
- $\{Y_t\}_{t\in\mathbb{Z}}$ is a stationary sequence of zero mean random elements with values in $\mathbb{H}$.

Hypothesis testing

We develop a methodology to test

$$H_0 : X_t = \mu + Y_t \quad \text{versus} \quad H_1 : X_t = \mu + s_t + Y_t$$

with an unknown $T \geq 2$.

Remark about T

- In practice, T can be assumed to be known or unknown depending on the particular situation.
- In many situations, the potential periodic signal is, for example, daily, weekly, monthly, or yearly.
- Even if T is known, it is still of interest to determine whether the periodic signal can be modelled using a single sinusoid or it has to be modelled by a superposition of several sinusoids.
- In some situations, it is very difficult to determine what the value of T could be (for example, solar cycles have an average duration of about 11 years).

Test statistic

Frequency domain approach

Our methodology is based on the frequency domain approach to the analysis of functional time series.

DFT and periodogram

Definition

The discrete Fourier transform (DFT) of $X_1, \ldots, X_n$ is defined by

$$\mathcal{X}_n(\omega_j) = n^{-1/2} \sum_{t=1}^{n} X_t e^{-it\omega_j}$$

for $n \geq 1$, where $\omega_j = 2\pi j/n$ with $j \in F_n = \{-\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor\}$ are the Fourier frequencies and $i = \sqrt{-1}$.

Definition

The periodogram operator of $X_1, \ldots, X_n$ is defined by

$$I_n(\omega_j) = \mathcal{X}_n(\omega_j) \otimes \mathcal{X}_n(\omega_j) = \langle \cdot, \mathcal{X}_n(\omega_j) \rangle \mathcal{X}_n(\omega_j)$$

for $n \geq 1$, where $\omega_j = 2\pi j/n$ with $j \in F_n$ are the Fourier frequencies.

Maximum of periodogram

The test statistic is given by

$$M_n = \max_{1 \le j \le q} \|I_n(\omega_j)\|_{op} = \max_{1 \le j \le q} \|\mathcal{X}_n(\omega_j)\|^2$$

for $n > 2$, where

(i) $\omega_j = 2\pi j/n$ with $1 \le j \le q = \lfloor n/2 \rfloor$;

(ii) $\| \cdot \|_{op}$ is the operator norm and $\| \cdot \|$ is the norm of the complexification of $\mathbb{H}$.

Intuition

Why the maximum of the periodogram?

Orthonormal basis for $\mathbb{C}^n$

- The vectors

$$e_j = n^{-1/2}\begin{pmatrix} e^{i\omega_j} & e^{i2\omega_j} & \dots & e^{in\omega_j} \end{pmatrix}'$$

  with $\omega_j = 2\pi j/n$ and $j \in F_n$ constitute an orthonormal basis for $\mathbb{C}^n$.
- Recall Euler's formula $e^{ix} = \cos x + i \sin x$ for $x \in \mathbb{R}$.
- For $x \in \mathbb{C}^n$,

$$x = \sum_{j \in F_n} a_j e_j,$$

  where

$$a_j = \langle x, e_j \rangle = n^{-1/2} \sum_{t=1}^{n} x_t e^{-it\omega_j}$$

  is the DFT of x at the frequency ω_j with $j \in F_n$.

Representation of periodic signals

Lemma

Suppose that $\{s_t\}_{t\in\mathbb{Z}}$ is a deterministic sequence with values in $\mathbb{H}$ such that $s_t = s_{t+T}$ and $\sum_{t=1}^{T} s_t = 0$ for all $t \in \mathbb{Z}$ with some $T \geq 2$. Then there exist $w_{11}, \dots, w_{1\lfloor T/2\rfloor} \in \mathbb{H}$ and $w_{21}, \dots, w_{2\lfloor T/2\rfloor} \in \mathbb{H}$ such that

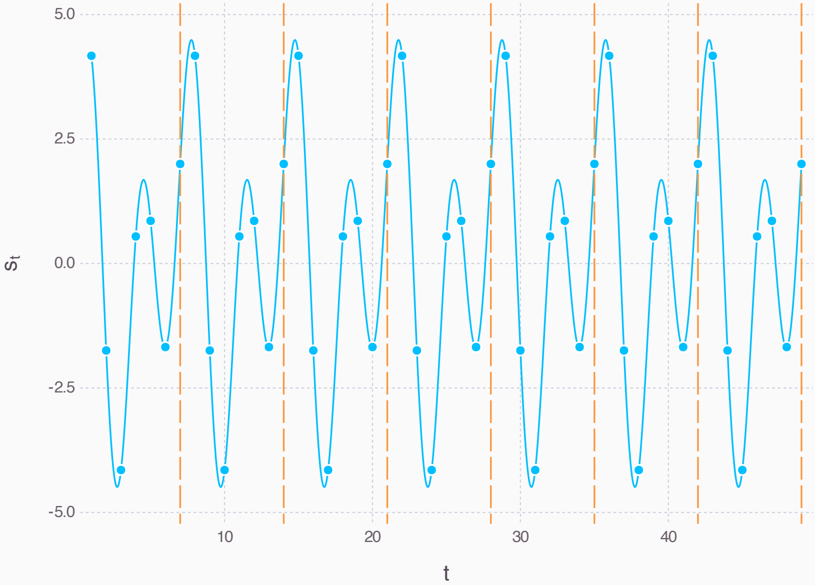
$$s_t = \sum_{k=1}^{\lfloor T/2\rfloor} \left[\cos\left(\frac{2\pi kt}{T}\right) w_{1k} + \sin\left(\frac{2\pi kt}{T}\right) w_{2k}\right]$$

for all $t \in \mathbb{Z}$. If, in addition, $n = Tm$, then

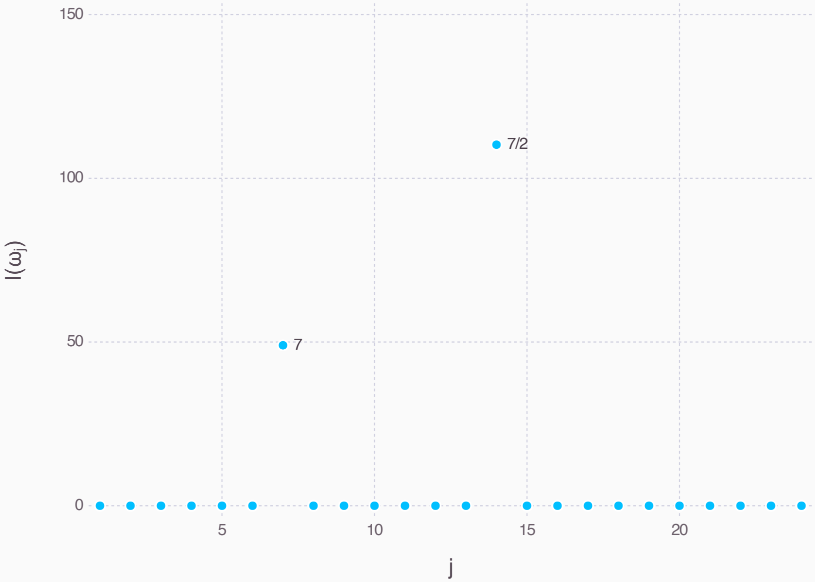
$$\mathcal{S}_n(\omega_j) = n^{-1/2} \sum_{t=1}^{n} s_t e^{-it\omega_j} = \begin{cases} n^{1/2}(w_{1k} - iw_{2k})/2, & j = km, \\ 0, & j \neq km, \end{cases}$$

where $k = 1, \dots, \lfloor T/2\rfloor$.

