Adaptive prediction for functional time series

Hassan Maissoro[1], Valentin Patilea[2] & Myriam Vimond[3]

[1] *CREST, Ensai, and Datastorm ; France, hassan.maissoro@datastorm.fr*
[2] *CREST, Ensai; France, valentin.patilea@ensai.fr*
[3] *CREST, Ensai; France, myriam.vimond@ensai.fr*

Résumé. Une procédure adaptative de prédiction de courbe pour une série temporelle fonctionnelle stationnaire est proposée. Les trajectoires des séries temporelles fonctionnelles sont supposées être irrégulières et sont observées avec erreur à des instants discrets. Notre prédicteur linéaire est basé sur le meilleur prédicteur linéaire sans biais (BLUP) et sur les estimateurs non paramétriques adaptatifs des fonctions de moyenne et d'autocovariance du processus. En d'autres termes, les fénêtres de lissage de ces estimateurs sont choisis de manière adaptative en fonction de la régularité locale des trajectoires. L'avantage d'une telle procédure sera une réduction du risque de prédiction par rapport aux procédures existantes.

Mots-clés. BLUP, Estimateur adaptatif, Function moyenne, Function d'autocovariance, Lissage à noyau.

Abstract. An adaptive procedure for curve prediction for a stationary functional time series is proposed. The sample paths of the functional times series are assumed to be irregular and are observed with error at discrete times in the domain. Our linear predictor is based on the best linear unbiased predictor (BLUP) and on the adaptive nonparametric mean and autocovariance functions estimators. That is, the bandwidth parameters of these estimators are chosen adaptively with respect to the local regularity of the sample paths. The benefit of such a procedure will be a reduction in risk prediction compared to existing procedures.

Keywords. BLUP, Adaptive estimator, Mean function, Autocovariance function, Kernel smoothing.

1 Introduction

In many applications, the observations corresponding to a curve are not all available at once. Consider, for example, electricity consumption curves in a household. The curve of the daily consumption is one element of the FTS. Predicting the curve could mean predicting the daily curve using the observations from the previous days. It could also mean predicting the future values on a daily curve using the observations from the previous days and the intra-day observations already available. However, the observations are in general noisy measurements of the curves, at discrete points in their domain, not necessarily regular or not necessarily the same from one curve to another. Our aim is to propose a simple and fast data-driven procedure for curve prediction for functional time series (FTS). The natural paradigm in stationary FTS is to consider the curves as sample paths of a stochastic process X, and to assume some kind of stochastic dependence between the curves.

In the literature, the most commonly studied case is where the curves are fully observed, see for example Jiao et al. (2023) and references therein. In the context of discretely observed FTS, a functional data recovery procedure have been already proposed by Rubín and Panaretos (2020) under the assumption that these functions admit at least one derivative. This procedure require estimators of the mean function and autocovariance function. However, in some cases, for example in the energy domain, the mean and autocovariance functions can be very irregular, of unknown irregularity. Several phenomena are naturally described by this type of data. This is the case for photovoltaic electricity production, which depends on the clouds. Thus, if the production of a photovoltaic park is observed for a sufficiently long period of time, it naturally generates data under the form of a set of irregular daily curves that are dependent on each other.

For irregular and weakly dependent curves, Maissoro et al. (2024) proposed new estimators of the mean and autocovariance functions that adapt to the local regularity of the underlying process X that generates the FTS $\{X_n\}$. Indeed the quality of the inference depends on the regularity of the underlying process that generated the trajectories. Here, a challenge is to take into account the local regularity parameters in order to adapt the curve prediction procedure in the context of FTS. We consider an adaptive predictor which combines the best linear unbiased predictor (BLUP) estimator and the adaptive optimal estimates of the irregular mean function and the autocovariances of the process to predict the unobserved part of the future curve.

In the following sections, we first formally present the FTS model and the data generation model we consider. Second, we introduce the adaptive linear predictor. Third, we explain how the linear predictor is estimated adaptively to the local regularity of the underlying process X that generates the FTS $\{X_n\}$.

