Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

On optimal prediction of missing functional data with memory

Van Bever, Germain; Ilmonen, Pauliina; Viitasaari, Lauri; Shafik, Nourhan; Sottinen, Tommi

Publication date:
2021

Document Version
Early version, also known as pre-print

Link to publication

Citation for pulished version (HARVARD):
Van Bever, G, Ilmonen, P, Viitasaari, L, Shafik, N & Sottinen, T 2021 'On optimal prediction of missing functional data with memory'.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. Oct. 2022

Submitted to the Annals of Statistics

ON OPTIMAL PREDICTION OF MISSING FUNCTIONAL DATA WITH MEMORY

By Pauliina Ilmonen* and Nourhan Shafik* and Tommi Sottinen† and Germain Van Bever‡ and Lauri Viitasaari§

Aalto University, University of Vaasa and University of Namur

This paper considers the problem of reconstructing missing parts of functions based on their observed segments. It provides, for Gaussian processes and arbitrary bijective transformations thereof, theoretical expressions for the L^2-optimal reconstruction of the missing parts. These functions are obtained as solutions of explicit integral equations. In the discrete case, approximations of the solutions provide consistent expressions of all missing values of the processes. In the case of Gaussian processes with a parametric covariance structure, the estimation can be conducted separately for each function, and yields nonlinear solutions in presence of memory. Simulated examples show that the proposed reconstruction indeed fares better than the conventional interpolation methods in various situations.

1. Introduction. In functional data analysis, the observed units are random curves $(Y^1)_{t\in\mathcal{I}}, \dots, (Y^J)_{t\in\mathcal{I}}$ defined on some domain $\mathcal{I}$. The standard setting, also adopted throughout this paper, is to assume that t represents the time at which the functions are observed and that $\mathcal{I} = [0, T] \subset \mathbb{R}$, for $T > 0$. There is a vast body of work in functional data analysis, which often extends classical multivariate techniques to this particular setting. See, for example, Ramsay and Silverman (2005), Ferraty (2011), Wang, Chiou and Müller (2016) and references therein.

While many procedures assume that the curves are fully observed, this will not be the case in most instances so that one has to proceed through a reconstruction step. The need for this step comes most often from the fact that only discretized measurements from each curve are available. More precisely, the collected data takes the form

$$Y_{ji} = Y^j_{t_{ji}},\ j = 1, \dots, J,\ i = 1, \dots, N_j,$$

where, in full generality, the measurement times t_{ji} could vary in number or location within the curves themselves. Recovering the curves from their discrete measurements has

*Aalto University School of Science, Department of Mathematics and Systems Analysis.

†University of Vaasa, Department of Mathematics and Statistics.

‡Corresponding author; University of Namur, Department of Mathematics and the Namur Institute for Complex Systems (naXys).

§Aalto University School of Business, Department of Information and Service Management.

MSC 2010 subject classifications: Primary 62R10; secondary 60G15; 62D10

Keywords and phrases: Gaussian processes, Missing data, Prediction, Regular conditional law

been extensively explored in the statistical literature. Reconstruction methods typically depend on assumptions on the generating process of the curves $Y(t)$. They also, and most importantly, depend on the way the functions are discretized, that is, on the nature of the locations t_{ji}.

The most classical example is to assume that $\{t_{ji} \ : \ j = 1, \dots, J, ; i = 1, \dots, N_j\}$ is increasingly dense in $[0, T]$ as $\min_j N_j$ gets larger. This will be the case, for example, for regularly observed data (i.e. $t_{ji} = iT/N$, $i = 1, \dots, N$ for some N) or for dense randomly observed data (i.e. $t_{ji} \sim \mathcal{U}([0, T])$ are i.i.d. uniform random locations in $\mathcal{I}$, with N_j large). In such situation, one typically proceeds with standard smoothing techniques, which include classical penalized regression on spline or Fourier basis functions (see Ferraty and Vieu (2006) or Ullah and Finch (2013)). In the more general setting of sparse functional data (for which N_j is small and t_{ji} are still i.i.d. uniform random variables), one can proceed with estimating the common mean and covariance functions of $(Y^j)_{t\in[0,T]}$ and reconstructing the functions under the normality assumption (see, for example, James, Hastie and Sugar (2000) and Yao, Müller and Wang (2005)).

Another situation which requires reconstructing the functional observation is the case of *fragmented data*. Such data arises in situations where the curves Y^j can not reasonably be observed regularly or randomly in the whole time domain $[0, T]$, but rather on a subinterval (or union thereof) $\Delta_j \subset [0, T]$. Each curve is observed, either discreetly or fully, on Δ_j. Treated as fragments of a general function, the objective is then to reconstruct Y^j on $[0, T]$. Proposals of curve reconstruction include Delaigle and Hall (2016), but also Delaigle and Hall (2013), in the classification setting or Liebl (2013) and Goldberg, Ritov and Mandelbaum (2014), for prediction purposes. Similarly, reconstructing the covariance operator of Y from fragmented data has been considered in Descary and Panaretos (2019) and Delaigle et al. (2020).

There also exists several proposals for *optimal* reconstruction of fragmented data. They include Kraus (2015), Liebl and Rameseder (2019) or, more recently, Kneip and Liebl (2020). In the latter, the authors construct a linear operator L which minimizes the mean squared error loss $\mathbb{E}[\{Y_s^j - [L((Y^j)_{t\in\Delta_j})]_s\}^2]$ at any $s \notin \Delta_j$. The estimation of L is based on $\{(Y^j)_{t\in\Delta_j} : j = 1, \dots, J\}$. Importantly, in all the references above, the reconstruction of Y^j is only possible given the knowledge of the full dataset.

The approach adopted in this paper breaks from the previous literature, in that it aims at providing optimal reconstruction of missing fragments $(Y^j)_{s\notin\Delta_j}$ solely based on $(Y^j)_{t\in\Delta_j}$ in the context of stochastic processes with memory. Throughout, we will assume that the observed set is a union of interval of the form

$$\Delta = [t_L^1, t_U^1] \cup [t_L^2, t_U^2] \cup \dots [t_L^n, t_U^n]. \tag{1.1}$$

More precisely, the problem under consideration is to estimate the conditional expectation $\hat{Y}_s^j := \mathbb{E}[Y_s^j | \mathcal{F}_\Delta^j]$, where Y_s^j is a missing value of the function Y^j and $\mathcal{F}_\Delta^j$ is the filtration generated by the observed parts of Y^j itself. Note that $(\hat{Y}^j)_{s\in[0,T]}$ is indeed the

optimal L^2 reconstruction of Y^j in the sense that it minimizes

$$\int_0^T \mathbb{E}[(Y_s^j - \hat{Y}_s^j)^2 | \mathcal{F}_\Delta^j] ds.$$

The contributions of this paper are threefold. In the first part, we provide an explicit expression for $\mathbb{E}[X_t|\mathcal{F}_\Delta]$, where $(X)_{t\in[0,T]}$ is a centered and separable Gaussian process that admits an integral representation. More precisely, under mild assumptions (discussed in Section 2), we make use of the Fredholm representation of X,

$$X_t = \int_0^T K(t,s) dW_s,$$

with $K(t,s) \in L^2([0,T]^2)$ a deterministic kernel and W a Brownian motion. This in turn allows us to derive a Wiener integral representation of $\mathbb{E}[X_t|\mathcal{F}_\Delta]$, where the kernel is shown to satisfy certain integral equations. Importantly, in such setting, the curve reconstruction can be conducted on a curve-by-curve basis and presents a straightforward empirical expression.

In the second part, we provide explicit expressions for $\mathbb{E}[Y_s^j|\mathcal{F}_\Delta^j]$ under the assumption that $Y_t^j = f(t, X_t^j)$, $t \in [0,T]$, is some bijective transformation of a Gaussian process $(X^j)_{t\in[0,T]}$. Also in this situation, we show that, once the integral representation of X^j is known, the regular conditional law of Y^j can be computed by solving certain integral equations. Note that processes of the type $Y_t = f(t, X_t)$ form a large and flexible class of models. Indeed, it is known (see, e.g. Viitasaari and Ilmonen, 2020) that the processes in this class can have arbitrary one-dimensional marginal distributions and approximate arbitrarily well any covariance structure.

Our third contribution is to provide a method to estimate both the bijective transformation and the solution to the integral equations, allowing us to estimate optimal L^2-predictors with minimal assumptions. Interestingly, if f is known (as is the case, for example, in the first part) the reconstruction can be conducted separately for each functional observation provided. This highlights the disentanglement between, on one side, the estimation of f which requires the whole dataset, and, on the other side, the reconstruction of $(X^j)_{t\notin\Delta_j}$ on the sole basis of $(X^j)_{t\in\Delta_j}$. We illustrate the estimation procedure for several Gaussian processes which include fractional Brownian motions. These processes have been studied intensively during the last decades, see, for example Mishura (2008).

The rest of the paper is organized as follows. In Section 2, we provide optimal predictors for Gaussian processes. Throughout Section 2, we assume that the underlying integral representation structure of the Gaussian process is known. In Section 3, we drop the assumption of Gaussianity and known related integral structures. In particular, we explain how our approach can be used to approximate the optimal predictor by using the observed data directly and with posing minimal assumptions. We end the paper with a simulation study provided in Section 4. The appendix to this paper is in two parts. In Appendix A, we recall some preliminaries on Gaussian analysis and provide some necessary results that

guarantee the existence of solutions to our integral equations. Finally, all the proofs are contained in Appendix B.

2. Bridge prediction laws for partially observed Gaussian processes. Let $X = (X_t)_{t\in[0,T]}$ be a centered, i.e. be such that $\mathbb{E}[X_t] = 0$ for all $t \in [0,T]$, and separable Gaussian process on a probability space $(\Omega, \mathcal{F}(X), \mathbb{P})$. Furthermore, we assume throughout that the variance of X is uniformly bounded, i.e.

$$\sup_{t\in[0,T]} \mathbb{E}[X_t^2] < \infty. \tag{2.1}$$

We stress that assuming uniformly bounded variance is a very mild condition. Most reasonable examples indeed fulfill (2.1). As a particular example, continuity of X (on the compact interval $[0,T]$, hence uniform continuity) is sufficient to guarantee both separability and (2.1).

