Adaptive Prediction for Functional Time Series

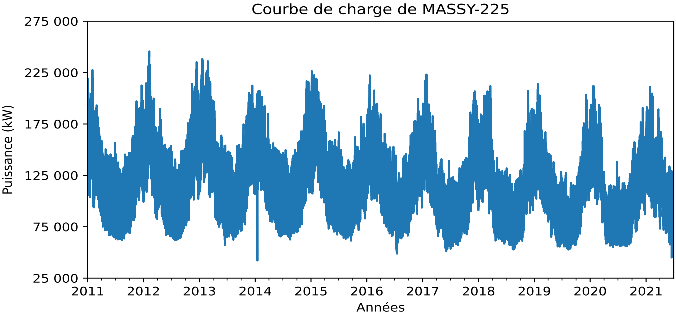
Hassan Maissoro, Valentin Patilea and Myriam Vimond

ISNPS, June 25-29 2024

Introduction (1/3)

Example of a connection point for the extraction and injection of electricity

- A set of N time-dependent curves, $X_n : [0, 1] \to \mathbb{R}$, $n = 1 \dots N$.



- The trajectories are irregular.
- We observe each curve with measurement errors.
- Regularity and final goal should be considered in reconstruction.

Introduction (2/3)

Observation scheme

For $n = 1, \ldots, N$, X_n is measured with error at discrete, randomly sampled points :

$$Y_{n,k} = X_n(T_{n,k}) + \sigma(T_{n,k})\varepsilon_{n,k}, \quad 1 \leq k \leq M_n,$$

- $\{X_n\}$ is a stationary process of $\mathcal{H} = \mathbb{L}^2[0,1]$,
- $M_1, \ldots, M_N \overset{i.i.d.}{\sim} M$ with expectation λ,
- the observation times $T_{n,k} \sim T$ are i.i.d.,
- $\varepsilon_{n,k} \sim \epsilon$ are independent centred errors with unit variance,
- $\{X_n\}$, $\{M_n\}$, $\{\varepsilon_{n,k}\}$, and $\{T_{n,k}\}$ are mutually independent.

Introduction (3/3)

Motivation

We aim to build a procedure for curve prediction that adapts to the local regularity of the trajectories for FTS in the context of weak dependence.

Using dependent curves measured with noise at random discrete points, our goal is to perform adaptive estimation of :

- the Best Linear Unbiased Predictor (BLUP) that is a combination of
- mean, covariance and autocovariance functions.

- For FTS, a functional data recovery has already been considered by Rubìn and Panaretos (2020) under the hypothesis that these functions admit at least one derivative.
- For irregular curves, Maissoro et al. (2024) proposed new estimators of the mean and autocovariance functions.

Outline

Adaptive linear predictor (1/6)

Let $\mu(t) = \mathbb{E}(X_n(t))$ and $\Gamma_\ell(s,t) = \mathbb{E}\left\{[X_0(s)-\mu(s)][X_\ell(t)-\mu(t)]\right\}$, for all $s,t \in I$ and $\ell \geq 0$. Moreover,

$$\mathbb{Y}_n = \left(Y_{n,1},\dots,Y_{n,M_n}\right)^\top, \quad \mathcal{Y}_{n_0,1} = \left(\mathbb{Y}_{n_0-1}^\top,\mathbb{Y}_{n_0}^\top\right)^\top, \quad \Sigma_n = \operatorname{diag}\left(\sigma^2(T_{n,1}),\dots,\sigma^2(T_{n,M_n})\right),$$

$$\mathcal{M}_{n_0,1} = \left(\mu(T_{n_0-1,1}),\dots,\mu(T_{n_0-1,M_{n_0-1}}),\mu(T_{n_0,1}),\dots,\mu(T_{n_0,M_{n_0}})\right)^\top.$$