2 The functional time series model

We consider a second-order functional time series (FTS) process $\{X_n\} = \{X_n(u),\ u \in I, n \in \mathbb{Z}\} \subset \mathcal{H}$, stochastically dependent with respect to the index n. Typically, I is a bounded domain on the real line on which the random functions are defined, and without loss of generality we consider $I = (0, 1]$. The space $\mathcal{H} = \mathbb{L}^2(I)$ is the Hilbert space of real-valued square integrable functions defined on I. Moreover, almost surely, the paths X_n are assumed to belong to the Banach space $\mathcal{C} = \mathcal{C}(I)$ of continuous functions, equipped with the sup-norm $\|\cdot\|_\infty$. The mean and the lag-ℓ, $\ell \geq 0$, (auto)covariance functions of $\{X_n\}$ are respectively

$$\mu(t) = \mathbb{E}(X(t)) \quad \text{and} \quad \Gamma_\ell(s,t) = \mathbb{E}\left\{[X_0(s) - \mu(s)][X_\ell(t) - \mu(t)]\right\}, \quad \forall s,t \in I.$$

Observation scheme. The data are obtained from sample paths realizations X_n, $n = 1, 2 \ldots,$ of a stationary time series X which are observed with error at discrete times. The data associated to the sample path X_n consists of the pairs $(Y_{n,i}, T_{n,i}) \in \mathbb{R} \times I$, representing the responses and the associated design point, which are assumed to be generated according to

$$Y_{n,i} = X_n(T_{n,i}) + \varepsilon_{n,i}, \qquad 1 \leq n \leq N, \;\; 1 \leq i \leq M_n. \tag{1}$$

Here, M_n is an integer which can be non random and common to several X_n, or randomly drawn from some distribution, independently of X. The $T_{n,i}$ are the design points for X_n, which can be non-random or randomly drawn from some distribution, independent of X and M. The case where $T_{n,i}$ are the same for several X_n, and implicitly M_n is fixed, is the so-called *common design* case. The case where the $T_{n,i}$ are random is referred to as the *independent design* case. The $\varepsilon_{n,i}$ are the measurement errors and

$$\varepsilon_{n,i} = \sigma(T_{n,i}) e_{n,i}, \quad 1 \leq i \leq M_n,$$

where $e_{n,i}$ are independent copies of a centred variable e with unit variance, and $\sigma(t)$ is an unknown bounded function that accounts for possible heteroscedastic measurement errors. The issues discussed here apply to both independent and joint design cases, unless otherwise stated.

3 Adaptive linear predictor

Let $t_0 \in I$ and $n_0 \in \{1, \ldots, N\}$ be fixed. Let $\mathcal{T}_n$ be the ordered set of design points $T_{n,i}$ where the curve X_n is measured with error as described in the model (1). Let

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

be the $M_n \times 1$ column matrix of the noisy measurements of X_n and the $M_n \times M_n$ diagonal covariance matrix of the noises $\varepsilon_{n,1}, \ldots, \varepsilon_{n,M_n}$, respectively. We aim to define the BLUP using the L lags from n_0, for $L < n_0$. Let us introduce the following notations,

$$\mathbb{M}_{n_0,L} = M_{n_0-L} + \ldots + M_{n_0-1} + M_{n_0}, \qquad \mathcal{Y}_{n_0,L} = (\mathbb{Y}_{n_0-L}^\top, \ldots, \mathbb{Y}_{n_0}^\top)^\top, \qquad \text{and}$$
$$\mathcal{M}_{n_0,L} = (\mu(T_{n_0-L,1}), \ldots, \mu(T_{n_0-L,M_{n_0-L}}), \ldots, \mu(T_{n_0,1}), \ldots, \mu(T_{n_0,M_{n_0}}))^\top.$$

Finally, the subscript M, T indicates a conditional operator (variance, covariance, mean) given M_n and $\mathcal{T}_n$, $n \geq 1$.

Following Robinson (1991), an estimator $\widehat{X}_{n_0}(t_0)$ of $X_{n_0}(t_0)$ given $\mathcal{Y}_{n_0,L}$ is BLUP if the following conditions hold true: $\widehat{X}_{n_0}(t_0)$ is a linear function of $\mathcal{Y}_{n_0,L}$; $\widehat{X}_{n_0}(t_0)$ is unbiased, *i.e.* $\mathbb{E}_{M,T}(\widehat{X}_{n_0}(t_0)) = \mu(t_0)$; and $\widehat{X}_{n_0}(t_0)$ has minimum mean square error among the class of linear unbiased estimators. The form of this linear predictor is

