Censored Functional Time Series: Theory and Application to Fair-Weather Atmospheric Electricity [1]

Tomáš Rubín[1], **Victor M. Panaretos**[2]

Ecole Polytechnique Fédérale de Lausanne, Switzerland

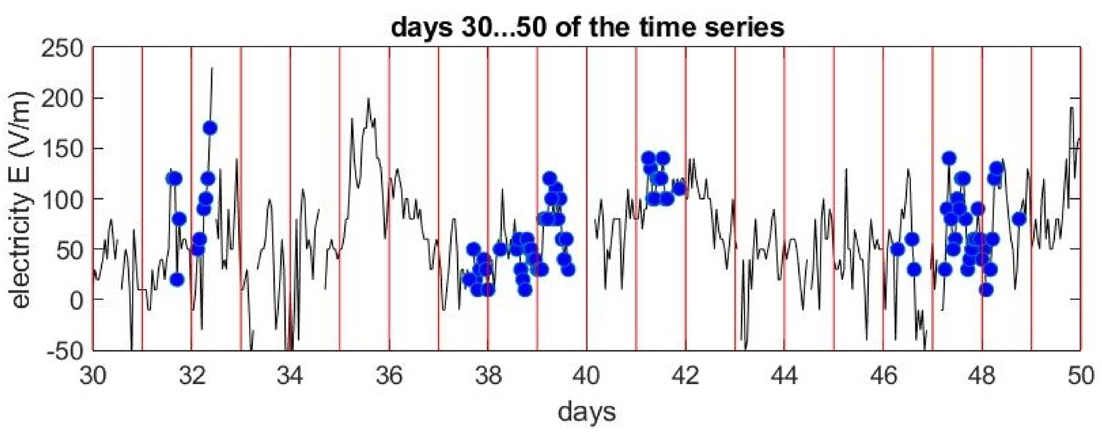
[1] *tomas.rubin@epfl.ch* [2] *victor.panaretos@epfl.ch*

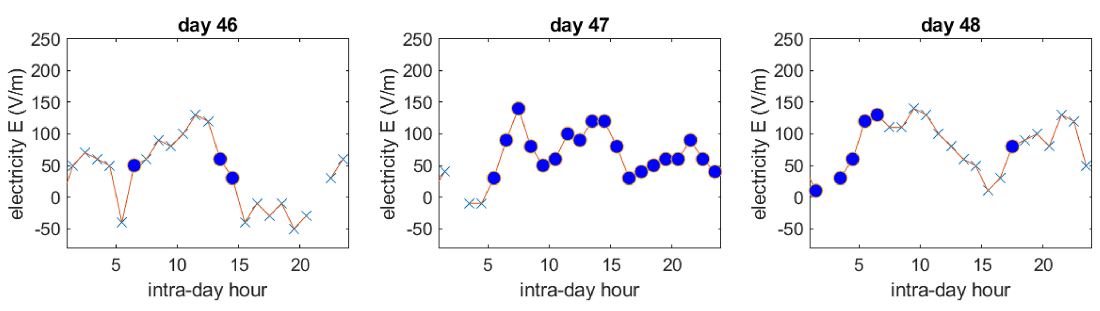
Summary

Functional time series (FTS) is a time-dependent sequence of random functions. The novelty of this contribution is in considering sparse observation protocol (censoring). We propose kernel-regression based estimators of the model dynamics, in particular the estimator of the spectral density. Furthermore, we show how to identify periodicities in data and how to recover (predict) the latent functional data themselves. The proposed methodology is demonstrated on fair-weather atmospheric electricity data.

1. Motivation: Atmospheric Electricity Data

- Atmospheric electric field Tashkent, Uzbekistan, measurements between years 1988 –- 1993 (source: [2]).
- Scalar time-series → cut into intraday profiles → separate the intraday variability and temporal dependence among consecutive days.
- Underlying continuous and smooth process = *functional* nature of data.