If a centered and separable Gaussian process X has an integrable variance function, i.e.

$$\int_0^T \mathbb{E}[X_t^2]dt < \infty, \tag{2.2}$$

then there exists (for details, we refer to Appendix A and references therein) a standard Brownian motion W and a deterministic kernel $K \in L^2\left([0,T]^2\right)$ such that

$$X_t = \int_0^T K(t,s)dW_s \tag{2.3}$$

holds in law. Note that (2.2) trivially holds under (2.1). Rephrasing, under (2.1), we assume that X belongs to the separable Hilbert space $L^2([0,T])$ almost surely.

REMARK 2.1. The integral in (2.3) stands for the Wiener integral with respect to the Brownian motion. In general, Wiener integrals

$$\int_0^T f(s)dW_s$$

can be defined for any function $f \in L^2([0,T])$. A more complete exposure on the topic is available in Appendix A.

For $t \in [0,T]$, let H_t^X be the first chaos of X, that is the closure (in $L^2(\Omega)$) of the linear space spanned by $X_s, s \in [0,t]$ (see also Definition A.5). We introduce the following assumption that stands throughout the article.

ASSUMPTION 2.1. The centered and separable Gaussian process X satisfies (2.1). Furthermore, there exist a kernel K and a Brownian motion W such that (2.3) holds exactly, and $H_T^X = H_T^W$.

Assumption 2.1 is a natural assumption and certainly not very restrictive. Indeed, whenever (2.1) is satisfied, we obtain that (2.3) holds in law. Assumption 2.1 therefore merely states that the underlying Brownian motion drives the process X directly and (2.3) holds as an equality instead of only as a representation in law. The additional condition $H_T^X = H_T^W$ is also very natural. Indeed, this simply means that, when the whole interval $[0,T]$ is concerned, the driving Brownian motion carries the same amount of randomness as X itself. In view of Karhunen-Loève expansion, this is essentially true as long as X is not finite dimensional – a case that is not particularly interesting for our purposes. We also remark that, in many interesting cases, the stronger relation $H_t^X = H_t^W$ holds for all $t \in [0,T]$. This is the case in all the examples considered in Section 4.

An interesting and widely studied subclass of processes satisfying Assumption 2.1 are called Volterra Gaussian processes. They admit a representation

$$X_t = \int_0^t K(t,s)dW_s. \tag{2.4}$$

The expression (2.3) is called the "Fredholm representation of X", stemming from the fact that X_t can only be computed with the knowledge of the kernel $K(t,s)$ for all $s \in [0,T]$. Similarly, (2.4) is coined the "Volterra representation of X" and its kernel $K(t,s)$ is of Volterra-type, that is, is a Fredholm kernel such that $K(t,s) = 0$ for $s > t$.

EXAMPLE 2.1. Let X be an H-self-similar Gaussian process, i.e. $X_{at} \overset{\mathcal{L}}{=} a^H X_t$ for all $a \geq 0$ and all $t \geq 0$. Then $H_t^X = H_t^W$ for all $t \in [0,T]$ if and only if X has a Volterra representation (2.4). The latter holds whenever X is purely non-deterministic. That is, when $\bigcap_{t>0} H_t^X$ is trivial, i.e. spanned only by constants. For details and for the proof of these facts, we refer to Yazigi (2015).

We are interested in the prediction law $X|\mathcal{F}_\Delta$, where

$$\Delta = [t_L^1, t_U^1] \cup [t_L^2, t_U^2] \cup \ldots \cup [t_L^n, t_U^n] \tag{2.5}$$

for some $0 \leq t_L^1 < t_U^1 < t_L^2 < t_U^2 < \ldots t_L^n < t_U^n \leq T$ and $\mathcal{F}_\Delta$ is the σ-algebra generated by X on Δ, i.e.

$$\mathcal{F}_\Delta = \sigma(X_t : t \in \Delta).$$

Setting $t_U^0 = 0$ and $t_L^{n+1} = T$, we aim at reconstructing the missing values on each missing subinterval $(t_U^k, t_L^{k+1}), k = 0, \ldots, n$. From a practical point of view, this means that we observe X_s for $s \in \Delta$ only, and our aim is to predict the missing values at $s \in [0,T] \setminus \Delta$.

The following result provides the best predictor in the $L^2(\Omega)$-sense.

THEOREM 2.1. *Let X be a stochastic process satisfying Assumption 2.1 and let $\mathcal{F}_\Delta = \sigma(X_t : t \in \Delta)$. Then for each $t \in [0,T] \setminus \Delta$ we have*

$$\hat{X}_t = \mathbb{E}\left[X_t|\mathcal{F}_\Delta\right] = \int_0^T f_t(s)dW_s, \tag{2.6}$$

where $f_t \in L^2([0,T])$ *is the solution of the (system of) Fredholm integral equations*

$$\int_0^T f_t(s)K(u_k,s)ds = \int_0^T K(t,s)K(u_k,s)ds, \tag{2.7}$$

for $k = 1,\dots,n$ *and for all* $t_L^k \le u_k < t_U^k$, *that satisfies*

$$\int_0^T f_t(s)g_\xi(s)ds = 0, \text{ for all } g_\xi \in L^2([0,T]) \text{ such that } \int_0^T g_\xi(s)K(u_k,s)ds = 0.$$

REMARK 2.2. In general, integral equations of the first kind such as (2.7) are ill-posed problems and uniqueness of the solution is not guaranteed. Indeed, from $H_T^X = H_T^W$ it follows that each $\xi \in H_T^X$ can be represented as $\xi = \int_0^T g_\xi(s)dW_s$ for some $g_\xi \in L^2([0,T])$. Consequently, we may choose any $\xi = \int_0^T g_\xi(s)dW_s \in H_T^X$ orthogonal to $\mathcal{F}_\Delta$, and, since then $\int_0^T g_\xi(s)K(u_k,s)ds = 0$ for $t_L^k \le u_k < t_U^k$, we observe that $f_t + g_\xi$ solves (2.7) whenever f_t does. As g_ξ can, in the general case, be chosen rather arbitrarily, the solution to (2.7) is clearly not unique. On the other hand, there is exactly one solution f_t of (2.7) such that (2.6) holds (see the proof of Theorem 2.1 in Appendix B). Indeed, existence and uniqueness of the conditional expectation allows to fix a unique representative within the set of solutions of (2.7).

REMARK 2.3. The conditional mean in (2.6) is expressed as a Wiener integral with respect to the driving Brownian motion W. Such an expression is useful when one is interested in theoretical properties of $\hat{X}$ or for simulation purposes. In practice, however, the driving Brownian motion is not observable on $[0,T]$ and thus one needs to transform (2.6) into something computable from the observations $X_s, s \in \Delta$, directly. The approximation of $\hat{X}_t$ in terms of the process X itself is discussed in Section 3.

The result below provides the regular law of X, conditional to $\mathcal{F}_\Delta$. It will prove useful when studying non-Gaussian processes in Section 3.

THEOREM 2.2. *Let* X *be a stochastic process satisfying Assumption 2.1 and let* $\mathcal{F}_\Delta = \sigma(X_t : t \in \Delta)$. *Then, the regular conditional law* $X|\mathcal{F}_\Delta$ *is Gaussian with random mean given by (2.6) and deterministic covariance given, for* $t,s \in [0,T] \setminus \Delta$, *by*

$$\rho(t,s|\Delta) = \int_0^T [K(t,u) - f_t(u)]\,[K(s,u) - f_s(u)]\,du, \tag{2.8}$$

where f *is the unique solution of (2.7) in Theorem 2.1.*

The next theorem particularizes Theorem 2.1 to the case of Volterra Gaussian processes.

THEOREM 2.3. *Let* X *be a Volterra process satisfying Assumption 2.1 with representation (2.4) and let* $\mathcal{F}_\Delta = \sigma(X_t : t \in \Delta)$. *Then, for each* $t \in [0,T] \setminus \Delta$ *we have*

$$\hat{X}_t = \mathbb{E}[X_t|\mathcal{F}_\Delta] = \int_0^T f_t(s)dW_s, \tag{2.9}$$

where, for each $k = 1, 2, \ldots, n$ *and* $t_L^k \leq s \leq t_U^k$, *the function* $f_t \in L^2([0,T])$ *is a solution of the recursive system of Volterra integral equations of the first kind given by, for* $t_L^k \leq u_k \leq t_U^k$,

$$\int_{t_L^k}^{u_k} [f_t(s) - K(t,s)]K(u_k,s)ds = \int_0^{t_L^k} [K(t,s) - f_t(s)]K(u_k,s)ds \tag{2.10}$$

with the boundary conditions

$$\int_{t_U^{k-1}}^{t_L^k} f_t(s)K(t_L^k,s)ds = \int_0^{t_L^k} K(t,s)K(t_L^k,s)ds - \int_0^{t_U^{k-1}} f_t(s)K(t_L^k,s)ds. \tag{2.11}$$

REMARK 2.4. As in the case of Theorem 2.1, the solution of (2.10)-(2.11) is not unique. However, only one amongst them satisfies (2.9). It can be determined imposing the same condition as in Theorem 2.1.

The following corollary covers the case $n = 2$, $t_L^1 = 0$, and $t_U^2 = T$, i.e. the case where $\Delta = [0, t_U^1] \cup [t_L^2, T]$ is such that only one subinterval is missing in $[0,T]$. In that situation, it provides the unique kernel f_t satisfying (2.9) of Theorem 2.3 under the hypothesis that $H_t^X = H_t^W$ for all $t \in [0,T]$.