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

where

$$B_{n_0,L} = \mathrm{Var}_{M,T}(\mathcal{Y}_{n_0,L})^{-1}\mathbb{E}_{M,T}\left([\mathcal{Y}_{n_0,L} - \mathcal{M}_{n_0,L}][X_{n_0}(t_0) - \mu(t_0)]\right),$$

where $\mathrm{Var}_{M,T}(\mathcal{Y}_{n_0,L}) = \mathbb{G}_{n_0,L} + \Sigma_{n_0,L}$, and,

$$\Sigma_{n_0,L} = \mathrm{diag}\left(\sigma^2(T_{n_0-L,1}), \dots, \sigma^2(T_{n_0-L,M_{n_0-L}}), \dots, \sigma^2(T_{n_0,1}), \dots, \sigma^2(T_{n_0,M_{n_0}})\right)$$

$$\mathbb{G}_{n_0,L} = \begin{pmatrix} G_0^{(n_0-L,n_0-L)} & G_1^{(n_0-L,n_0-L+1)} & \cdots & G_L^{(n_0-L,n_0)} \\ G_1^{(n_0-L+1,n_0-L)} & G_0^{(n_0-L+1,n_0-L+1)} & \cdots & G_{L-1}^{(n_0-L+1,n_0)} \\ \vdots & \vdots & \ddots & \vdots \\ G_L^{(n_0,n_0-L)} & G_{L-1}^{(n_0,n_0-L+1)} & \cdots & G_0^{(n_0,n_0)} \end{pmatrix}$$

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

Let us point out that $\widehat{X}_{n_0}(t_0)$ becomes the PACE predictor when $L = 0$, see Yao et al. (2005). To further elaborate, for simplicity we hereafter study the best linear predictor using only the observations for the current curve and the $L = 1$ lagged curve. The case $L > 1$ can be handled similarly at the cost of more complex matrix algebra. When $L = 1$, we have

$$\mathrm{Var}_{M,T}\left(\mathcal{Y}_{n_0,1}\right) = \begin{pmatrix} G_0^{(n_0-1,n_0-1))} + \Sigma_{n_0-1} & G_1^{(n_0-1,n_0))} \\ G_1^{(n_0,n_0-1))} & G_0^{(n_0,n_0))} + \Sigma_{n_0} \end{pmatrix} \in \mathbb{R}^{M_{n_0}+M_{n_0-1}} \times \mathbb{R}^{M_{n_0}+M_{n_0-1}},$$

and

$$\mathrm{Cov}_{M,T}\left(\mathcal{Y}_{n_0,1}, X_{n_0}(t_0)\right) = \begin{pmatrix} \Gamma_1(T_{n_0-1,1}, t_0) \\ \vdots \\ \Gamma_1(T_{n_0-1,M_{n_0-1}}, t_0) \\ \Gamma_0(T_{n_0,1}, t_0) \\ \vdots \\ \Gamma_0(T_{n_0,M_{n_0}}, t_0) \end{pmatrix} \in \mathbb{R}^{M_{n_0}+M_{n_0-1}}.$$

We propose adaptive estimators for the vector $\mathrm{Cov}_{M,T}\left(\mathcal{Y}_{n_0,1}, X_{n_0}(t_0)\right)$ and the matrix $\mathrm{Var}_{M,T}\left(\mathcal{Y}_{n_0,1}\right)$. It is worth noting that the nonparametric estimator of $G_0^{(n,n)} + \Sigma_n$ can be easily constructed by the 'first smooth, then estimate' approach, and it does not require a diagonal correction due to the noise variance, as would have been the case for the estimation of Γ_0. More precisely, outside the diagonal we estimate $\mathbb{E}[X(s)X(t)]$ by the average of the products of the nonparametric estimates of $X(s)$ and $X(t)$. For the diagonal values of the matrix $G_0^{(n,n)} + \Sigma_n$ we simply smooth the squares of the centered $Y_{n,i}$, $1 \leq i \leq M_n$.

The adaptive estimation of the linear predictor involves the estimation of the matrix $\text{Var}_{M,T}(\mathcal{Y}_{n_0,1})$ and the vector $\text{Cov}_{M,T}(\mathcal{Y}_{n_0,1}, X_{n_0}(t_0))$ adaptive to the regularity of the stationary process X that generates the FTS $\{X_n\}$. This is equivalent to estimating the mean function μ, the covariance function Γ_0 and the lag-1 autocovariance function Γ_1, which adapt to the local regularity of X. We assume that the sample paths X_n are not differentiable almost surely. Functional data from fields such as energy, environment, chemistry and physics, medical devices, meteorology, are often very irregular and thus considering non-differentiable curves X_n seems realistic.