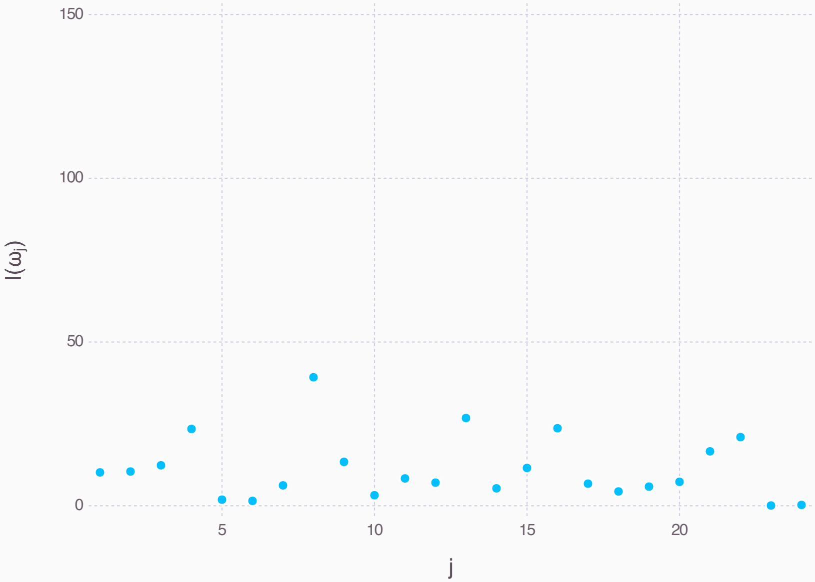
Periodic signal ($T = 7$, $w_{11} = 2$, $w_{22} = 3$, $n = 49$)



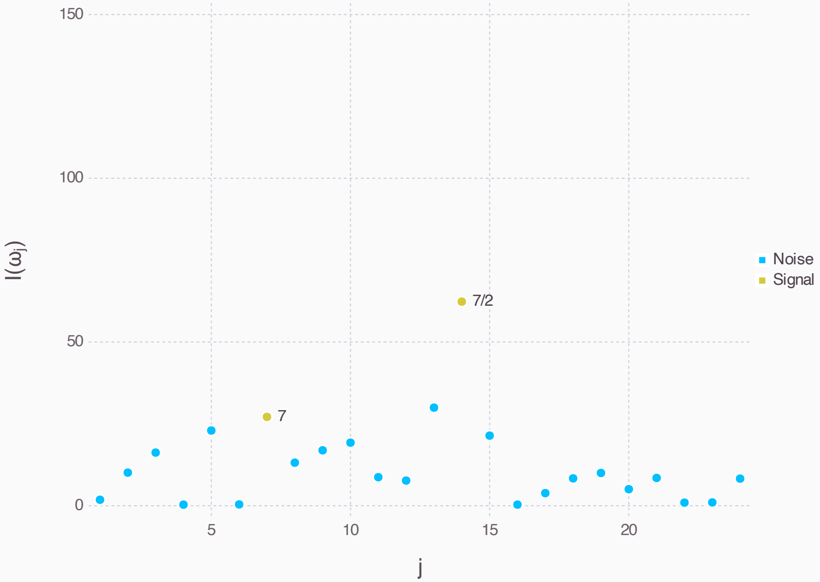
Periodogram of periodic signal ($T = 7$, $w_{11} = 2$, $w_{22} = 3$, $n = 49$)



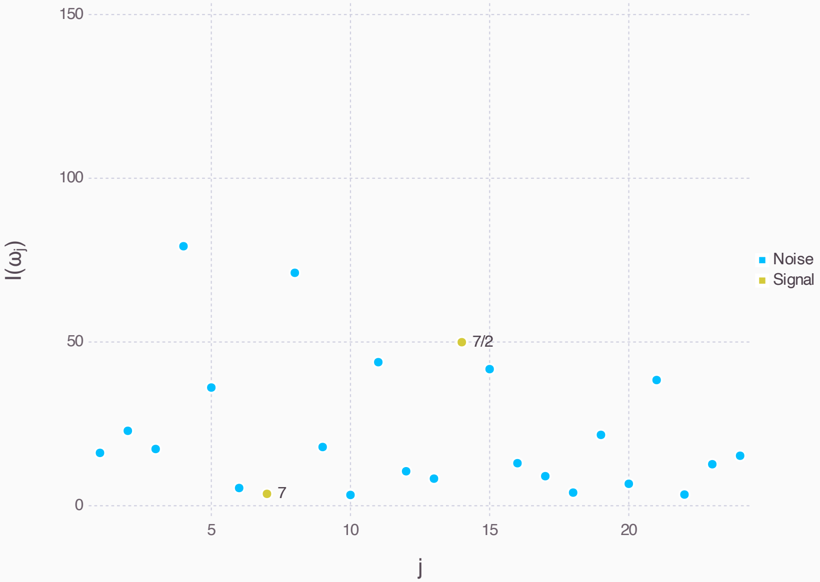
Periodogram of $N(0, 10)$ white noise



Periodogram of periodic signal plus $N(0, 10)$ white noise



Periodogram of periodic signal plus $N(0, 25)$ white noise



Maximum of periodogram

The test statistic is given by

$$M_n = \max_{1 \leq j \leq q} \|\mathcal{X}_n(\omega_j)\|^2$$

for $n > 1$.

- Small values of M_n indicate that there is no periodic component.
- Large values of M_n indicate that there is a periodic component.
- We need a criterion to decide when M_n is small and when M_n is large.

Main results

Results under Gaussianity in the univariate case

- The usefulness of the maximum of the periodogram for detecting periodicities is well known (Fisher [1929]).
- First results were established under the assumption of Gaussianity.
- An alternative approach is to establish the asymptotic distribution of the appropriately standardized M_n under some general conditions.

Results under Gaussianity in the univariate case (cont.)

If $X_1, \ldots, X_n$ are iid standard normal random variables,

$$M_n - \log q \xrightarrow{d} G \quad \text{as} \quad n \to \infty,$$

where $q = \lfloor n/2 \rfloor$ and G is the standard Gumbel distribution with the CDF given by

$$F(x) = \exp\{-\exp^{-x}\}$$

for $x \in \mathbb{R}$.