Definition. Let $t_0 \in I$ and $n_0 \in \{1,\dots,N\}$ be fixed. Following Robinson (1991), the BLUP of $X_{n_0}(t_0)$ given $\mathcal{Y}_{n_0,1}$ is :

$$\widehat{X}_{n_0}(t_0) = \widehat{\mu}(t_0) + \widehat{B}_{n_0,1}^\top(\mathcal{Y}_{n_0,1} - \widehat{\mathcal{M}}_{n_0,1}),$$

$$\text{where} \quad B_{n_0,1} = \begin{pmatrix} G_0^{(n_0-1,n_0-1)}+\Sigma_{n_0-1} & G_1^{(n_0-1,n_0)} \\ G_1^{(n_0,n_0-1)} & G_0^{(n_0,n_0)}+\Sigma_{n_0} \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_1(T_{n_0-1,1},t_0) \\ \vdots \\ \Gamma_1(T_{n_0-1,M_{n_0-1}},t_0) \\ \Gamma_0(T_{n_0,1},t_0) \\ \vdots \\ \Gamma_0(T_{n_0,M_{n_0}},t_0) \end{pmatrix},$$

and $G_\ell^{(n,n')} = \left(\Gamma_\ell(T_{n,i},T_{n',j})\right)_{1\leq i\leq M_n, 1\leq j\leq M_{n'}}$.

Estimation. Put a hat on to get an estimate...

Adaptive linear predictor (2/6)

Local Regularity Parameters

Definition. The process X admits a *local regularity* at $t \in I$, with local exponent $H_t \in (0,1)$ and Hölder constant $L_t > 0$, if

$$\mathbb{E}\left[(X(u) - X(v))^2\right] \approx L_t^2 |u - v|^{2H_t},$$

for all u, v satisfying $t - \Delta/2 \leq u \leq t \leq v \leq t + \Delta/2$ for some $\Delta > 0$.

Adaptive linear predictor (2/6)

Local Regularity Parameters

Definition. The process X admits a *local regularity* at $t \in I$, with local exponent $H_t \in (0,1)$ and Hölder constant $L_t > 0$, if

$$\mathbb{E}\left[(X(u) - X(v))^2\right] \approx L_t^2 |u - v|^{2H_t},$$

for all u, v satisfying $t - \Delta/2 \leq u \leq t \leq v \leq t + \Delta/2$ for some $\Delta > 0$.

Estimation. We use some nonparametric estimates $\widetilde{X}_n$ to recover the X_n's. For any u, v close to t, let

$$\widehat{\theta}(u, v) = \frac{1}{N} \sum_{n=1}^{N} \left\{ \widetilde{X}_n(v) - \widetilde{X}_n(u) \right\}^2 .$$

Our estimators of H_t and L_t^2 are defined as empirical counterparts of their respective definition. Let $t_1 = t - \Delta/2$, $t_3 = t + \Delta/2$. The estimators of H_t and L_t^2 are

$$\widehat{H}_t = \frac{\log(\widehat{\theta}(t_1, t_3)) - \log(\widehat{\theta}(t_1, t))}{2\log(2)} \qquad \text{and} \qquad \widehat{L_t^2} = \frac{\widehat{\theta}(t_1, t_3)}{\Delta^{2\widehat{H}_t}}.$$

Concentration bounds. Under $\mathbb{L}^{\mathbf{p}}_{\mathcal{C}} -$ **m-approximability** by Maissoro et al. (2024).

Adaptive linear predictor (3/6)

Weak dependency assumption

Let $\{X_n\}_{n\in\mathbb{Z}}$ be a stationary FTS, with continuous paths, on $I = [0,1]$:

- $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$: space of square integrable functions ;
- $(\mathcal{C}, \|\cdot\|_\infty)$: space of continuous functions on I.

The space $\mathbb{L}^p_{\mathcal{C}}$ is the space of $\mathcal{C}-$valued random element X such that

$$\nu_p(\|X\|^p_\infty) = \left(\mathbb{E}\left[\|X\|^p_\infty\right]\right)^{1/p} < \infty.$$

Weak dependency assumption : $\{X_n\}_n$ is $\mathbb{L}^{\mathbf{p}}_{\mathcal{C}}$ $-$ **m-approximable**.