4 Estimation of the adaptive linear predictor

We use the procedure of adaptive mean and autocovariance function estimation proposed by Maissoro et al. (2024) under the assumption that X admits a local regularity and that the FTS $\{X_n\}$ satisfy a weak dependency assumption. We first give an insight into the local regularity and weak dependence assumptions before proving the definition of these estimators.

Local regularity parameters. The assumption that X admits a *local regularity* in the neighbourhood of $t \in I$ allows estimates based on $\{X_n\}$, which adapt to the regularity of the underlying process X. The local regularity for stationary FTS was considered by Maissoro et al. (2024) and was proposed and studied by Golovkine et al. (2022) in the context of independent samples of random curves. The process X admits a *local regularity* at $t \in I$ with a local exponent $H_t \in (0, 1)$ and a local Hölder constant L_t^2 if a constant $\beta > 0$ exists and for any $t \in I$ such that

$$\mathbb{E}\left[\{X(u) - X(v)\}^2\right] = L_t^2 |u - v|^{2H_t} \{1 + O(|u - v|^\beta)\}, \tag{2}$$

when $u \leq t \leq v$ lie in a small neighborhood of t. The procedure for estimating the local regularity parameters H_t and L_t^2 as well as examples of processes satisfying (2), including but not limited to the Functional AutoRegressive process of order one where the white noise belongs to the class of multifractional Brownian motion, can be found in Maissoro et al. (2024).

Weak dependence. We use the concept of $\mathbb{L}^p_{\mathcal{C}} - m-$approximability considered by Maissoro et al. (2024) for random processes valued in $(\mathcal{C}, \|\cdot\|_\infty)$. It is a redefinition of the concept of $\mathbb{L}^p_{\mathcal{H}} - m-$approximability introduced by Hörmann and Kokoszka (2010) in the context of random processes valued in $\mathcal{C}$ instead of $\mathcal{H}$. As a result, the weak dependence between the curves X_n is transferred to the sequences $\{X_n(t)\}$ for all $t \in I$. Consequently, this general notion of weak dependence allows the localised study of FTS, which is necessary for inference on local regularity parameters as well as on the mean and autocovariance function estimators. The underlying idea of this weak dependence is to approximate $\{X_n\}$ with an m-dependent sequence $\{X_n^{(m)}, m \geq 1\}$ such that, for any $n \in \mathbb{Z}$, the sequence $\{X_n^{(m)}, m \geq 1\}$ converges to X_n in some manner as $m \to \infty$. Therefore, the behavior of the original process can be determined by observing the behavior of its coupled m-dependent sequences, provided they are sufficiently close to the original process. A formal definition and examples of common FTS models that satisfy $\mathbb{L}^p_{\mathcal{C}} - m-$approximability, including but not limited to a revised FAR(1) process, the Functional Linear Process, the Product Model, and the Functional ARCH, can be found in Maissoro et al. (2024).

Adaptive mean and (auto)covariance estimation. Let us consider a stationary FTS $\{X_n\} \subset \mathbb{L}^p_{\mathcal{C}}$, for some $p \geq 4$, defined on $I = (0,1]$. Let $\widehat{X}_1(t;h), \widehat{X}_2(t;h), \dots, \widehat{X}_N(t;h)$ be the Nadaraya-Watson (NW) estimators of $X_1(t), X_2(t), \dots, X_N(t)$ obtained using a bandwidth parameter h considered in some set of bandwidths $\mathcal{H}_N$. Maissoro et al. (2024) follow the '*smooth first, then estimate*' approach and define the estimator of the mean and the autocovariance function $\Gamma_\ell(s,t)$ with $s \neq t$ using empirical estimators where the original trajectories are replaced with the NW ones. For the mean function, at each $t \in I$ an optimal bandwidth parameter h^*_μ is estimated by minimising an estimated version of a risk function $R_\mu(t;h)$ over the grid $\mathcal{H}_N$ and used to estimate $\mu(t)$. The risk function is a sharp upper bound on the quadratic risk of the $\widehat{\mu}_N(t;h)$ estimator of the mean function, which is defined below. For any $t \in I$, let

$$\pi_n(t;h) = 1 \quad \text{if} \quad \sum_{i=1}^{M_n} \mathbb{1}\{|T_{n,i} - t| \leq h\} \geq 1, \qquad \text{and} \quad \pi_n(t;h) = 0 \text{ otherwise},$$

where $\mathbb{1}\{\cdot\}$ denote the indicator function. The adaptive mean function estimator is $\widehat{\mu}^*_N(t) = \widehat{\mu}_N(t; h^*_\mu)$ where