General results in the univariate case

- Walker [1965] conjectured that the same result holds provided that the moments up to some sufficiently high order exist.
- Walker [1965] also stated that no proof was known at the time and that the problem of constructing one is undoubtedly extremely difficult.
- Davis and Mikosch [1999] proved that the limit indeed remains the same provided that $E|X_1|^s < \infty$ with some $s > 2$ using a Gaussian approximation technique due to Einmahl [1989].

Our results

- Our main result is an extension of the result of Davis and Mikosch [1999] to real separable Hilbert spaces.
- The main ingredient of our proof is a powerful Gaussian approximation developed by Chernozhukov, Chetverikov, and Kato [2017].
- Our results allow us to propose several methodologies to detect periodic signals in Hilbert space valued time series when the length of the period is unknown.

Suppose that $\{Y_t\}_{t\in\mathbb{Z}}$ is a linear process with values in $\mathbb{H}$ given by

$$Y_t = \sum_{k=-\infty}^{\infty} a_k(\varepsilon_{t-k})$$

for each $t \in \mathbb{Z}$, where

- $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are iid zero mean random elements with values in $\mathbb{H}$;
- $\{a_k\}_{k\in\mathbb{Z}} \subset L(\mathbb{H})$.

Assumptions

Assumption 1

i) $E\|\varepsilon_0\|^r < \infty$ where $r > 2$ if $\dim \mathbb{H} < \infty$ and $r \geq 4$ otherwise;

ii) the eigenvalues λ_k of $E[\varepsilon_0 \otimes \varepsilon_0]$ are distinct and the sequence $\{k\lambda_k\}_{k\geq 1}$ is ultimately non-increasing;

iii) some technical conditions on the decay rate of $\{\lambda_k\}_{k\geq 1}$.

Assumption 2

i) $\sum_{k\neq 0} \log(|k|)\|a_k\| < \infty$;

ii) $A^{-1}(\omega)$ exists for each $\omega \in [-\pi, \pi]$, where $A(\omega) = \sum_{k=-\infty}^{\infty} a_k e^{-ik\omega}$ with $\omega \in [-\pi, \pi]$ is the transfer function;

iii) $\sup_{\omega\in[0,\pi]} \|A^{-1}(\omega)\| < \infty$.

Main result

Theorem

Under H_0 and Assumptions 1 and 2, we have that

$$\lambda_1^{-1}\left(\max_{1\le j\le q}\|A^{-1}(\omega_j)\mathcal{X}_n(\omega_j)\|^2 - b_n\right) \xrightarrow{d} G \quad \text{as} \quad n\to\infty,$$

where

- $A(\omega_j) = \sum_{k=-\infty}^{\infty} a_k e^{-ik\omega_j}$ with $j = 1, \dots, q$;
- $b_n = \lambda_1 \log q - \lambda_1 \sum_{j=2}^{\infty} \log(1 - \lambda_j/\lambda_1)$;
- G is the standard Gumbel distribution with the CDF given by $F(x) = \exp\{-\exp\{-x\}\}$ for $x \in \mathbb{R}$.

FAR(1)

$\{Y_t\}_{t\in\mathbb{Z}}$ is an FAR(1) model given by

$$Y_t = \rho(Y_{t-1}) + \varepsilon_t = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{t-j})$$

for $t \in \mathbb{Z}$ with $\rho \in L(\mathbb{H})$.

Assumption 3

i) There is an $n_0 \geq 1$ such that $\|\rho^{n_0}\| < 1$;

ii) $\hat{\rho}$ is an estimator of ρ such that

$$\|\hat{\rho} - \rho\|_{op} = o_p(1/\tau_n')$$

as $n \to \infty$ with $\tau_n' \geq \log n$.

The transfer function, residuals and their eigenvalues

- $\{\hat{\varepsilon}_k\}_{2 \leq k \leq n}$ are the residuals given by

$$\hat{\varepsilon}_k = X_k - \hat{\rho}(X_{k-1})$$

  for $k = 2, \ldots, n$.
- $\{\hat{\lambda}_j\}_{j \geq 1}$ are the eigenvalues of

$$\frac{1}{n-1} \sum_{k=2}^{n} \hat{\varepsilon}_k \otimes \hat{\varepsilon}_k.$$

- The transfer function and its inverse are given by

$$A(\omega) = (I - e^{-i\omega}\rho)^{-1} \quad \text{and} \quad A^{-1}(\omega) = I - e^{-i\omega}\rho$$

  respectively for $\omega \in [-\pi, \pi]$.

Theorem

Under H_0 and Assumptions 1 and 3,

$$G_n := \hat{\lambda}_1^{-1} \max_{1 \le j \le q} \|(I - e^{-i\omega_j}\hat{\rho})(\mathcal{X}_n(\omega_j))\|^2 - \log q + \max\left\{ \sum_{j=2}^{\tau_n} \log(1 - \hat{\lambda}_j/\hat{\lambda}_1), c_n \right\} \xrightarrow{d} \mathcal{G}$$

as $n \to \infty$, where $\{\tau_n\}_{n \ge 1} \subset \mathbb{N}$ and $\{c_n\}_{n \ge 1} \subset \mathbb{R}$ are sequences that satisfy certain technical conditions.

Consistency

Theorem

Under H_1,

$$G_n/\ell_n \xrightarrow{p} \infty \quad as \quad n \to \infty$$

for any positive sequence $\ell_n = o(n)$ as $n \to \infty$ provided certain technical conditions are satisfied.

Empirical study

Simulation setting

- We simulate functional time series that are stationary and behaves similarly as the original PM10 data.
- The periodic component in the simulation study is given by

$$s_t(u) = a\cos(2\pi t/d),$$

  where $u \in [0, 1]$ and $d - 2$ is a Poisson distributed random variable P_λ with $\lambda = 5$ or $\lambda = 15$.
- a is equal to 0 (no periodic signal), 1 or 2.

Empirical rejection rates

		$a = 0\ (\equiv H_0)$			$a = 1$			$a = 2$		
	α	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\lambda = 5$	$n = 100$	0.049	0.022	0.004	0.867	0.805	0.670	1.000	0.999	0.994
	$n = 200$	0.074	0.034	0.005	0.990	0.983	0.972	1.000	1.000	1.000
	$n = 500$	0.091	0.052	0.011	1.000	1.000	0.999	1.000	1.000	1.000
$\lambda = 15$	$n = 100$	0.067	0.030	0.004	0.260	0.172	0.072	0.837	0.773	0.629
	$n = 200$	0.069	0.030	0.006	0.585	0.488	0.312	0.987	0.975	0.926
	$n = 500$	0.093	0.044	0.007	0.990	0.979	0.946	1.000	1.000	1.000

Transforming data into curves

- The data is preprocessed in the following way:
    - the missing values are linearly interpolated;
    - the negative values are set to 0 so that the square root transformation can be performed;
    - the raw observations are transformed into curves using the `R` package `fda` and the function `Data2fd()` with 21 Fourier basis functions.
- We use the PCA based estimator of ρ ('Bosq [2000]).
- The tuning parameter k_n which determines the number of principal components used in the estimation procedure is selected so that k_n principal components explain more than 99% of the variance in our dataset.