> $\mathbb{L}^{\mathbf{p}} - \mathbf{m}$-approximation for $\mathcal{H}-$valued functional data was introduced by Hörmann and Kokoszka (2010).

Example. *FAR(1) is* $\mathbb{L}^p_{\mathcal{C}}$ $-$ *m-approximable.*

Adaptive linear predictor (4/6)

Adaptive mean estimation. Let $\mu(t) = \mathbb{E}(X_n(t))$.

- A naive estimator of $\mu(t)$ is an average of nonparametric estimator of the curves.
- **The objective :** estimation of $\mu(t)$ by adaptive smoothing of X_n.
- The proposed estimator is $\widehat{\mu}_N(t; h^*_\mu)$, with

$$\widehat{\mu}_N(t; h) = \sum_{n=1}^{N} \frac{\pi_n(t; h)}{P_N(t; h)} \widehat{X}_n(t; h), \qquad \text{where}$$

  - $\pi_n(t; h) = 1$ if there is at least one $T_{n,i} \in [t - h, t + h]$ and 0 otherwise,
  - $P_N(t; h) = \sum_{n=1}^{N} \pi_n(t; h)$,
  - and $\widehat{X}_n(t; h)$ is Nadaraya-Watson estimator with bandwidth h.
- h^*_μ minimises a sharp bound of the quadratic risk of $\widehat{\mu}_N(t; h)$.

Adaptive linear predictor (5/6)

Adaptive mean estimation. More precisely, we consider

$$\mathbb{E}_{M,T}\left[\left(\widehat{\mu}_N(t;h)-\mu(t)\right)^2\right] \leq 2R_\mu(t;h), \quad \text{where}$$

$$R_\mu(t;h) = L_t^2 h^{2H_t}\mathbb{B}(t;h,2H_t) + \sigma^2(t)\mathbb{V}_\mu(t;h) + \mathbb{D}_\mu(t;h)/P_N(t;h),$$

and define $\quad h_\mu^* \in \underset{h\in\mathcal{H}_N}{\arg\min}\, \widehat{R}_\mu(t;h) \qquad$ with $\quad \widehat{R}_\mu(t;h) = R_\mu(t;h,\widehat{H}_t,\widehat{L}_t^2,\widehat{\sigma}^2(t)).$

Let $t \in I$. Under some assumptions we have

$$\widehat{R}_\mu(t;h) = \mathcal{O}_\mathbb{P}\left\{h^{2H_t} + (N\lambda h)^{-1} + N^{-1}\right\},$$
$$h_\mu^* = \mathcal{O}_\mathbb{P}\left\{(N\lambda)^{-\frac{1}{1+2H_t}}\right\},$$

and the estimator $\widehat{\mu}_N(t;h_\mu^*)$ satisfies

$$\widehat{\mu}_N^*(t) - \mu(t) = \mathcal{O}_\mathbb{P}\left\{(N\lambda)^{-\frac{H_t}{1+2H_t}} + N^{-1/2}\right\}.$$

Adaptive linear predictor (6/6)

Adaptive lag$-\ell(\ell > 0)$ autocovariance estimation. As for the mean,

- the *'first smooth, then estimate'* estimator is $\widehat{\Gamma}_{N,\ell}(s,t;h_s^*,h_t^*)$, where
- (h_s^*,h_t^*) minimises a sharp bound of the quadratic risk of $\widehat{\Gamma}_{N,\ell}(s,t;h_s,h_t)$.