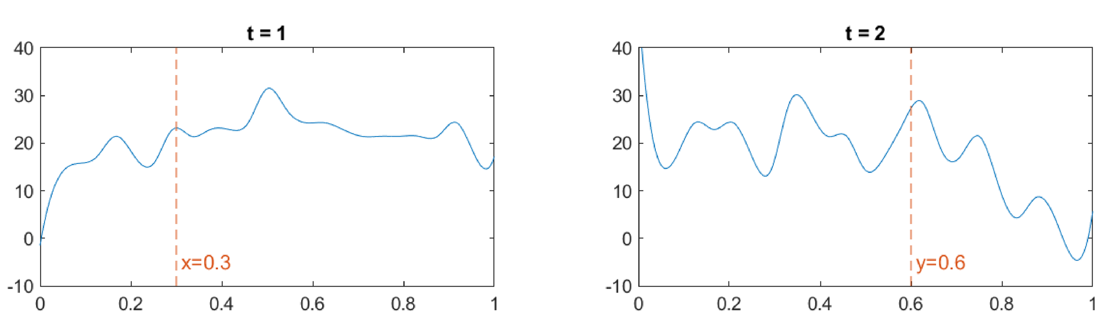




- **Problem:** the measurements are reliable only in fair-weather conditions ($\leq 20\,km/h$ wind speed, cloudless sky) — blue circles in the above plots. The unfair-weather (crosses) points are discarded.
- Thus the time-series is in fact censored, i.e. sparsely observed.
- **Questions & aims**:
  - separate the intra-day and inter-day dependence,
  - interpret the dynamics of the time-series,
  - fill-in the gaps in the data, remove noise, construct confidence bands.
- **The dataset in numbers**:
  - 5 years = 1826 days,
  - 12997 fair-weather measurements (29 %),
  - 1118 days with at least 1 fair-weather measurement (61 %),
  - 7.1 fair-weather measurements per day on average,
  - 11.6 fair-weather measurements per day on average among the days where there is at least one measurement.

2. Functional Time Series Framework

- **Functional time series (FTS)**: ordered and time dependent collection of random smooth curves $X_t(\cdot), t = 1, \dots, T$. $x \mapsto X_t(x), x \in [0, 1]$, is a function/curve.



- Assuming stationarity in the time variable t, we may define.
  - Mean function $\mu(x) = \mathbb{E}[X_t(x)]$.
  - Lag-0 covariance kernel $R_0(x, y) = \text{cov}(X_t(x), X_t(y))$.
  - Lag-h autocovariance kernel $R_h(x, y) = \text{cov}(X_{t+h}(x), X_t(y))$.
- The **spectral analysis**: a classical result in scalar time-series: the autocovariance function is 1-1 to its Fourier transform, the spectral density; also for functional time series, [3].
  - Spectral density kernel $f_\omega(x, y) = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} R_h(x, y)e^{-\mathrm{i}\omega h}$ at frequency $\omega \in (-\pi, \pi)$
  - Inverse formula $R_h(x, y) = \int_{-\pi}^{\pi} f_\omega(x, y)e^{\mathrm{i}\omega h}$
- The novelty of this contribution is in considering **sparse observation scheme** (censoring protocol) possibly with measurement errors. The dataset is

$$Y_{tj} = X_t(x_{tj}) + \epsilon_{tj}, \quad j = 1, \dots, N_t, \quad t = 1, \dots T$$

where x_{tj} are independent measurement locations and ϵ_{tj} are iid homoscedastic errors with variance $\sigma^2 > 0$.