COROLLARY 2.1. *Let* X *be a stochastic process satisfying Assumption 2.1 and let* $\mathcal{F}_\Delta = \sigma(X_t : t \in \Delta)$, *with* $\Delta = [0, t_U] \cup [t_L, T]$. *Suppose further that* $H_t^X = H_t^W$ *for all* $t \in [0,T]$. *Then for all* $t \in [t_U, t_L]$ *we have*

$$\hat{X}_t = \mathbb{E}[X_t|\mathcal{F}_\Delta] = \int_0^{t_U} K(t,s)dW_s + c(t)\int_{t_U}^{t_L} K(t_L,s)dW_s + \int_{t_L}^{T} g_t(s)dW_s, \tag{2.12}$$

where $c(t)$ *is given by*

$$c(t) = \frac{\int_{t_U}^{t} K(t,s)K(t_L,s)ds}{\int_{t_U}^{t_L} K^2(t_L,s)ds} \tag{2.13}$$

and $g_t(s)$ *is the solution to the Volterra integral equation of the first kind given by*

$$\int_{t_L}^{u} g_t(s)K(u,s)ds = \int_{t_U}^{t_L} \left[K(t,s) - c(t)K(t_L,s)\right] K(u,s)ds. \tag{2.14}$$

The expression (2.12) decomposes the dependence of $\hat{X}_t$ on the process X in three sub-intervals. Note, however, that the assumption $H_t^X = H_t^W$ for all $t \in [0,T]$, guarantees

$$\int_0^{t_U} K(t,s)dW_s + c(t)\int_{t_U}^{t_L} K(t_L,s)dW_s = \int_0^{t_U} K(t,s) - c(t)K(t_L,s)dW_s + c(t)X_{t_L},$$

so that (2.12) rewrites

$$\hat{X}_t = \mathbb{E}[X_t|\mathcal{F}_\Delta] = \Big(\int_0^{t_U} K(t,s) - c(t)K(t_L,s)dW_s\Big) + \Big(c(t)X_{t_L} + \int_{t_L}^{T} g_t(s)dW_s\Big).$$

In particular, $\int_0^{t_U} K(t,s) - c(t)K(t_L,s)dW_s \in H_{t_U}^X$ is measurable as well. This removes any explicit integration within the missing interval $[t_U, t_L]$. Finally, note that computing $\hat{X}_t$ remains difficult as, indeed, the quantity $\int_{t_L}^T g_t(s)dW_s$ can not easily be expressed as a function of the observable process X. We address this remaining difficulty in the empirical case using a discretization argument; see Section 3.

The process $\hat{X}_t$ detailed in Corollary 2.1 is a Gaussian bridge, since $\hat{X}_{t_U} = X_{t_U}$ and $\hat{X}_{t_L} = X_{t_L}$. Indeed, this can be seen by setting $t = t_U$ in (2.12), which in turn gives $c(t_U) = 0$ and g_{t_U}, a solution to $\int_{t_L}^u g_{t_U}(s)K(u,s)ds = 0$, leading to $g_{t_U} \equiv 0$. This shows

$$\hat{X}_{t_U} = \int_0^{t_U} K(t_U, s)dW_s = X_{t_U}.$$

Similarly, for $t = t_L$, we again have $\int_{t_L}^u g_{t_L}(s)K(u,s)ds = 0$, giving $g_{t_L} \equiv 0$. Also, $c(t_L) = 1$ and the first two terms give

$$\hat{X}_{t_L} = \int_0^{t_U} K(t_L, s)dW_s + \int_{t_U}^{t_L} K(t_L, s)dW_s = X_{t_L}.$$

REMARK 2.5. Note that, in the general situation of Theorem 2.1, the conditional expectation $\hat{X}_t = \mathbb{E}[X_t|\mathcal{F}_\Delta]$ is a (generalized) Gaussian bridge (see Sottinen and Yazigi (2014)). Note also that, even in the setting of Theorem 2.3 with $n \geq 3$, showing that $\hat{X}_{t_U^i} = X_{t_U^i}$, $i = 1, \ldots, n$ and $\hat{X}_{t_L^j} = X_{t_L^j}$, $j = 1, \ldots, n$ is not as straightforward (compared to the case $n = 2$, that is, under Corollary 2.1). The same argument as above holds for $t = t_L^n$ and $t = t_U^1$, though. This is not the case for $t = t_U^j$, $j = 2, \ldots, n$ (or $t = t_L^i$, $i = 1, \ldots, n-1$), which cannot be covered by simply plugging into (2.10). It can be showed that the only solution to (2.10) and (2.11) satisfying the unicity condition satisfies $f_{t_U^k}(s) = K(t_U^k, s)$ and $f_{t_L^k}(s) = K(t_L^k, s)$, for all $k = 1, \ldots, n$. This gives $\hat{X}_{t_U^k} = X_{t_U^k}$ and $\hat{X}_{t_L^k} = X_{t_L^k}$ as expected.

The following corollary provides the future prediction laws by setting $n = 1$ and $t_L^1 = 0$.

COROLLARY 2.2. *Let X be a stochastic process satisfying Assumption 2.1 and let $\mathcal{F}_\Delta = \sigma(X_t : t \in \Delta)$, with $\Delta = [0, t_U]$. Suppose further that $H_t^X = H_t^W$ for all $t \in [0, T]$. Then for all $t > t_U$ we have*

$$\hat{X}_t = \mathbb{E}[X_t|\mathcal{F}_\Delta] = \int_0^{t_U} K(t, s)dW_s. \tag{2.15}$$

REMARK 2.6. Corollary 2.2 can be viewed as a generalization of the results provided in Sottinen and Viitasaari (2017), where the authors studied the future prediction law of fractional Brownian motions. Contrary to the latter, we here rely on functional analytic argument of invertibility of bounded linear operators. We also note that, without the additional assumption $H_t^X = H_t^W$, it is not clear whether (2.15) is measurable with respect to $\mathcal{F}_\Delta$.

EXAMPLE 2.2. Let W be a Brownian motion, i.e. $K(t,s) = \mathbf{1}_{s\le t}$. By direct computations, for $t \in (t_U^{j-1}, t_L^j)$, the solution is given by

$$f_t(s) = \mathbf{1}_{s\le t_U^{j-1}} + \frac{t - t_U^{j-1}}{t_L^j - t_U^{j-1}} \mathbf{1}_{t_U^{j-1}<s<t_L^j}.$$

Thus

$$\hat{W}_t = W_{t_U^{j-1}} + \frac{t - t_U^{j-1}}{t_L^j - t_U^{j-1}} (W_{t_L^j} - W_{t_U^{j-1}}),$$

i.e. the optimal predictor is linear within each missing interval.

Finally, we consider the approximation of (2.15) in practice, where X is only observed on the discrete set

$$\Delta_N = \{t_j : t_j \in \Delta, j = 1, 2, \dots, N\} \subset \Delta.$$

Let $X_N \in \mathbb{R}^N$ be the observed vector with components $(\mathbf{X}_N)_j = X_{t_j}$, $t_j \in \Delta_N$. Let $\mathbf{R}_N \in \mathbb{R}^{N\times N}$ denote the covariance matrix of $\mathbf{X}_N$ and let $\mathbf{b}_N(t) \in \mathbb{R}^N$, $t \in [0,T]$ denote the vector consisting of covariances $\mathbb{E}[X_t X_{t_j}], t_j \in \Delta_N$. Using representation (2.3), we thus have

$$(\mathbf{R}_N)_{ij} = \int_0^T K(t_i, s) K(t_j, s) ds$$

and

$$(\mathbf{b}_N(t))_j = \int_0^T K(t, s) K(t_j, s) ds.$$

THEOREM 2.4. *Suppose X satisfies Assumption 2.1 and that $\boldsymbol{R}_N$ has full rank. Then for any $t \in [0,T] \setminus \Delta_N$ we have*

$$\hat{X}_{t,N} = \mathbb{E}\left[X_t | \mathcal{F}_{\Delta_N}\right] = \sum_{t_j \in \Delta_N} a_j(t) X_{t_j}, \tag{2.16}$$

where $\boldsymbol{a}(t) = (a_1(t), \dots, a_N(t))^T$ is given by

$$\boldsymbol{a}(t) = \boldsymbol{R}_N^{-1} \boldsymbol{b}_N(t). \tag{2.17}$$

Moreover, if the sequence $\mathcal{F}_{\Delta_N}, N \ge 1$ increases to $\mathcal{F}_\Delta$, then

$$\hat{X}_{t,N} \to \mathbb{E}\left[X_t | \mathcal{F}_\Delta\right] \tag{2.18}$$

in $L^2(\Omega)$.

It is worth emphasizing that Theorem 2.4 provides a ready-to-use practical approach for prediction based on the covariances $\mathbf{R}_N$ and $\mathbf{b}_N$ only. Moreover, the assumption that $\mathcal{F}_{\Delta_N}$ is increasing is also natural, as this includes, e.g., the case where $\Delta_N \subset \Delta_{N+1}$ and, at each step, the new sample point is chosen uniformly from Δ and $N \to \infty$. Note that, in general, it is a challenging task to find the correct solution to the integral equations in Theorem 2.1 or Theorem 2.3. The convergence provided in Theorem 2.4 avoids this difficulty.

3. Partially Observed Functional Data. In this section, we break from the Gaussian case and consider a general family of processes $(Y_t)_{t\in[0,T]}$ such that the marginal distribution of Y_t may be arbitrary and such that $(Y_t)_{t\in[0,T]}$ display any (approximation of a) covariance structure. This can be achieved by setting $Y_t = f(t, X_t)$, for X a Gaussian process satisfying Assumption 2.1. Discussion on these processes can be found in Viitasaari and Ilmonen (2020).

The following result provides an explicit expression of the optimal $L^2(\Omega)$-reconstruction in this general setting.

THEOREM 3.1. *Let $Y_t = f(t, X_t)$, where X is a Gaussian process satisfying Assumption 2.1 and the mapping $x \mapsto f(t, x)$ is bijective for all $t \in [0, T]$. Then for any $t \in [0, T] \setminus \Delta$ we have*

$$\mathbb{E}\left[Y_t|\mathcal{F}_\Delta^Y\right] = \int_{-\infty}^{\infty} f\big(t, \hat{X}_t + z\sqrt{\rho(t, t|\Delta)}\big)\phi(z)dz, \tag{3.1}$$

where $\mathcal{F}_\Delta^Y = \sigma(Y_t : t \in \Delta)$, $\hat{X}_t$ is provided in Theorem 2.1, $\rho(t, t|\Delta)$ is given by (2.8) and ϕ denotes the standard normal density.

Theorem 3.1 puts in perspective the role played by f in the reconstruction of the missing parts of Y. Should f be known, then one could simply recover $X_t = (f(t, \cdot))^{-1}(Y_t)$ and apply Theorem 2.1. Therefore, in this situation, it is not required to sample several processes to be able to reconstruct a single one. This reconstruction can be conducted on a curve-by-curve basis as in Section 2.