PM10 time series

- We plot the points $(j, G_n(j))$ with $j = 1, \dots, q = 1998$ and

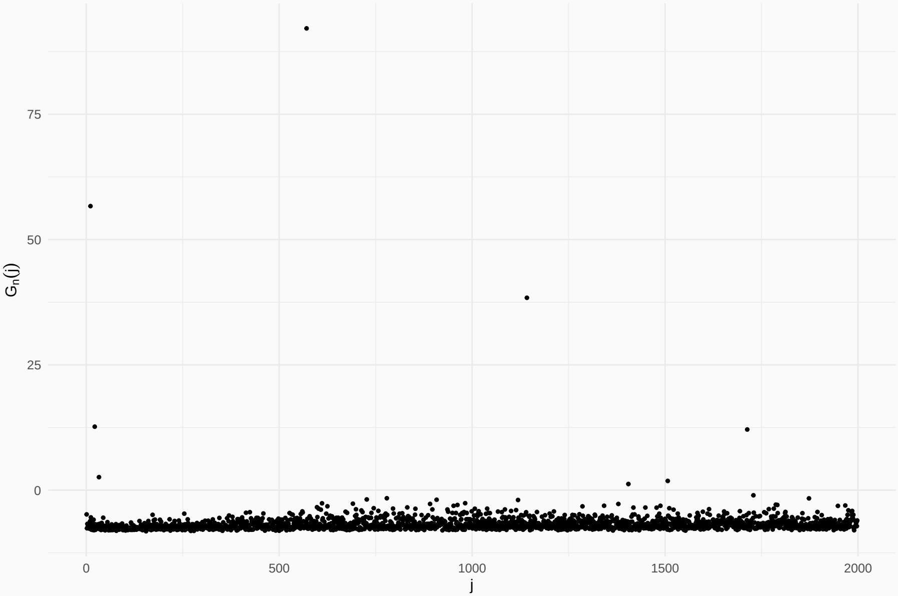
$$G_n(j) := \lambda_1^{-1} \|(I - e^{-i\omega_j}\hat{\rho})(\mathcal{X}_n(\omega_j))\|^2 - \log q + \max\Big\{\sum_{j=2}^{\tau_n} \log(1 - \hat{\lambda}_j/\hat{\lambda}_1), c_n\Big\},$$

  where $n = 3997$.