Let $s,t \in I$. Under some assumptions we have

$$h_s^* = \mathcal{O}_{\mathbb{P}}\left(\max\left\{(N\lambda^2)^{-\frac{H_t}{H_s\{2H_t+1\}+H_t}}, (N\lambda)^{-\frac{1}{2H_s+1}}\right\}\right),$$
$$h_t^* = \mathcal{O}_{\mathbb{P}}\left(\max\left\{(N\lambda^2)^{-\frac{H_s}{H_s\{2H_t+1\}+H_t}}, (N\lambda)^{-\frac{1}{2H_t+1}}\right\}\right),$$

and the estimator $\Gamma_{N,\ell}^*(s,t) = \widehat{\Gamma}_{N,\ell}(s,t;h_s^*,h_t^*)$ satisfies

$$\widehat{\Gamma}_{N,\ell}^*(s,t) - \Gamma_\ell(s,t) = \mathcal{O}_{\mathbb{P}}\left((N\lambda^2)^{-\frac{H_sH_t}{H_s\{2H_t+1\}+H_t}} + (N\lambda)^{-\frac{H_s}{2H_s+1}} + (N\lambda)^{-\frac{H_t}{2H_t+1}} + N^{-1/2}\right).$$

Better rates than those using a single bandwidth, as in Golovkine et al. (2021).

Application (1/5)

We simulate a FAR(1) where the WN are i.i.d. *multifractional Brownian motion* (see Stoev and Taqqu (2006)) paths with :

- a logistic H_t function and $L_t^2 = 1$,
- a kernel $\Psi(s, t)$ estimated from voltage curves (see Hebrail and Berard (2012)).

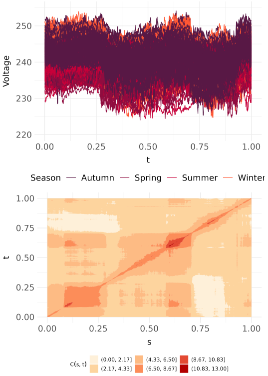


Figure – Cov

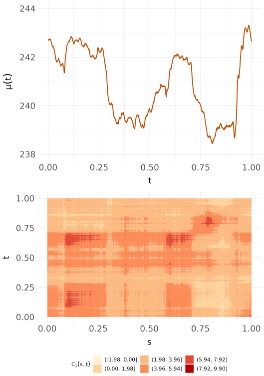


Figure – lag-1 Autocov

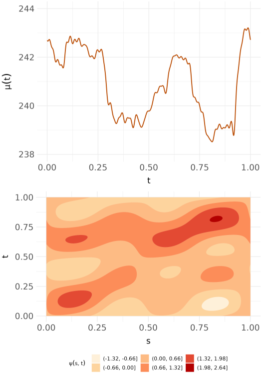
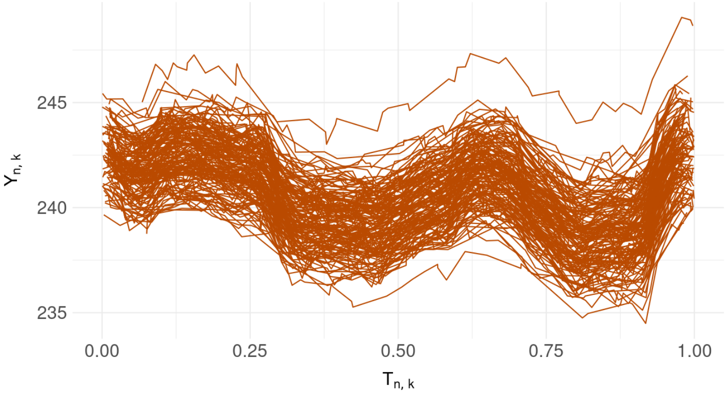