On the other hand, if f is not known, as is arguably the case in most instances, then one should proceed with its estimation. The rest of this section considers the general situation of a sample of functional observations with missing intervals. It presents an estimating procedure for $\mathbb{E}\left[Y_t|\mathcal{F}_\Delta^Y\right]$ based on a preliminary estimation of f.

REMARK 3.1. Throughout, it is enough to assume that f is injective since only its inverse is of interest. Without loss of generality, we can further assume that it is surjective on its image set.

Let $(Y_t^j)_{t\in\Delta_j}, j = 1, 2, \ldots, J$, be J independent observations where, for each j, the observed part Δ_j is of the form (2.5). Heuristically, if we have sufficiently many observations Y_t^j for each $t \in [0, T]$, then we can estimate the function $f(t, X_t)$ and, by invertibility of f, recover the *proxy* values for X_t^j. From these values, we can then estimate the covariance of the underlying process X. After that, we can estimate the prediction $\hat{X}_t$ and $\rho(t, t|\Delta)$, from which we get the prediction $\hat{Y}_t$ by replacing f, $\hat{X}_t$, and $\rho(t, t|\Delta)$ by their estimated values in (3.1). In order to make this heuristic argument precise, we start by posing the following assumptions. For our purposes, we state the following assumption on f.

ASSUMPTION 3.1. The process $(Y)_{t\in[0,T]}$ is of the form

$$Y_t = f(t, X_t), \tag{3.2}$$

where, for each $t \in [0,T]$, the transformation $z \mapsto f(t,z)$ is continuous and bijective.

Note that, the continuity of f in Assumption 3.1 is equivalent to assuming that the distribution of Y_t has full support, provided that the distribution of X_t is continuous and has full support. Respectively, the bijectivity of f amounts to assuming that the distribution of Y_t is continuous under the same hypothesis on X_t. Note that the latter trivially holds in the context of the Gaussian processes X used in this section. Remark also that no particular behavior in t is assumed. In particular, the classical functional model $Y_t = h(t) + X_t$ (where each observation is a noisy version of a common mean h) always satisfies Assumption 3.1, regardless whether h is continuous or not. Finally, note that, under Assumption 3.1, the continuity of f allows to approximate (3.1) using standard Riemann-Stieltjes integration.

In general, the model $Y_t = f(t, X_t)$ is not identifiable. Indeed, for any $c_t > 0$, $t \in [0,T]$ it holds,

$$g(t, Z_t) = f(t, X_t),$$

where $g(t,z) = f(t, c_t z)$ and $Z_t = X_t/c_t$ is a Gaussian process. To ensure identifiability, one then needs to impose standardization constraints on X_t. We impose the following assumption.

ASSUMPTION 3.2. The variance function $t \mapsto \mathbb{E}X_t^2$ of X is known.

For example, imposing $\mathrm{Var}(X_t) = 1$ for all $t \in [0,T]$ amounts to choosing $c_t = \sqrt{\mathbb{E}[X_t]}$ above. We also note that any other similar condition ensuring the identifiability of the model would suffice. As another example, we could impose the restriction $\mathrm{Var}(Y_t) = \mathbb{E}[X_t^2]$. In this situation as the variance of Y_t can be estimated from the observations, we could recover the variance function of X as well. Under this assumption, the transformation f can be expressed in a way suitable for future estimation. Since $x \mapsto f(t,x)$ is bijective, the proof follows easily and is hence omitted.

Note that, if, for some t, $\mathbb{E}[X_t^2] = 0$, then $\mathrm{Var}(Y_t) = 0$, so that $Y_t = c$ with probability one. One can then trivially predict $\mathbb{E}\left[Y_t | \mathcal{F}_\Delta^Y\right] = c$. In the following, we only consider predictions for values of t such that $\mathbb{E}[X_t^2] > 0$.

PROPOSITION 3.1. *Let $t \in [0,T]$ be fixed and let F_t and Q_t denote the cumulative distribution function and quantile function of the non constant random variable Y_t, respectively. Then, under Assumption 3.1, we have*

$$f(t,z) = Q_t\left(\Phi_t(z)\right)$$

and

$$z = \Phi_t^{-1}\left(F_t(z)\right),$$

where Φ_t *is the cumulative distribution function of* $N(0, \mathbb{E}X_t^2)$.

We are now ready to introduce our estimation procedure for the predictor $\mathbb{E}\big[Y_t^j | \mathcal{F}_{\Delta_j}^Y\big]$, $t \in [0,T] \setminus \Delta_j$. We also assume that we are given the discrete net $\pi_N \subset [0,T]$ of evaluation points and the corresponding observed values $Y_{t_j}, t_j \in \Delta_j \cap \pi_N$. For $t \in \pi_N$, let $K(t) = \{j : t \in \Delta_j \cap \pi_N\}$ denote the indices of observations Y_j defined at t. Similarly, for $(t,s) \in \pi_N^2$, let $J(t,s) = \{j : t,s \in \Delta_j \cap \pi_N\}$ denote the indices of observations Y_j defined at both t and $s \in \Delta_j$. With a slight abuse of notation, we omit the dependence in t or (t,s) and simply use K and J for all $t,s \in \pi_N$. The estimation is done by using the following steps.

(i) **Estimation of f and construction of a proxy of the process X**:

For $t \in \pi_N$, define the empirical distribution function based on the observations $Y_t^k, k \in K$ by

$$\hat{F}_{t,K}(z) = \frac{1}{|K|+1} \sum_{k \in K} \mathbf{1}_{Y_t^k \le z}.$$

Note that here the summation is over all $|K| := |K(t)|$ observed values Y_t^k, and we divide with $|K|+1$ instead of the usual $|K|$ to ensure that $\hat{F}_{t,K}(Y_t^k) \in (0,1)$ for any observed value Y_t^k. Let $\hat{Q}_{t,K}$ denote the empirical quantile function of Y_t and define the estimator $\hat{f}$ of f by setting, for each $t \in \pi_N$ and $z \in \mathbb{R}$,

$$\hat{f}_K(t,z) = \hat{Q}_{t,K}\left(\Phi_t(z)\right). \tag{3.3}$$

The proxy $X_{t,K}^{\mathrm{prox}}$ for the value X_t is defined by

$$X_{t,K}^{\mathrm{prox}} = \Phi_t^{-1}\left(\hat{F}_{t,K}(Y_t)\right). \tag{3.4}$$

Since $\hat{F}_{t,K}(Y_t) \in (0,1)$, $X_{t,K}^{\mathrm{prox}} \in (-\infty, \infty)$ is well-defined.

(ii) **Estimation of the covariance**:

For $t,s \in \pi_N$, we estimate the covariances $\hat{R}_J(t,s)$ by

$$\hat{R}_J(t,s) = \frac{1}{|J|} \sum_{k \in J} X_{t,J}^{k,\mathrm{prox}} X_{s,J}^{k,\mathrm{prox}}, \tag{3.5}$$

where $X_{t,J}^{k,\mathrm{prox}}$ denotes the proxy of the value X_t^k associated to observations Y_t^k. For $\Delta \cap \pi_N \subset \pi_N$, let $\hat{\mathbf{R}}_{N,\Delta}$ be the matrix collecting the covariance estimators $\hat{R}_J(t,s)$ for $t,s \in \Delta \cap \pi_N$ and let $\hat{\mathbf{R}}_N := \hat{\mathbf{R}}_{N,\pi_N}$. Similarly, for a fixed $t \in \pi_N$ and $\Delta \cap \pi_N \subset \pi_N$, we denote by $\hat{\mathbf{b}}_{N,\Delta}(t)$ the vector consisting of covariance estimates $\hat{R}_J(t,s), s \in \Delta \cap \pi_N$.

(iii) **Estimation of the predictors $\hat{X}_t^j$**:

For a given j and $t \in \pi_N \setminus \Delta_j$, approximate (2.16) by

$$\hat{X}_{t,N}^j = (a_j(t))' X_{\Delta_j,K}^{\text{prox}}, \tag{3.6}$$

where

$$a_j(t) = \hat{\mathbf{R}}_{N,\Delta_j}^{-1} \hat{\mathbf{b}}_{N,\Delta_j}(t)$$

and $X_{\Delta_j,K}^{\text{prox}}$ is the vector containing the $X_{s,K}^{\text{prox}}$ for $s \in \Delta_j \cap \pi_N$ in the same order as in $\hat{\mathbf{b}}_{N,\Delta_j}(t)$. Note that here we have assumed explicitly that $\hat{\mathbf{R}}_{N,\Delta_j}$ is of full-rank. Under assumptions of Theorem 2.4 this holds provided that, for every $(t,s) \in \pi_N^2$, $|J(t,s)|$ is large enough, ensuring $\hat{\mathbf{R}}_N$ is close to $\mathbf{R}_N$ (see Corollary B.3). Note that, in practice, if $\hat{\mathbf{R}}_{N,\Delta_j}$ is not of full-rank, then one could, for example, replace the inverses with general Moore-Penrose inverses.

(iv) **Estimation of $\rho(t,t|\Delta \cap \pi_N)$**:

For $\Delta \cap \pi_N \subset \pi_N$, we define a centered Gaussian vector Z_N with covariance $\hat{\mathbf{R}}_N$. By simulating M vectors $Z_N^m, m = 1, 2, \dots M$, we can define the estimator $\hat{\rho}_M(t,t|\Delta \cap \pi_N)$ by

$$\hat{\rho}_M(t,t|\Delta \cap \pi_N) = \frac{1}{M} \sum_{m=1}^{M} \left(Z_{t,N}^m - \hat{Z}_{t,N}^m \right)^2, \tag{3.7}$$

where $\hat{Z}_{t,N}^m$ is the prediction of $Z_{t,N}^m$ computed through (2.16).

(v) **Approximation of the integral**:

The final step of our construction requires estimating the integral in (3.1). For any continuous function $h \in L^1(\mathbb{R}, \phi(z)dz)$, let $L > 0$ and approximate the integral

$$I_\phi(h) := \int_{-\infty}^{\infty} h(z)\phi(z)dz \qquad \text{with} \qquad I_{\phi,L}(h) = \int_{-L}^{L} h(z)\phi(z)dz.$$

For practical or simulation purposes this can be further approximated by Riemann sums

$$I_{\phi,L}(h) = \sum_{z_k} h(z_k)\phi(z_k)|z_k - z_{k-1}|,$$

where the sequence z_k is a partition of $[-L, L]$ such that $\max_k |z_k - z_{k-1}| \to 0$ as $L \to \infty$.