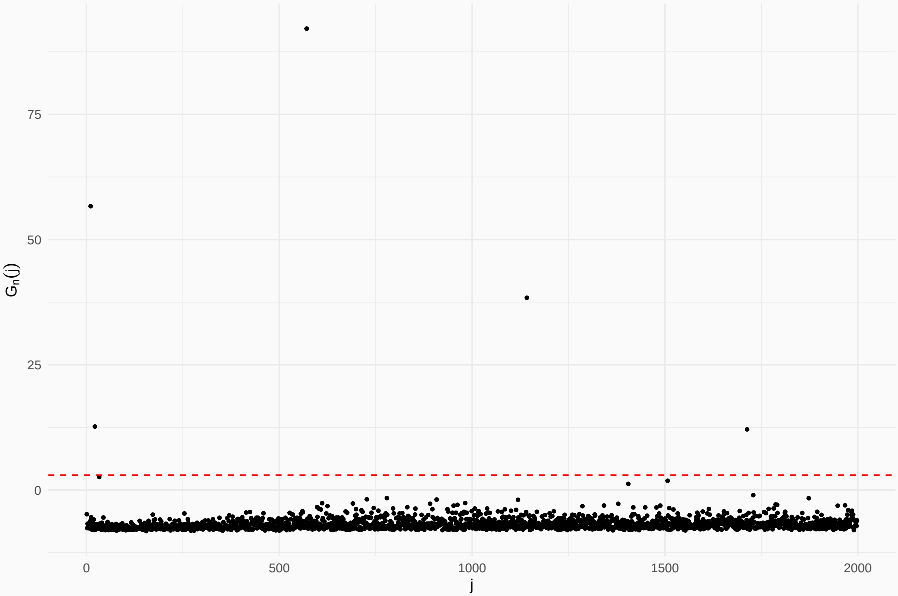
- Observe that

$$G_n = \max_{1 \le j \le q} G_n(j).$$

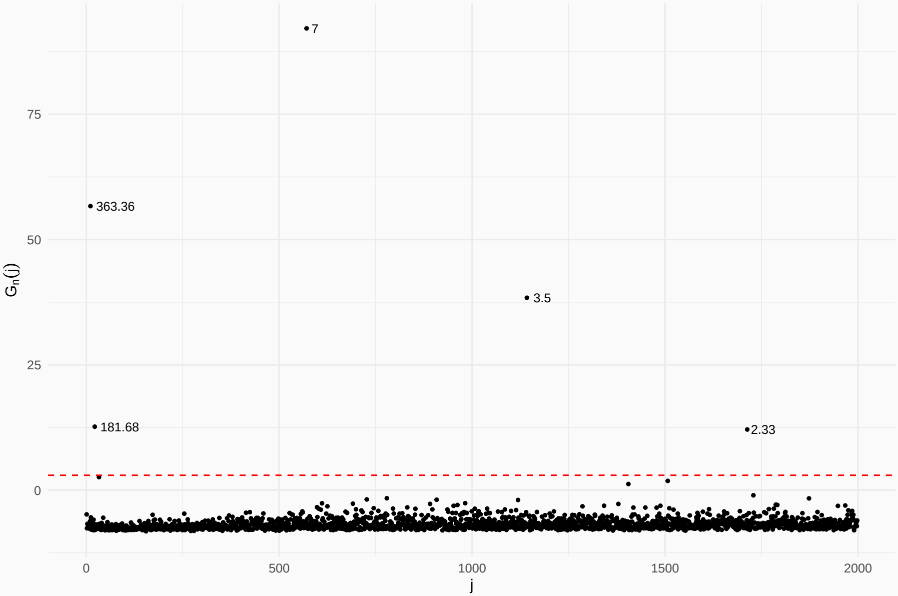
PM10 time series



PM10 time series



PM10 time series



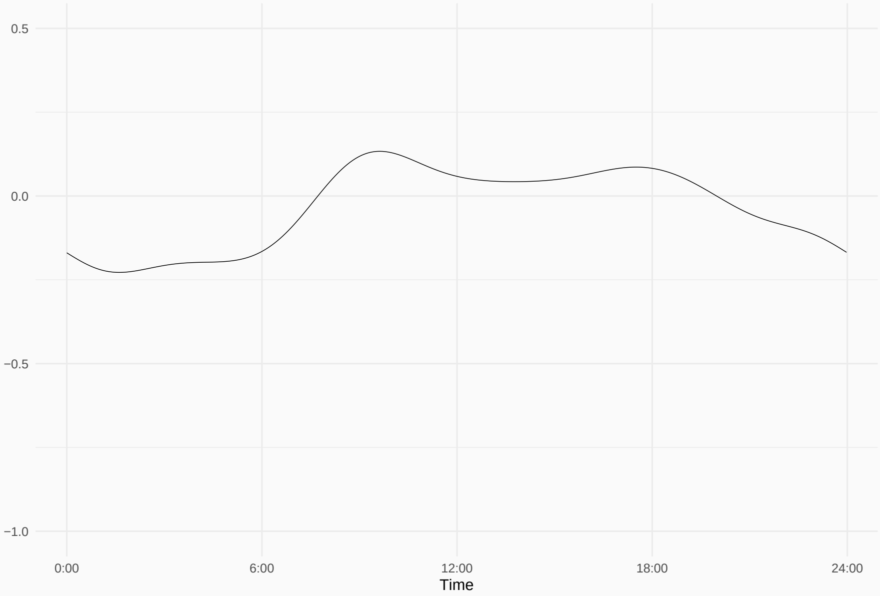
Estimation of periodic signals

The natural estimators of w_{1k} and w_{2k} are given by

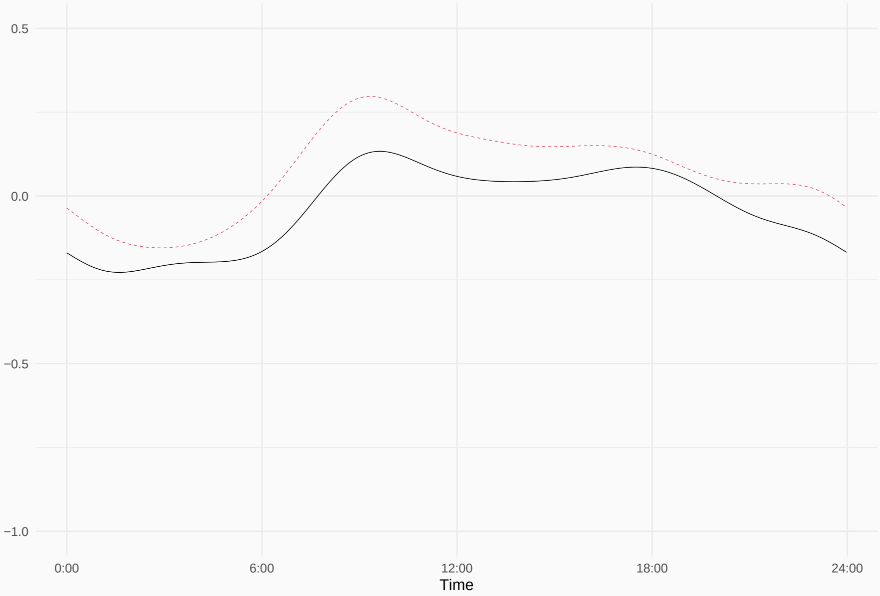
$$\hat{w}_{1k} = \frac{2}{n}\sum_{t=1}^{n} X_t \cos(2\pi kt/T) \quad \text{and} \quad \hat{w}_{2k} = \frac{2}{n}\sum_{t=1}^{n} X_t \sin(2\pi kt/T)$$

with $k = 1, \ldots, \lfloor T/2 \rfloor$.