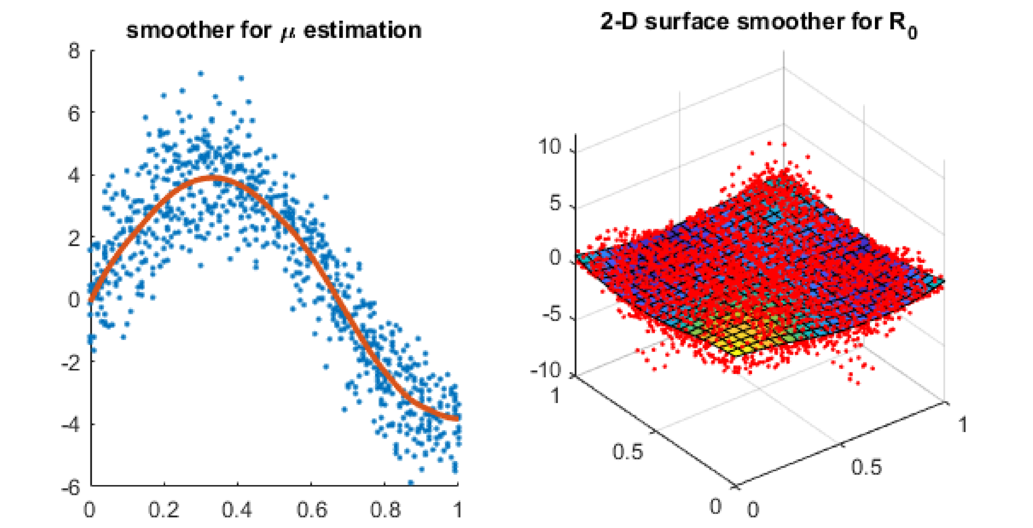
3. Estimation of the Model Dynamics

3.1 Estimation of $\mu(\cdot)$ and $R_h(\cdot, \cdot)$

- We estimate the model dynamics components by kernel smoothing methods à la Yao et al. [4].
- Estimator of $\mu(\cdot)$ — local linear smoother over (x_{tj}, Y_{tj}).
- Estimators of second order dynamics is based on "raw covariances"

$$G_{h,t}(x_{t+h,j}, x_{tk}) \stackrel{def}{=} \left(x_{t+h,j} - \hat\mu(x_{t+h,j})\right)\left(x_{tk} - \hat\mu(x_{tk})\right)$$
$$\mathbb{E}\left[G_{h,t}(x_{t+h,j}, x_{tk})\right] \approx R_h\left(x_{t+h,j}, x_{tk}\right) + \sigma^2 1_{[h=0,j=k]}$$

- Estimator of $R_0(\cdot, \cdot)$ — 2-D local linear smoother over $(x_{tj}, x_{tk}, G_{0,t}(x_{tj}, x_{tk}))$ for $j \neq k$.



- Estimator of $R_h(\cdot, \cdot)$ — 2-D local linear smoother over $(x_{t+h,j}, x_{tk}, G_{h,t}(x_{t+h,j}, x_{tk}))$.
- Estimator of σ^2 — local quadratic on the diagonal.

3.2 Estimation of Spectral Density

- Recall the Bartlett's estimate for scalar time series

$$\hat f_\omega = \frac{1}{2\pi} \sum_{h=-Q}^{Q} W_h \hat c_h e^{-\mathrm{i}\omega h}$$

where $\hat c_h$ is the empiric autocovariance, the Bartlett's weights are $W_h = \frac{1-|h|}{Q}$ for the lag $|h| < Q$ and Q is the Bartlett's span parameter.
- Estimator of $f_\omega(\cdot, \cdot)$ for given $\omega \in (-\pi, \pi)$ — complex-valued 2-D local linear smoother over

$$\left(x_{t+h,j}, x_{tk}, G_{h,t}(x_{t+h,j}, x_{tk})e^{-\mathrm{i}\omega h}\right)$$

each with the Bartlett's weights $W_h = \frac{1-|h|}{Q}$ for $|h| < Q$.
- The parameter Q must be:
  - moderate ($Q \sim T^{1/3}$) for estimation of $\{f_\omega(\cdot, \cdot)\}_{\omega\in(-\pi,\pi)}$,
  - very high ($Q \approx T/2$) when periodicity discovery is of interest (undersmoothing).