We are now ready to define the predictor for Y_t^j.

DEFINITION 3.1. Let $t \in \pi_N \setminus \Delta_j$. For given L and M, define

$$\hat{Y}_{t,L,M,K,N}^j = I_{\phi,L}\left(\hat{h}_M(t,\cdot) \right), \tag{3.8}$$

where

$$\hat{h}_M(t,z) = \hat{f}_K\left(t, \hat{X}^j_{t,N} + z\sqrt{\hat{\rho}_M(t,t|\Delta_j \cap \pi_N)}\right),$$

and $\hat{f}_K$, $\hat{X}^j_{t,N}$, and $\hat{\rho}_M$ are given by (3.3), (3.6), and (3.7), respectively.

The following result shows that the prediction can, under our model assumptions, be done using the observed data only.

THEOREM 3.2. *Suppose that the non constant random variable Y_t satisfies Assumptions (3.1) and (3.2) and let $\hat{Y}_{t,L,M,K,N}$ be given by (3.8). Let $J^* = \min_{(t,s)\in\pi_N} |J(t,s)|$. Assume $\mathcal{F}_{\Delta\cap\pi_N}$ increases to $\mathcal{F}_\Delta$ as $N \to \infty$. Then,*

$$\lim_{L\to\infty}\lim_{N\to\infty}\lim_{J^*\to\infty}\lim_{M\to\infty} \hat{Y}_{t,L,M,K,N} = \mathbb{E}\left[Y_t|\mathcal{F}^Y_\Delta\right]$$

in probability.

REMARK 3.2. In the result above, note that $J^* \leq |K|$. The assumption $J^* \to \infty$ is required to ensure that $\hat{\mathbf{R}}_N \to \mathbf{R}_N$.

REMARK 3.3. Typically, in the functional data analysis approach, the object of study is $Y_t = h(t) + \varepsilon_t$ for $t \in \Delta$ and ε_t independent errors. The first standard aim, in order to apply functional techniques, is to reconstruct $h(t)$ from the discrete observations. The model $Y_t = f(t, X_t)$ studied in this paper allows such reconstruction. Indeed, one can consider $\varepsilon_t \sim X_t = W_t - W_{t-\delta}$ for some δ and with W_t a Brownian motion.

REMARK 3.4. Note that, the discrete approximation described above is very general. If one further assumes that the processes are of a parametric nature (as is the case in the Simulations section below), then step (ii) can be simplified. Indeed, one can then estimate the relevant parameters and estimate the associated covariance structure by plugging-in the parameter estimates.

4. Simulation Study. In this section, we illustrate the theoretical results from Sections 2 and 3. For the sake of simplicity, we assume throughout that $T = 1$ and that the missing part of the processes is of length 0.2. That is $\Delta = [0, t_U] \cup [t_L, 1]$, with $t_L - t_U = 0.2$. The general setting considered here further assumes that the process X_t is discretized at equidistant locations $t_i = \frac{i}{N}$, $i = 1, \ldots, N$ of $[0, 1]$. We measure the loss incurred from estimating X_N on $\bar{\Delta} = [0, 1] \setminus \Delta$ with $\hat{X}_N$ via

$$L(X_N, \hat{X}_N) = \frac{1}{0.2n} \sum_{t_i \in \bar{\Delta}} \left(X_{t_i} - \hat{X}_{t_i}\right)^2,$$

that is, the discretized approximation of the L^2 loss on $\bar{\Delta}$ based on the full processes X and $\hat{X}$, $L(X, \hat{X}) = \int_{\bar{\Delta}} (X_t - \bar{X}_t)^2 dt$. Different simulation settings are considered.

Simulation setting 1: fractional Brownian motions (fBm): This setting assumes that $X_t = B_t^H$ is a fractional Brownian motion, that is, a centered Gaussian process with covariance

$$R(t,s) = \frac{1}{2}\left[t^{2H} + s^{2H} - |t-s|^{2H}\right].$$

The Hurst parameter $H \in (0,1)$ gives the Hölder and self-similarity index of the process, allowing varying roughness and memory. The case $H = 1/2$ corresponds to (classical) Brownian motions. It is known that fractional Brownian motions satisfy Assumption 2.1 for any $H \in (0,1)$, and their kernels $K_H(t,s)$ are completely known (see, e.g., Biagini et al. (2008) and references therein).

Simulation settings 2 and 3: bifractional Brownian motions (bBm): A bifractional Brownian motion is a centered Gaussian process with covariance

$$R(t,s) = \frac{1}{2^K}\left[(t^{2H} + s^{2H})^K - |t-s|^{2HK}\right].$$

The parameters are such that $H \in (0,1)$, $K \in (0,2)$ and $HK \in (0,1)$. Note that, for $K = 1$, one recovers a fBm. These Gaussian processes were introduced to model situations where small increments tend to be stationary, but large increments do not. That is, only the small scale behavior tends to be close to a fBm. Moreover, the case $HK = \frac{1}{2}$ provides an interesting model in mathematical finance, as then one obtains a process that has non-trivial quadratic variation similar to the standard Brownian motion, while possessing memory. For details on bifractional Brownian motion, see the monograph Tudor (2013). It is known that bifractional Brownian motions are purely non-deterministic and HK-self-similar. Consequently, Assumption 2.1 is satisfied despite the fact that the kernel is not explicitly known. In the simulations, values $K = 2/3$ (setting 2) and $K = 3/4$ (setting 3) are considered.

Simulation setting 4: independent mixed fractional Brownian motions (imfBm): A mixed fractional Browian motion consists of a sum of a standard Brownian motion and a fractional Brownian motion, i.e.

$$X_t = W_t + B_t^H.$$

In this setting, we assume that the processes W and B^H are independent, a case in which the covariance is known explicitly and is given by

$$R(t,s) = \min(s,t) + \frac{1}{2}\left[t^{2H} + s^{2H} - |t-s|^{2H}\right].$$

Especially with $H > \frac{1}{2}$, the model is widely used in mathematical finance. It is known that independent mixed fractional Brownian motions satisfy Assumption 2.1. See Mishura (2008).

Figure 1 displays, for each simulation setting and each $H \in \{0.2, 0.5, 0.8\}$ a random observation X_t together with the optimal prediction $\hat{X}_{t,N}$ in the interval $\bar{\Delta} = [0.4, 0.6]$. The reconstruction was conducted according to Theorem 2.4. In particular, for illustration

purposes, we assume that the covariance R_N is known and is used in the reconstruction of $X_{t,N}$ on $\bar{\Delta}$.

As evidenced by Figure 1, the curves become smoother for larger self-similarity indices. Furthermore, when the self-similarity index is 0.5 (that is, in cases 1 and 4, for $H = 0.5$), the optimal reconstruction yields straight lines, as displayed. The other cases show the impact of the memory present in the processes on the optimal prediction. Note again that, in the context of Theorem 2.4, $\hat{X}_{t,N}$ is based on the observed parts of X only and that the reconstruction can be conducted independently for each curve.

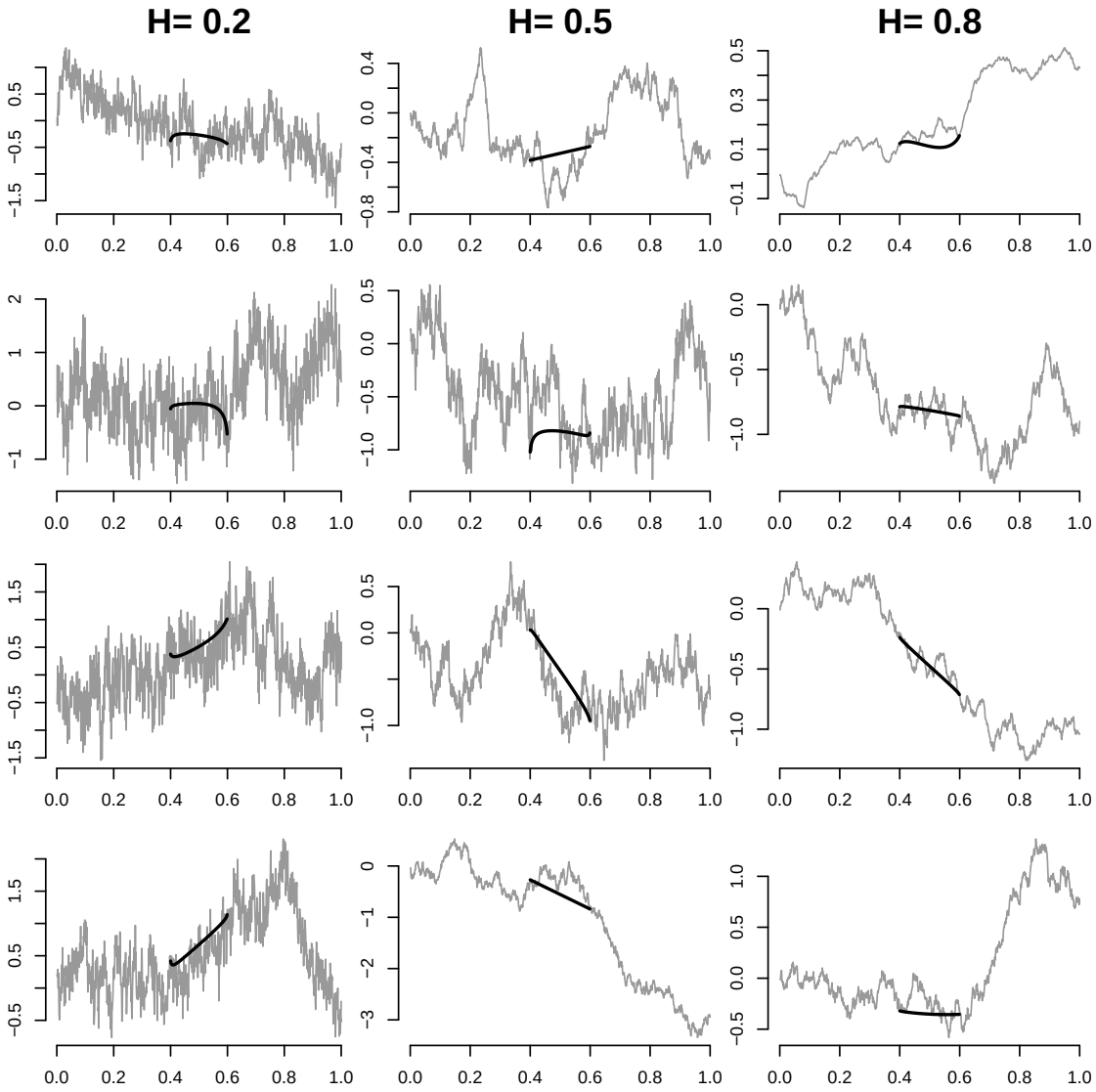

FIG 1. *Gaussian processes (in grey) generated according to each simulation setting (row 1: fBm, row 2-3 : bBm, for $K = 2/3$ and $K = 3/4$, row 4: imfBm) for Hurst indices $H = 0.2$ (left), $H = 0.5$ (middle) and $H = 0.8$ (right). The optimal reconstructed curves on $\bar{\Delta} = [0.4, 0.6]$ are displayed in black.*