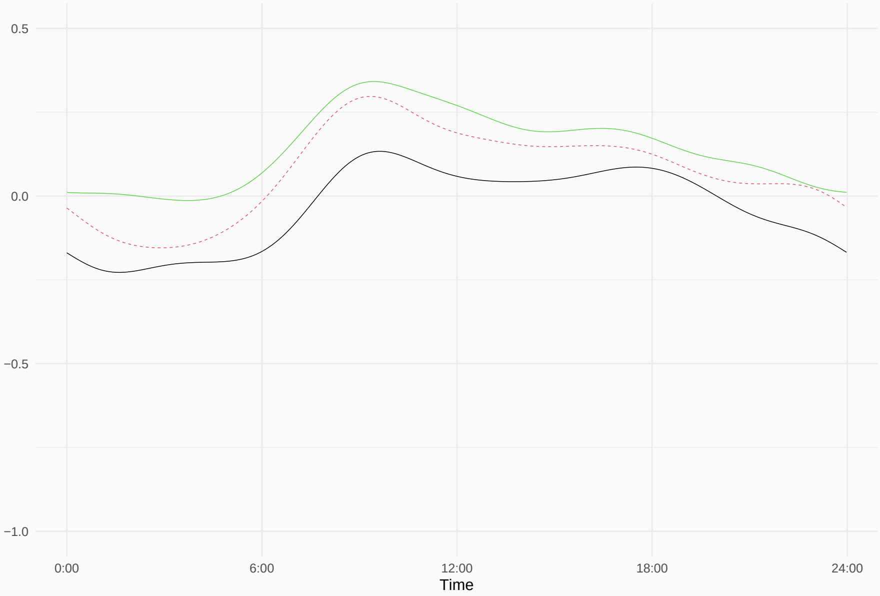
Weekly periodic component



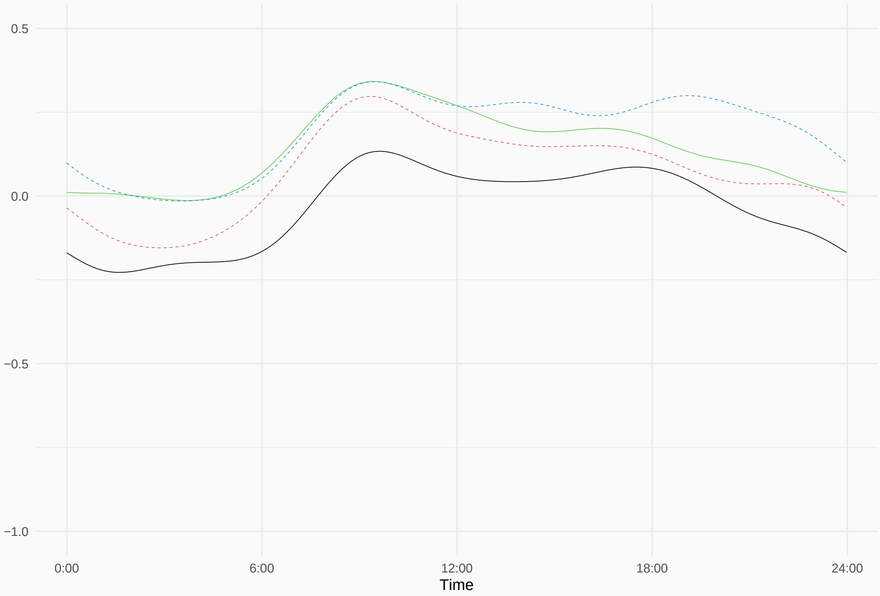
Weekly periodic component



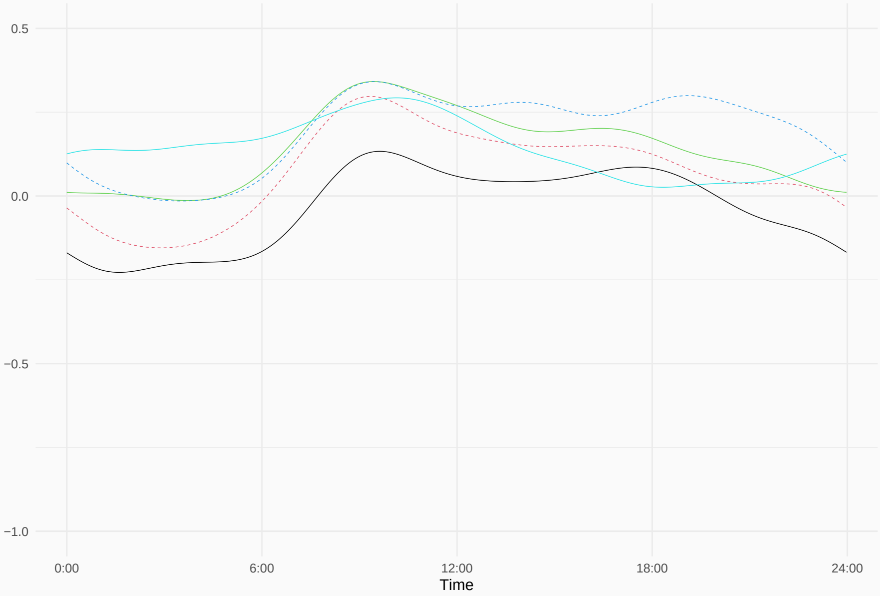
Weekly periodic component



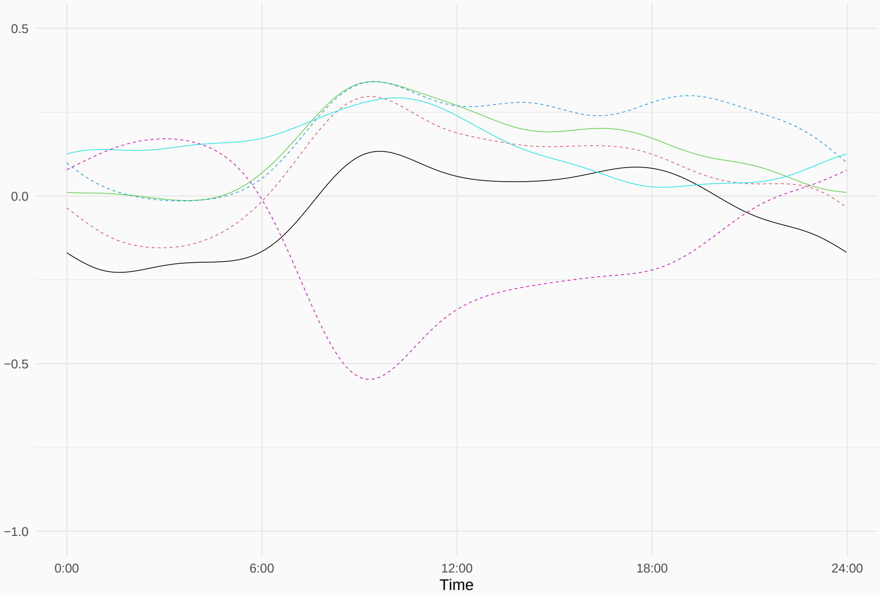
Weekly periodic component



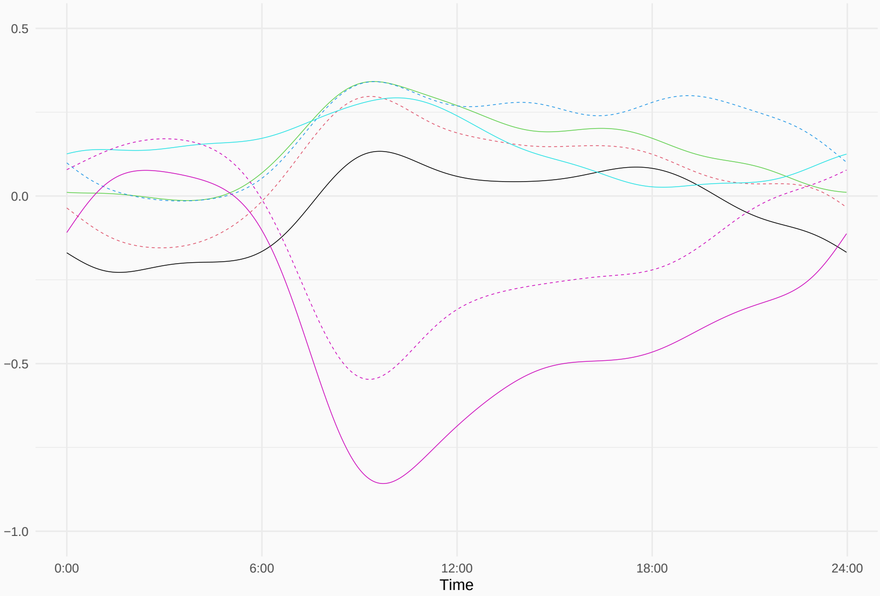
Weekly periodic component



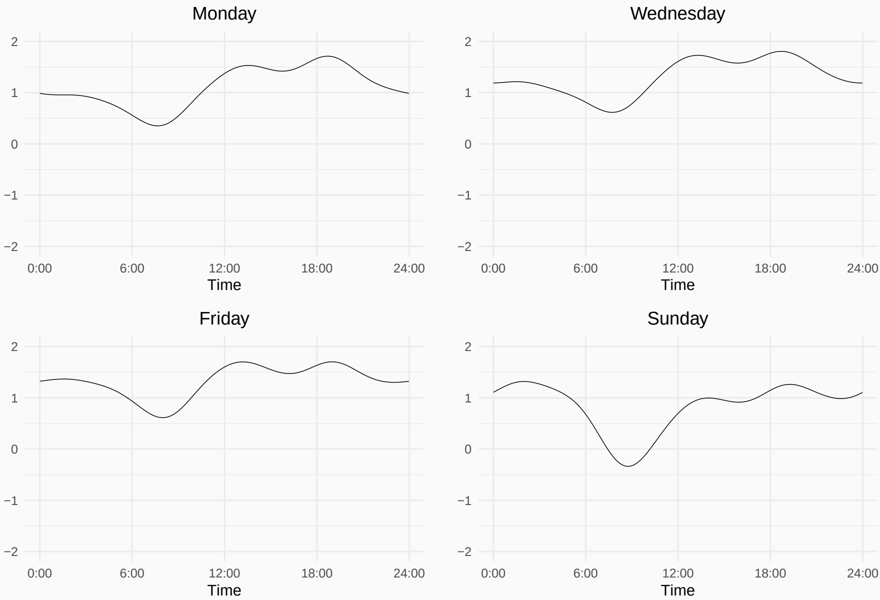
Weekly periodic component



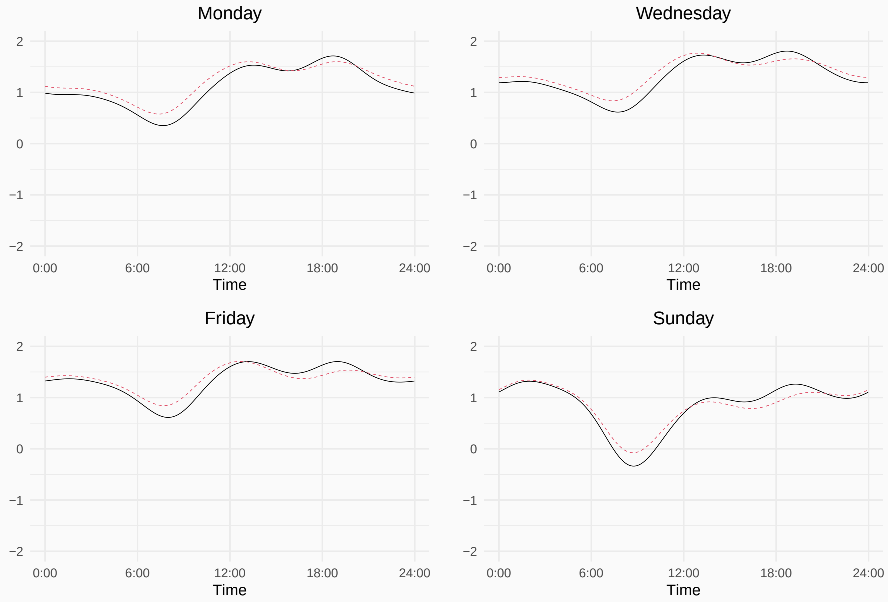
Weekly periodic component



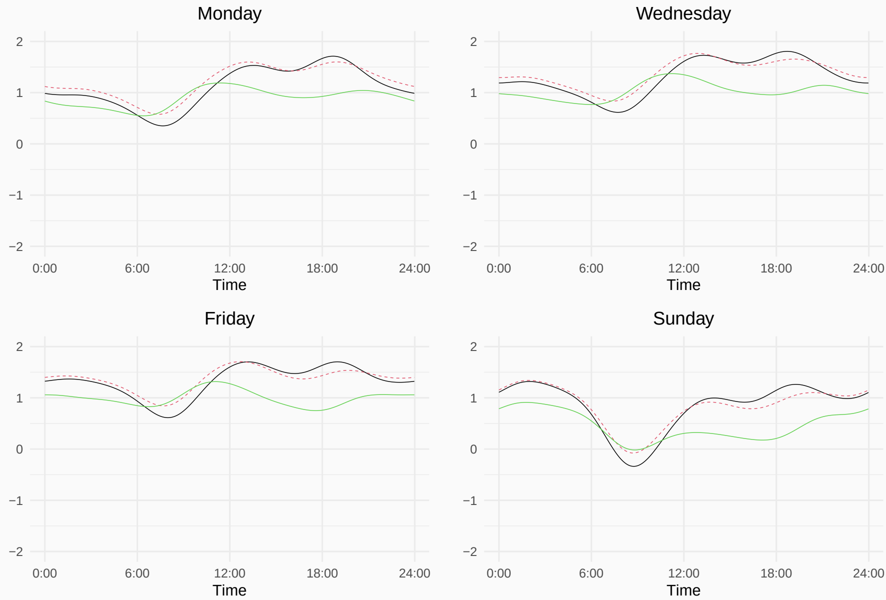
Yearly periodic component



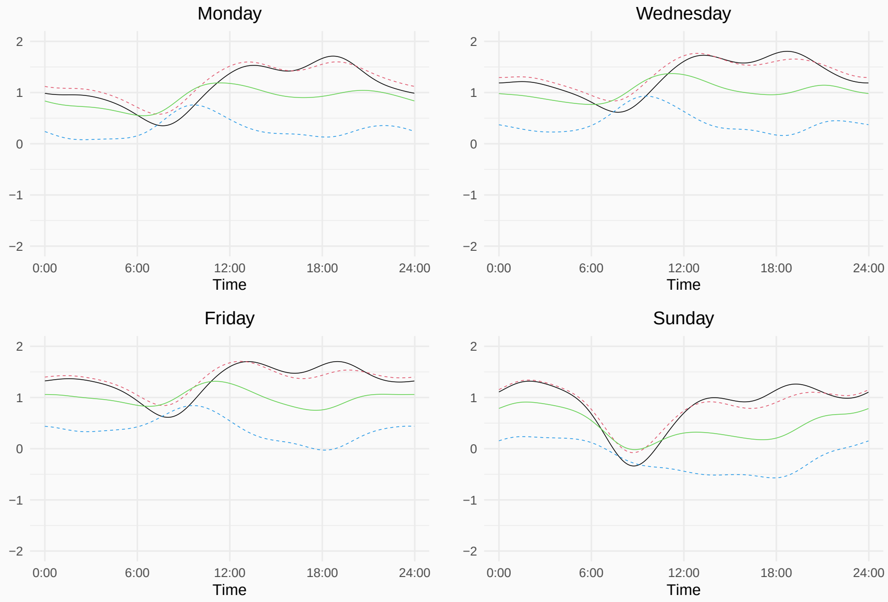
Yearly periodic component



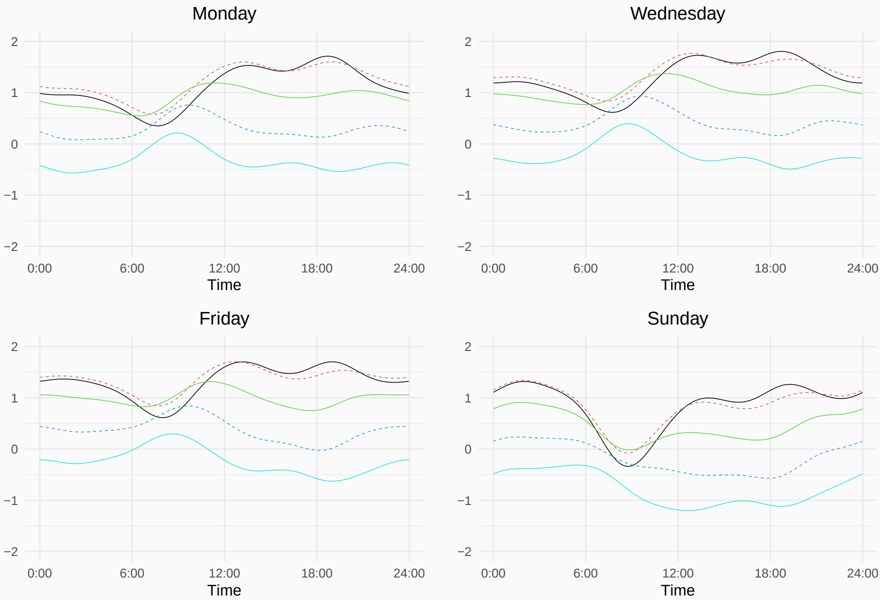
Yearly periodic component



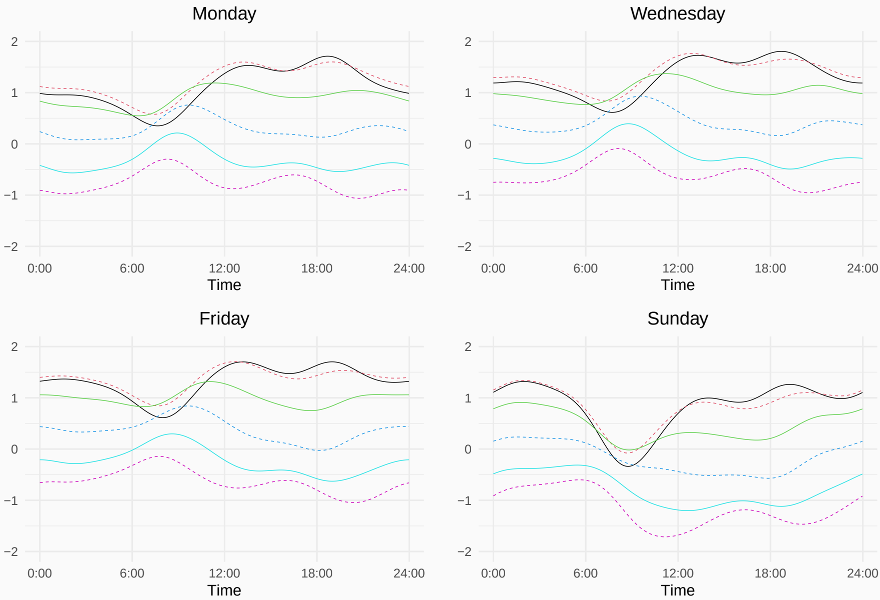
Yearly periodic component



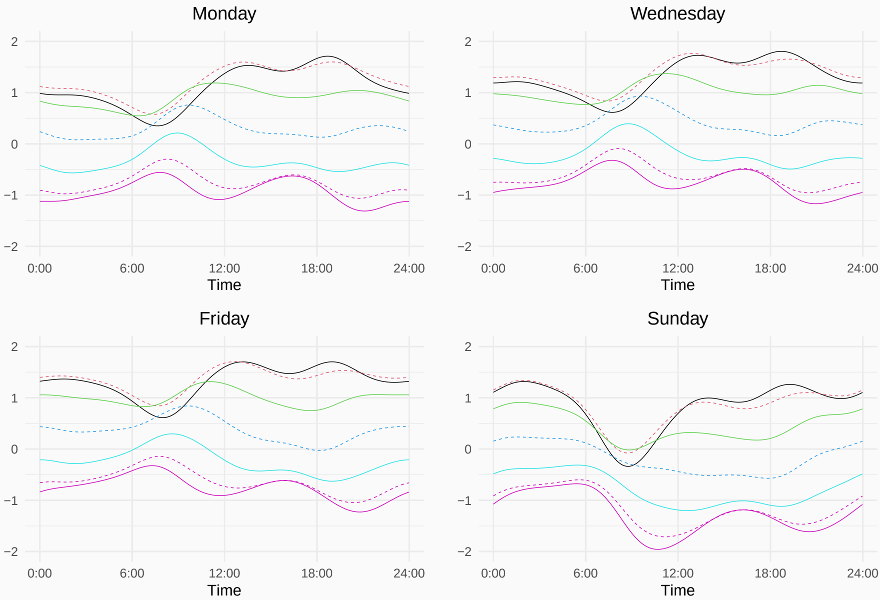
Yearly periodic component



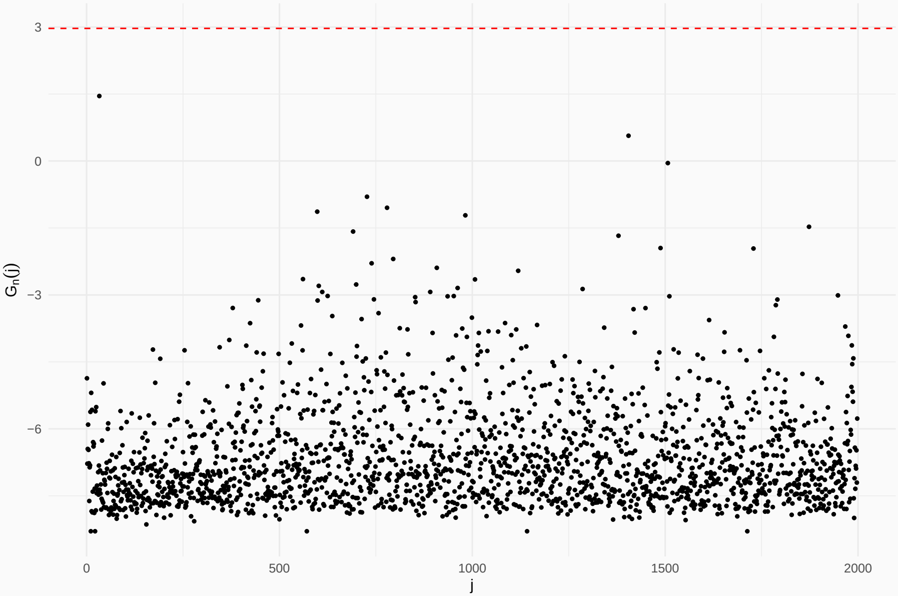
Yearly periodic component



Yearly periodic component



Deseasonalized data



Summary

Summary