Figure – Kernel $\Psi(s, t)$

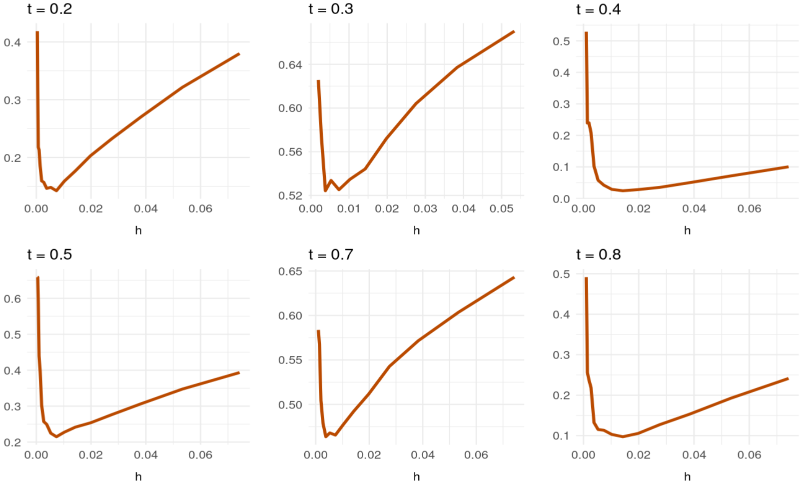
Application (2/5)

Generate curves for $N = 150$ and $\lambda = 40$ with additional normal noise of standard deviation 0.25.



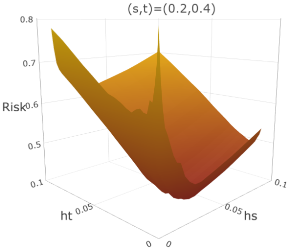
Applications (3/5)

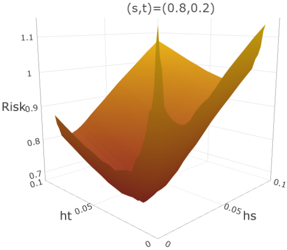
Adaptive mean function estimation. $\widehat{R}_\mu(t; h)$ at some locations :

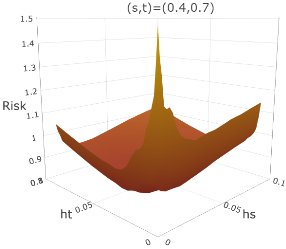


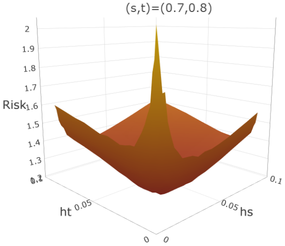
Application (4/5)

Adaptive mean function estimation. $\widehat{R}_\Gamma(s, t; h_s, h_t)$ at some locations :







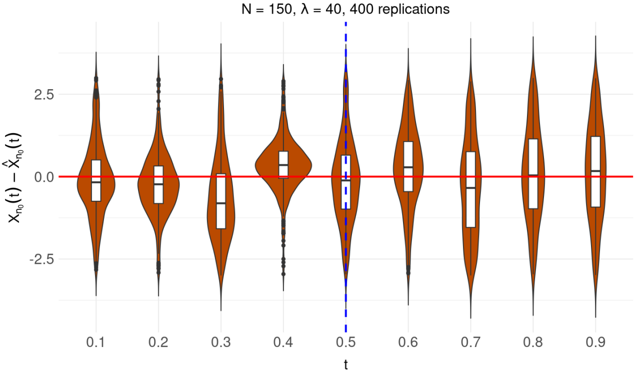


Applications (5/5)

Application : Adaptive BLUP estimation.

Generate a FTS with $N = 150$ and $\lambda = 40$ and regenerate the 150-th curve 400 times

- Estimate the BLUP using only points observed before 0.5.
- Obtaining satisfactory results.



Take home message

Adaptive predictor which combines

1. The best Linear Unbiased Predictor (BLUP) estimator.
2. The estimation of local regularity parameters for FTS.
3. The adaptive optimal estimates of mean, covariance and autocovariance.

Work in progress...

- Advanced empirical study on BLUP,
- Establishing confidence bands for the estimates, etc.

Take home message

Adaptive predictor which combines

1. The best Linear Unbiased Predictor (BLUP) estimator.
2. The estimation of local regularity parameters for FTS.
3. The adaptive optimal estimates of mean, covariance and autocovariance.

Work in progress...

- Advanced empirical study on BLUP,
- Establishing confidence bands for the estimates, etc.

Thanks for your attention !