3.3 Theoretical Results & Practice

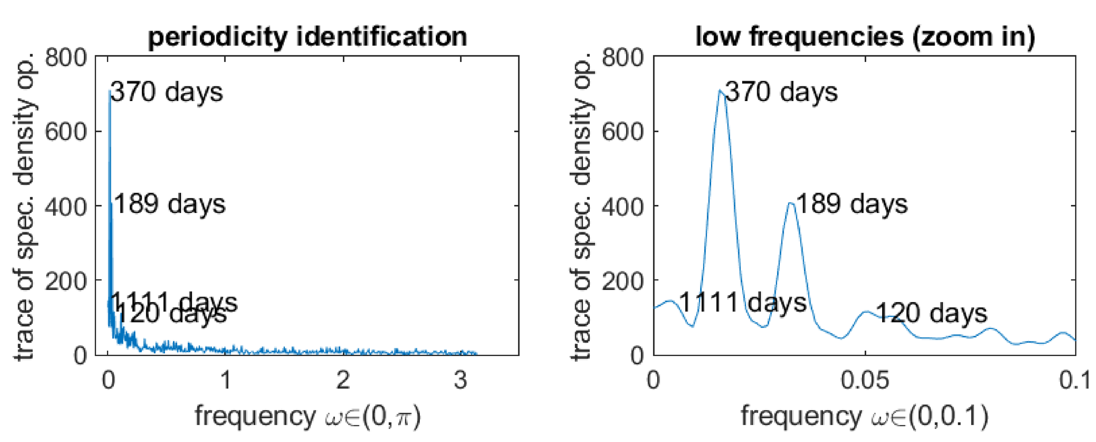
- We proved the asymptotic behaviour (convergence rates) of the all above estimators under some smoothness assumptions and cumulant mixing conditions.
- We checked the sub-asymptotic performance on simulations.
- We suggested cross-validation driven selection procedure for tuning parameters (smoothing bandwidths).

4. Functional Data Recovery

- Once the spectral density is estimated, we obtain the estimate of the space-time covariance by the inverse Fourier formula.
- Conditioned on the observed data, the functional data as the latent process can be recovered by the best linear unbiased predictor (BLUP).
- Assuming Gaussianity, we can construct pointwise and simultaneous confidence bands for each functional datum.

5. Results for Atmospheric Electricity Data

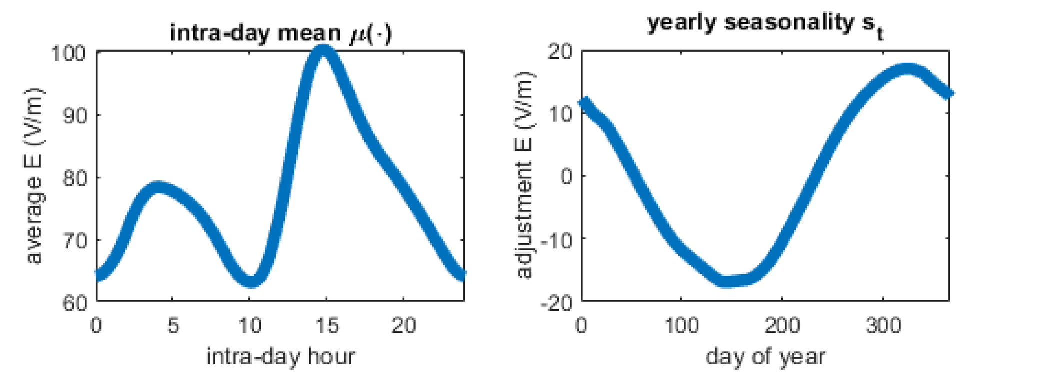
- Undersmoothed estimator of spectral density to investigate the periodicities:

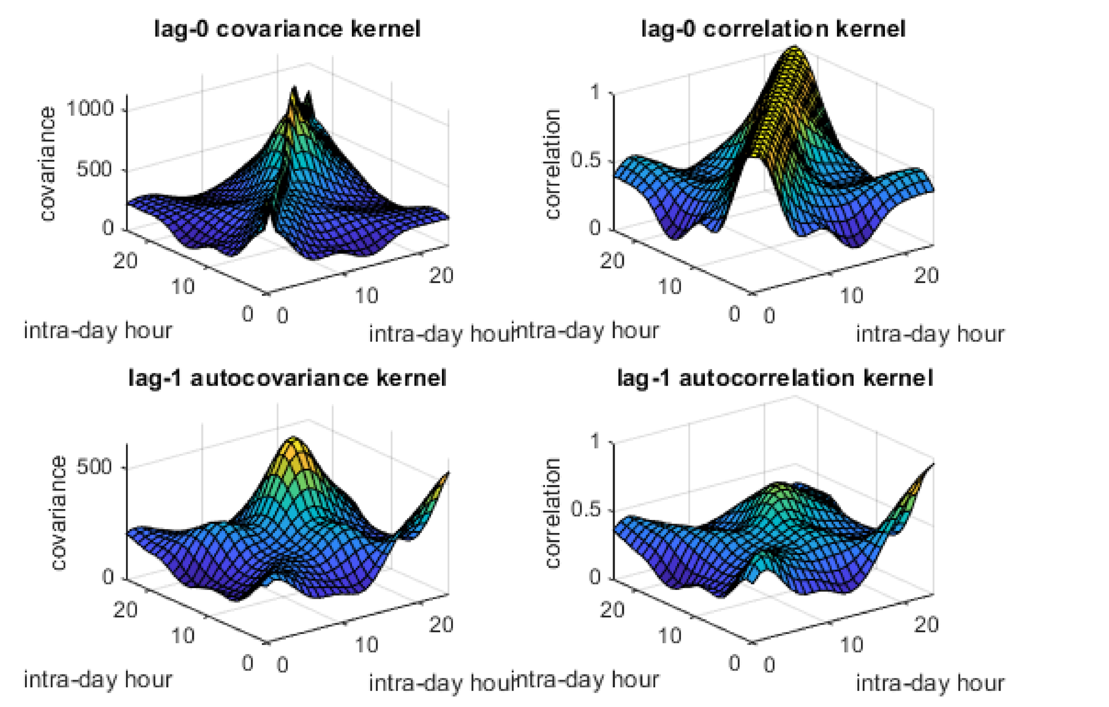


- There is an obvious yearly periodicity which we opt to model deterministically. We propose the additive model:

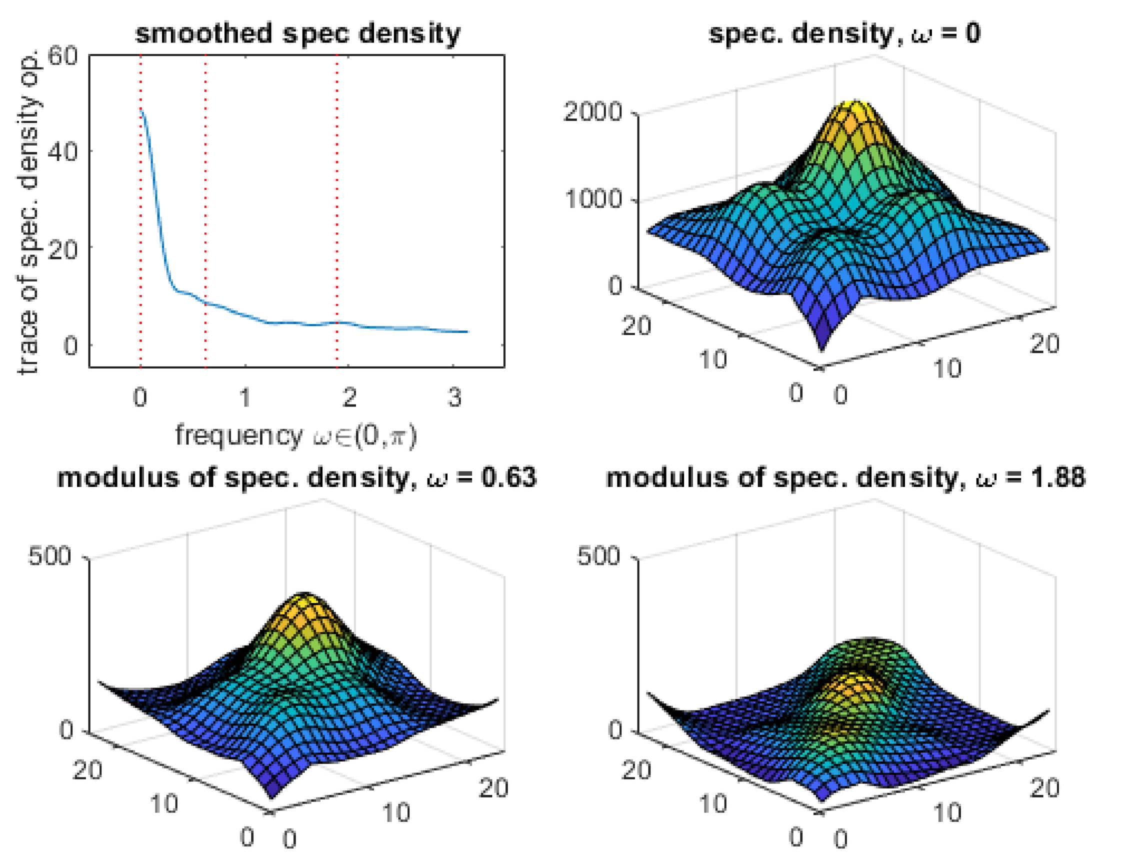
$$Y_{tj} = \mu(x_{tj}) + s_t + X_t(x_{tj}) + \epsilon_{tj}$$