- A general test for periodic signals in Hilbert space valued time series when the length of the period is unknown.
- The appropriately standardized maximum of the periodogram converges in distribution to the standard Gumbel distribution.
- A weekly as well as a yearly periodic components are detected in the PM10 data.
- The periodic signals in the PM10 data are not pure sinusoids but are actually driven by several sinusoids.

https://imada.sdu.dk/u/characiejus/

References (1/2)

D. Bosq. *Linear Processes in Function Spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag New York, 2000.

Clément Cerovecki, Vaidotas Characiejus, and Siegfried Hörmann. The maximum of the periodogram of a sequence of functional data. *Journal of the American Statistical Association*, 0(0):1–9, 2022. doi: 10.1080/01621459.2022.2071720. URL `https://doi.org/10.1080/01621459.2022.2071720`.

V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45:2309–2352, 2017.

Richard A. Davis and T. Mikosch. The maximum of the periodogram of a non-Gaussian sequence. *The Annals of Probability*, 27:522–536, 1999.

Uwe Einmahl. Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of Multivariate Analysis*, 28(1):20 – 68, 1989.

Ronald A. Fisher. Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A. Containing Papers of a Mathematical and Physical Character*, 125(796): 54–59, August 1929. doi: https://doi.org/10.1098/rspa.1929.0151.

References (2/2)

A. Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism*, 3(1):13–41, 1898.

A.M. Walker. Some asymptotic results for the periodogram of a stationary time series. *Journal of the Australian Mathematical Society*, 5:107–128, 1965.

G. T. Walker. *Correlation in seasonal variations of weather, III : on the criterion for the reality of relationships or periodicities*, volume 21 of *Memoirs of the India Meteorological Department*. Meteorological Office, 1914.

G. Udny Yule. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226:267–298, 1927.