In order to assess the quality of the reconstruction, the following experiment was con-

ducted. In each simulation setting, $n = 500$ observations $X_1, \dots, X_n$ were generated and observed on the regular grid π_N for $N = 250$. For each $j = 1, \dots, 500$, $Y_j = f(X_j)$ was assumed to be observed on $\Delta_j = [0,1] \setminus [t_{U,j}, t_{L,j}]$, with $t_{U,j}$ chosen uniformly within the interval $[0.1, 0.7]$ and $t_{L,j} = t_{U,j} + 0.2$. This resulted in uniformly missing intervals of length 0.2 within $[0.1, 0.9]$. The functions used in this simulation setting were $f(X_j) = X_j$ and $f(X_j) = \exp(X_j)$ (referred henceforth as the identity and exponential case, respectively).

We compared our approach to several classical curve reconstruction techniques. More precisely, the competitors considered were:

- spline smoothing (using 9 basis splines of order 2);
- Fourier smoothing (using 13 basis elements); and
- linear interpolation.

Note that, for spline and Fourier smoothing, the number of basis elements were based on cross-validation. For our approach, we consider three scenarios. In the first two, the parametric family of the Gaussian processes are known. The first scenario assumes that both f and H are fixed, so that, we can use Theorem 2.4 as a baseline. In the second, we assume that f is known. The Hurst index H is estimated as in Coeurjolly (2000) based on the observed data points and used in the estimation of the covariance matrix in step (ii). Finally, the last scenario uses the algorithm detailed in Section 3, without assuming neither knowledge on f, H nor the process distribution.

The following figures display boxplots of the losses $L(Y_{j,N}, \hat{Y}_{j,N})$ for each simulation setting and for each value of $H \in \{0.2, 0.5, 0.8\}$. The cases $f(X) = X$ and $f(X) = \exp(X)$ are displayed in Figures 2 and 3, respectively. The predictions are displayed as follows: spline smoothing (S9) in red, Fourier smoothing (F13) in green, linear interpolation (I) in yellow and our proposals in different hues of blue. The case where the Gaussian process is fully known or when H is estimated are denoted by (G) and (GH) respectively. The cases where the underlying processes are unknown and where f is estimated are denoted by (IdH) and (ExpH) (in the identity and exponential case, respectively).

Simulation results clearly highlight the optimality of the proposed reconstruction. Remarkably, the cost of not knowing H seems to remain negligible. In the identity case, the estimation of f does not change the precision much. The case $f = \exp$ seems to be slightly more unstable, as the function has to be estimated pointwise. Note that, the algorithm still provides competitive estimation of the missing parts. Naturally, if the type of the underlying parametric process is known, it is advisable to avoid estimating the full function f and restrict to the estimation of the model parameters. In our examples, larger values of the Hurst index H, corresponding to processes with stronger memory, yield more accurate reconstruction for all techniques. In the cases where the optimal reconstruction is the linear interpolation (simulation settings 1 and 4, for $H = 0.5$ and $f = \mathrm{id}$), our approach and the linear interpolation indeed coincide. In all other cases, there is a clear gain in using the optimal approach. The spline and Fourier smoothing do poorly in general,

and get worse for lower H values. This is explained as these interpolations are known to accurately approximate functions only in their observed parts and lower H values provide rougher curves.

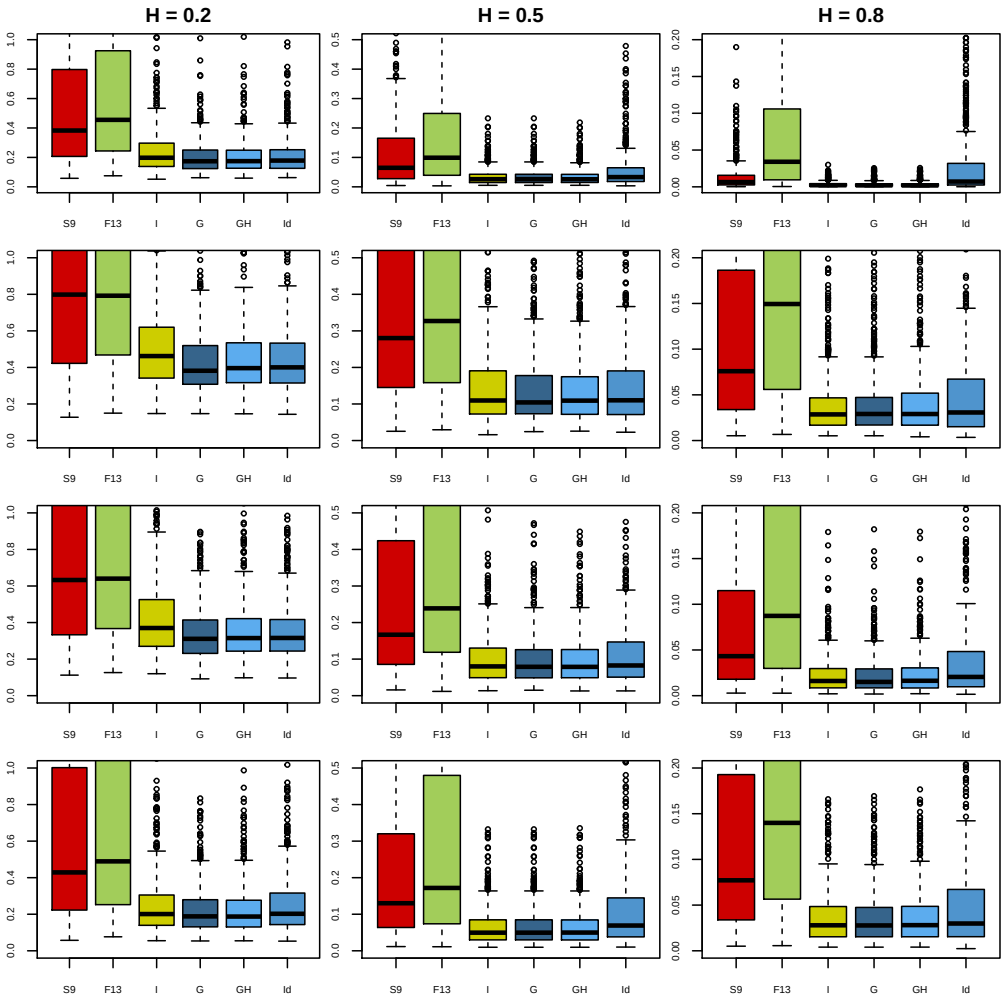

FIG 2. *Boxplots of the L^2 loss, for $f = id$, for the four simulations settings described above (one per row), for $H = 0.2$, 0.5 or 0.8 (per column). The reconstruction methods are, from left to right: (red) spline smoothing using 9 basis elements of order 2 (S9); (green) Fourier smoothing using 13 basis elements (F13); (yellow) the reconstruction obtained by linearly interpolating X_{t_U} to X_{t_L}; (blue) our proposal assuming the process is Gaussian with the right H value (left, G), with an estimated H (middle, GH) or with an estimated f and H (right, IdH).*

APPENDIX A: PRELIMINARIES ON GAUSSIAN ANALYSIS

In this section we recall some preliminaries on Gaussian analysis and processes. For details on the topic, we refer to Janson (1997); Nualart (2006); and Sottinen and Viitasaari (2016).

Suppose that X is a centered and separable Gaussian process on some compact interval