where $\mu(\cdot)$ is the intra-day mean, s_t is yearly seasonality adjustment, and $X_t(x)$ is now a zero-mean stationary FTS.
- The model dynamics components $\mu(\cdot)$, s_t, $\{f_\omega(\cdot, \cdot)\}_{\omega\in(-\pi,\pi)}$, as well as lag-$0$ and lag-1 (auto)covariance kernels for illustration, were estimated by our kernel-regression estimates.

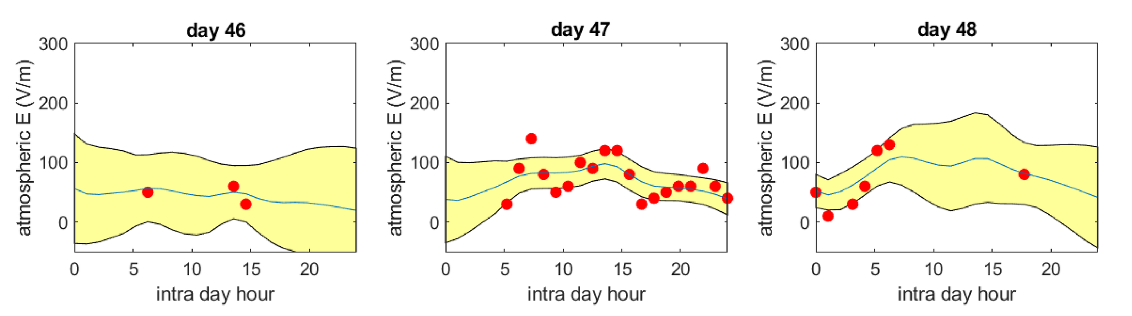




- Spectral density is estimated by the proposed Barlett-like smoothing estimator with the Barlett span parameter $Q = 20$.



- Once the spectral density is estimated, we recover the functional data and construct the simultaneous 95%-confidence bands.



References

[1] Tomáš Rubín and Victor M Panaretos. Sparsely observed functional time series: Estimation and prediction. *arXiv preprint arXiv:1811.06340*, 2018.

[2] Hannes Tammet. A joint dataset of fair-weather atmospheric electricity. *Atmospheric Research*, 91(2-4):194–200, 2009.

[3] Victor M. Panaretos and Shahin Tavakoli. Fourier analysis of stationary time series in function space. *The Annals of Statistics*, 41(2):568–603, 2013.

[4] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.