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

The bandwidth h^*_μ is chosen to minimize an estimated version of $R_\mu(t;h)$, namely

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

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

$$\mathbb{V}_\mu(t;h) = \sum_{n=1}^{N} \frac{\pi_n(t;h)}{P_N^2(t;h)} c_n(t;h) \max_{1\leq i\leq M_n} |W_{n,i}(t;h)|, \text{ with } c_n(t;h) = \sum_{i=1}^{M_n} |W_{n,i}(t;h)|,$$

$$\mathbb{B}(t;h,\alpha) = \sum_{n=1}^{N} \frac{\pi_n(t;h)}{P_N(t;h)} c_n(t;h) b_n(t;h,\alpha), \text{ with } b_n(t;h,\alpha) = \sum_{i=1}^{M_n} \left|\frac{T_{n,i}-t}{h}\right|^\alpha |W_{n,i}(t;h)|,$$

$$\mathbb{D}_\mu(t;h) = \mathbb{E}\left[\{X_0(t)-\mu(t)\}^2\right] + 2\sum_{\ell=1}^{N-1} p_\ell(t;h)\mathbb{E}\left(\{X_0(t)-\mu(t)\}\{X_\ell(t)-\mu(t)\}\right),$$

$$\text{where} \quad p_\ell(t;h) = \sum_{i=1}^{N-\ell} \frac{\pi_i(t;h)\pi_{i+\ell}(t;h)}{P_N(t;h)},$$

and a simple and consistent choice of the estimator of errors' variance $\sigma^2(t)$ at t is

$$\widehat{\sigma}^2(t) := \frac{1}{N}\sum_{n=1}^{N} \frac{1}{2}\left(Y_{n,i(t)} - Y_{n,i(t)+1}\right)^2,$$

where, for each n, $i(t), i(t)+1$ are the indices of the two closest domain points $T_{n,i}$ to t.

The adaptive autocovariance estimator is $\Gamma^*_{N,\ell}(s,t) = \Gamma^*_{N,\ell}(s,t;h^*_\Gamma)$ where

$$\widehat{\Gamma}_{N,\ell}(s,t;h) = \sum_{n=1}^{N-\ell} \frac{\pi_n(s;h)\pi_{n+\ell}(t;h)}{P_{N,\ell}(s,t;h)} \widehat{X}_n(s;h)\widehat{X}_{n+\ell}(t;h) - \mu^*_N(s)\mu^*_N(t),$$

where $\widehat{X}_n(s;h)$ and $\widehat{X}_{n+\ell}(t;h)$ are Nadaraya-Watson (NW) estimators of $X_n(s)$ and $X_{n+\ell}(t)$ respectively, and

$$P_{N,\ell}(s,t;h) = \sum_{n=1}^{N-\ell} \pi_n(s;h)\pi_{n+\ell}(t;h).$$

Like in the case of mean estimation, the optimal bandwidth h^*_Γ is defined as the optimum of an explicit risk bound, similar to $\widehat{R}_\mu(t;h)$. As a variation of the existing method, we further investigate how two bandwidth parameters can be allowed in autocovariance estimation.

References

Golovkine, S., Klutchnikoff, N., and Patilea, V. (2022). Learning the smoothness of noisy curves with application to online curve estimation. *Electron. J. Stat.*, 16(1):1485–1560.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.*, 38(3):1845–1884.

Jiao, S., Aue, A., and Ombao, H. (2023). Functional time series prediction under partial observation of the future curve. *J. Amer. Statist. Assoc.*, 118(541):315–326.

Maissoro, H., Patilea, V., and Vimond, M. (2024). Learning the smoothness of weakly dependent functional time series. *Working Paper CREST-Ensai. arXiv preprint arXiv:202402.xxxxx.*

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statist. Sci.*, 6(1):15–51. With comments and a rejoinder by the author.

Rubín, T. and Panaretos, V. M. (2020). Sparsely observed functional time series: estimation and prediction. *Electron. J. Stat.*, 14(1):1137–1210.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.*, 100(470):577–590.