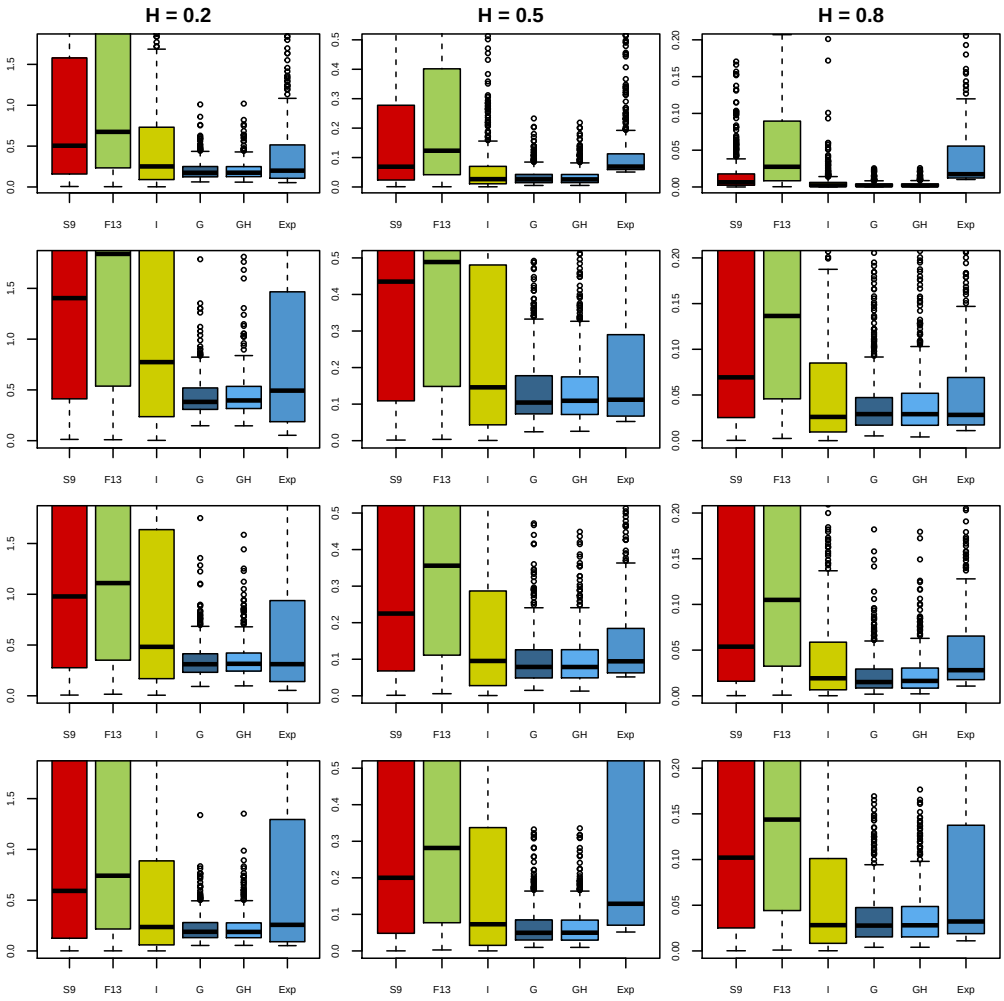

FIG 3. *Boxplots of the L^2 loss, for $f = \exp$, for the four simulations settings described above (one per row), for $H = 0.2$, 0.5 or 0.8 (per column). The reconstruction methods are, from left to right: (red) spline smoothing using 9 basis elements of order 2 (S9); (green) Fourier smoothing using 13 basis elements (F13); (yellow) the reconstruction obtained by linearly interpolating X_{t_U} to X_{t_L}; (blue) our proposal assuming the process is Gaussian with the right H value (left, G), with an estimated H (middle, GH) or with an estimated f and H (right, IdH).*

$[0, T]$. It is known (Sottinen and Viitasaari, 2016) that if the covariance R of X is of trace class, i.e. $R(t, t) = Var(X_t) \in L^2([0, T])$, then X admits a Fredholm representation (2.3)

$$X_t = \int_0^T K(t, s) dW_s,$$

where W is a standard Brownian motion and $K \in L^2([0, T]^2)$. In particular, this is the case whenever X is continuous.

DEFINITION A.1 (Operator associated with a kernel). The associated operator to a

kernel $\Gamma \in L^2([0,T]^2)$, also denoted by Γ, is defined as $\Gamma : L^2([0,T]) \to L^2([0,T])$ with

$$(\Gamma f)(t) = \int_0^T f(s)\Gamma(t,s)\,ds. \tag{A.1}$$

DEFINITION A.2 (Isonormal process). The isonormal process associated with X, also denoted by X, is the Gaussian family $\{X(h), h \in \mathcal{H}\}$, where the Hilbert space $\mathcal{H} = \mathcal{H}(R)$ is the closed span of $\{\mathbf{1}_t := \mathbf{1}_{[0,t)},\ t \le T\}$, closed for the inner product on $\mathcal{H}$ defined via $\langle \mathbf{1}_t, \mathbf{1}_s \rangle_{\mathcal{H}} := R(t,s)$.

By Definition A.2, $X(h)$ can be viewed as the image of $h \in \mathcal{H}$ in the isometry that extends the relation

$$X(\mathbf{1}_t) := X_t$$

linearly. This allows us to define the Wiener integral with respect to X.

DEFINITION A.3 (Wiener integral). $X(h)$ is the *Wiener integral* of the element $h \in \mathcal{H}$ with respect to X, which is denoted by

$$\int_0^1 h(t)\,dX_t := X(h).$$

REMARK A.1. Note that, it may happen that the space $\mathcal{H}$ is not a space of functions, but instead it has to be considered as a space of generalised functions, see Pipiras and Taqqu (2001).

DEFINITION A.4. The adjoint associated operator Γ^* of a kernel $\Gamma \in L^2([0,T]^2)$ is defined by linearly extending the relation

$$\Gamma^* \mathbf{1}_t = \Gamma(t,\cdot).$$

EXAMPLE A.1. Let h be a step function of the form

$$h(s) = \sum_{k=1}^n \alpha_k \mathbf{1}_{(t_{k-1},t_k]},$$

where $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $0 \le t_0 < t_1 < \ldots t_n \le T$. Then

$$\Gamma^* h(s) = \int_0^T h(t)\Gamma(dt,s).$$

Definition A.4 allows us to provide transfer principle for Wiener integrals. For details, see Sottinen and Viitasaari (2016).

THEOREM A.1 (Transfer principle for Wiener integrals). *Let X be a separable centered Gaussian process with representation (2.3) and let $h \in \mathcal{H}$. Then,*

$$\int_0^T h(t)\,dX_t = \int_0^T K^* h(t)\,dW_t.$$

Finally, we recall two important spaces related to a Gaussian process X.

DEFINITION A.5. The first chaos of X, denoted by $H^X = H_T^X$, is defined as the Hilbert space spanned by $X_s, s \in [0,T]$ and closed for the $L^2(\Omega)$-distance. Similarly, H_t^X denotes the Hilbert space spanned by $X_s, s \in [0,t]$.

DEFINITION A.6 (Cameron-Martin space). The Cameron-Martin space of X, denoted by CM_X, is the space of functions defined by

$$f(t) = \mathbb{E}[X_t \xi], \quad \xi \in H^X.$$

The Cameron-Martin space is a Hilbert space when equipped with the inner product

$$\langle f_1, f_2 \rangle_{CM_X} = \mathbb{E}[\xi_1 \xi_2],$$

where $f_i(\cdot) = \mathbb{E}[X.\xi_i], i = 1,2$. Moreover, all the spaces $\mathcal{H}$, H^X, and CM_X are isometric and separable whenever X is separable. More specifically, the isometry between $\mathcal{H}$ and H^X is given by the relation

$$\int_0^T h(t) dX_t \in H^X, \quad h \in \mathcal{H}.$$

Note that the Cameron-Martin space is also known as a Reproducing Kernel Hilbert Space (RKHS). In this instance, the covariance is the reproducing kernel, i.e., is defined via $f(t) = \langle f, R(t,.) \rangle_{RKHS}$.

APPENDIX B: PROOFS

B.1. Proofs related to Section 2. For a linear operator $A : X \mapsto Y$, where X and Y are some vector spaces, the *null space* and *range* are defined by

$$\mathcal{N}(A) = \{x \in X : Ax = 0\} \subset X \quad \text{and} \quad \mathcal{R}(A) = \{y \in Y : Ax = y \text{ for some } x \in X\} \subset Y.$$

If $\mathcal{N}(A) = \{0\}$, then the inverse operator $A^{-1} : \mathcal{R}(A) \mapsto X$ exists on $\mathcal{R}(A)$. In particular, if $\mathcal{R}(A) = Y$, then the inverse A^{-1} exists on the entire domain Y. The following non-trivial result, originally due to Banach, is applied to prove Theorem B.2.

THEOREM B.1. *Let $A : X \mapsto Y$ be a bounded linear operator such that $\mathcal{N}(A) = \{0\}$ and $\mathcal{R}(A) = Y$. Then the inverse operator A^{-1} is linear and bounded as well.*

COROLLARY B.1. *Let $A : X \mapsto Y$ be a bounded linear operator such that $\mathcal{N}(A) = \{0\}$ and $\mathcal{R}(A) = Y$. Then for any $y \in Y$ the equation*

$$Ax = y$$

has a unique solution $x \in X$.

In order to prove Theorem 2.1, we need the following preliminary results.

THEOREM B.2. *Let X be given by (2.3) and suppose further that $H_T^X = H_T^W$. Then, the associated operator $K : L^2([0,T]) \mapsto L^2([0,T])$ satisfies $\mathcal{N}(K) = \{0\}$ and $\mathcal{R}(K) = CM_X \subset L^2([0,T])$. In particular, $K : L^2([0,T]) \mapsto CM_X$ is a bounded linear operator that has a bounded and linear inverse $K^{-1} : CM_X \mapsto L^2([0,T])$.*

PROOF. The fact that $H_T^X = H_T^W$ ensures that the operator $K^* : \mathcal{H} \mapsto L^2([0,T])$ is bijective. Indeed, each $\xi \in H_T^X$ satisfies

$$\xi = \int_0^T f(s)dX_s = \int_0^T (K^*f)(s)dW_s = \int_0^T g(s)dW_s,$$

for some $g \in L^2([0,T])$. Conversely, each element $\phi \in H_T^W$, with representation

$$\phi = \int_0^T h(s)dW_s, \quad h \in L^2([0,T]),$$

belongs to H_T^X as well. Thus for each $f \in L^2([0,T])$ there exists $\xi \in H_T^X$ such that

$$(Kf)(t) = \int_0^T K(t,s)f(s)ds = \mathbb{E}\left[X_t\xi\right].$$

This gives directly that $\mathcal{R}(K) = CM_X$, and that if $Kf \equiv 0$, then

$$\xi = \int_0^T f(s)dW_s = 0$$

almost surely, implying $f \equiv 0$. Thus $\mathcal{N}(K) = \{0\}$. Moreover, it is clear from the isometry

$$\|Kf\|_{CM_X} = \|f\|_{L^2([0,T])}$$

that $K : L^2([0,T]) \mapsto CM_X$ is a bounded linear operator. Theorem B.1 concludes the proof. □

REMARK B.1. Since K^* is invertible, we can transform our prediction into a Wiener integral with respect to X by using

$$\hat{X}_t = \int_0^T f_t(s)dW_s = \int_0^T \left((K^*)^{-1} f_t\right)(s)dX_s. \tag{B.1}$$

However, even in some well-studied cases such as the fractional Brownian motion, the expression (B.1) is impractical as the operators K^* and $(K^*)^{-1}$ might be complicated.

COROLLARY B.2. *Let X be given by (2.3) and suppose further that $H_T^X = H_T^W$. Then, for any $g \in CM_X$, the integral equation of the first kind*

$$(Kf)(t) = \int_0^T K(t,s)f(s)ds = g(t)$$

has a unique solution $f \in L^2([0,T])$.

PROOF. The claim follows directly from Theorem B.2 and Corollary B.1. □

We are now in position to prove our first main theorem.

PROOF OF THEOREM 2.1. Let $t \in [0,T] \setminus \Delta$. It is well-known that $\hat{X}_t$ is Gaussian. Moreover, since $\hat{X}_t \in H_T^X$, the assumption $H_T^X = H_T^W$ implies that the representation

$$\hat{X}_t = \int_0^T f_t(s) dW_s$$

holds. By definition of the conditional expectation, which is an orthogonal projection, $\hat{X}_t - X_t$ is orthogonal to all random variables of the form

$$Y = \sum_{k=1}^n \alpha_k X_{t_L^k} + \sum_{k=1}^n \beta_k (X_{u_k} - X_{t_L^k}), \tag{B.2}$$

where $u_k \in [t_L^k, t_U^k]$ for all $k = 1, \dots, n$. By linearity, it follows that $\hat{X}_t - X_t$ is orthogonal to all Gaussian random variables measurable with respect to $\mathcal{F}_\Delta$, which in turn implies, by Gaussianity, that $\hat{X}_t - X_t$ is orthogonal to $L^2(\Omega, \mathbb{P}, \mathcal{F}_\Delta)$.

Using (2.3) in (B.2) and the fact that $\hat{X}_t - X_t = \int_0^T f_t(s) - K(t,s) dW_s$, the orthogonality condition gives

$$0 = \int_0^T [f_t(s) - K(t,s)] \sum_{k=1}^n \left[\alpha_k K(t_L^k, s) + \beta_k \left(K(u_k, s) - K(t_L^k, s) \right) \right] ds.$$

As this holds for all $\alpha_k, \beta_k \in \mathbb{R}$, we obtain that, for all $k = 1, \dots, n$ and $t_L^k \le u_k \le t_U^k$,

$$\int_0^T f_t(s) K(u_k, s) ds = \int_0^T K(t,s) K(u_k, s) ds.$$

This shows that (2.7) holds.

For any g_ξ such that $\int_0^T g_\xi(s) K(u_k, s) ds = 0$, the same reasoning shows that $\int_0^T g_\xi(s) dW_s$ is orthogonal to $\mathcal{F}_\Delta$. The extra condition $\int_0^T g_\xi(s) f_t(s) ds = 0$ ensures that $\int_0^T f_t(s) dW_s$ is measurable with respect to $\mathcal{F}_\Delta$. The uniqueness of f_t then follows from the uniqueness of the conditional expectation $\hat{X}_t = \int_0^T f_t(s) dW_s$. This completes the proof. □

PROOF OF THEOREM 2.2. It is known that $X|\mathcal{F}_\Delta$ is Gaussian with Gaussian random mean and deterministic covariance given by

$$\hat{X}_t = \mathbb{E}[X_t | \mathcal{F}_\Delta] \quad \text{and} \quad \rho(t,s|\Delta) = \mathbb{E}\left[(X_t - \hat{X}_t)(X_s - \hat{X}_s) \right],$$

respectively. See, e.g. (Janson, 1997, Chapter 9). The mean is provided in Theorem 2.1. Similarly, the representation (2.8) for ρ follows from

$$X_t - \hat{X}_t = \int_0^T K(t,u) - f_t(u) dW_u$$

and the Itô isometry

$$\mathrm{Cov}\left(\int_0^T f_1(s)dW_s, \int_0^T f_2(s)dW_s\right) = \int_0^T f_1(s)f_2(s)ds.$$

□

PROOF OF THEOREM 2.3. By Theorem 2.1, for any $1 \leq k \leq n$ and $t_L^k \leq u_k \leq t_U^k$, it holds that

$$\int_0^T f_t(s)K(u_k, s)ds = \int_0^T K(t, s)K(u_k, s)ds.$$

With the convention $t_U^0 = 0$, we have

$$\int_{t_U^{k-1}}^{t_L^k} f_t(s)K(t_L^k, s)ds = \int_0^{t_L^k} K(t, s)K(t_L^k, s)ds - \int_0^{t_U^{k-1}} f_t(s)K(t_L^k, s)ds$$

and

$$\int_{t_L^k}^{u_k} [f_t(s) - K(t, s)]K(u_k, s)ds = \int_0^{t_L^k} [K(t, s) - f_t(s)]K(u_k, s)ds.$$

The result follows. □

PROOF OF COROLLARY 2.1. In the setting of this corollary, equations (2.10) and (2.11) rewrite

$$\int_0^u [f_t(s) - K(t, s)]\, K(u, s)ds = 0, \quad 0 \leq u \leq t_U, \tag{B.3}$$

$$\int_{t_U}^{t_L} f_t(s)K(t_L, s)ds = \int_0^{t_L} K(t, s)K(t_L, s)ds - \int_0^{t_U} f_t(s)K(t_L, s)ds, \tag{B.4}$$

and

$$\int_{t_L}^u [f_t(s) - K(t, s)]\, K(u, s)ds = \int_0^{t_L} [K(t, s) - f_t(s)]\, K(u, s)ds, \quad t_L \leq u \leq T. \tag{B.5}$$

The solution to (B.3) is given by $f_t(s) = K(t, s), s \leq t_U$. It is unique since $H_t^X = H_t^W$ allows to use Corollary B.2. (B.4) therefore rewrites

$$\int_{t_U}^{t_L} f_t(s)K(t_L, s)ds = \int_0^{t_U} K(t, s)K(t_L, s)ds.$$

Let $f_t(s) = \tilde{f}_t(s) + c(t)K(t_L, s)$. It follows from (2.13) that

$$\int_{t_U}^{t_L} \tilde{f}_t(s)K(t_L, s)ds = 0$$

and since $H_t^X = H_t^W$, we must have $\tilde{f}_t(s) = 0$ for all $t_L < s < t_U$. Indeed, otherwise $\int_0^T f_t(s)dW_s \in H_s^W \setminus H_{t_L}^W = H_s^X \setminus H_{t_L}^X$ for some $t_L < s < t_U$, and hence $\hat{X}_t$ would not be measurable. Finally, (2.14) follows by plugging $f_t(s)$ for $s < t_L$ into (B.5). □

PROOF OF COROLLARY 2.2. Theorem 2.1 implies that $\hat{X}_t = \int_0^T f_t(s)dW_s$, where f_t satisfies, for $0 \le u \le t_U$,

$$\int_0^u f_t(s)K(u,s)ds = \int_0^u K(t,s)K(u,s)ds.$$

Since $H_t^X = H_t^W$ for all $t \in [0,T]$, Corollary B.2 implies that the unique solution is given by $f_t(s) = K(t,s)$. □

PROOF OF THEOREM 2.4. Let $t \in [0,T] \setminus \Delta_N$. As in the proof of Theorem 2.3, the random variable $\hat{X}_{t,N}$ is Gaussian and measurable with respect to Δ_N. Thus, it has the representation (2.16) and $\hat{X}_{t,N} - X_t$ is orthogonal to all random variables

$$Y = \sum_{t_k \in \Delta_N} \alpha_k X_{t_k}, \quad \alpha_k \in \mathbb{R}.$$

It follows that

$$0 = \mathbb{E}\left[(\hat{X}_{t,N} - X_t)Y\right] = \sum_{t_k \in \Delta_N} \alpha_k \mathbb{E}\left[(\hat{X}_{t,N} - X_t)X_{t_k}\right]$$

and hence

$$\mathbb{E}\left[(\hat{X}_{t,N} - X_t)X_{t_k}\right] = 0.$$

Now representation (2.16) gives

$$\mathbb{E}\left[\hat{X}_{t,N}X_{t_k}\right] = \sum_{t_j \in \Delta_N} a_j(t)(\mathbf{R}_N)_{jk}$$

which gives us (2.17). In order to prove (2.18), we set

$$M_N = \mathbb{E}\left[X_t | \mathcal{F}_{\Delta_N}\right].$$

Since $\mathcal{F}_{\Delta_N} \subset \mathcal{F}_{\Delta_{N+1}}$, it follows that M_N is a martingale (with respect to $(\mathcal{F}_{\Delta_N})_{N\ge1}$). By Jensen's inequality and the fact that X_t has bounded variance, we get

$$\sup_{N\ge1} \mathbb{E}M_N^2 \le \mathbb{E}X_t^2 < \infty.$$

Hence M_N is uniformly bounded in $L^2(\Omega)$. Thus we may apply the martingale convergence theorem giving us $M_N \to M_\infty$ almost surely and in $L^2(\Omega)$. From this and the fact that $\mathcal{F}_{\Delta_N}$ increases to $\mathcal{F}_\Delta$ it follows that

$$M_\infty = \hat{X}_t = \mathbb{E}\left[X_t | \mathcal{F}_\Delta\right].$$

This concludes the proof. □

B.2. Proofs related to Section 3.

PROOF OF THEOREM 3.1. Since $x \mapsto f(t,x)$ is bijective for all $t \in [0,T]$, it follows that $\mathcal{F}_\Delta^Y = \mathcal{F}_\Delta^X$. The result now follows by writing

$$f(t, X_t) = f(t, \hat{X}_t + X_t - \hat{X}_t)$$

which yields, by the independence of $\hat{X}_t$ and $X_t - \hat{X}_t$ and the fact that $X_t - \hat{X}_t \sim N(0, \rho(t,t|\Delta))$, that

$$\mathbb{E}\left[f(t, X_t)|\mathcal{F}_\Delta\right] = \int_{-\infty}^{\infty} f(t, \hat{X}_t + z)\phi_{\rho(t,t|\Delta)}(z)dz,$$

where $\phi_{\rho(t,t|\Delta)}$ denotes the density function of $N(0, \rho(t,t|\Delta))$. The claim now follows from a change of variable. □

PROPOSITION B.1. *Let $t \in \pi_N$ and let $X_{t,K}^{prox}$ be given by (3.4). Then, for any $p \geq 1$, we have, almost surely and in $L^p(\Omega)$*

$$X_{t,K}^{prox} \to X_t \tag{B.6}$$

as $|K| \to \infty$.

PROOF. For simplicity, we suppress t from the notation and simply write e.g. X_K^{prox} instead of $X_{t,K}^{\text{prox}}$. With this simplified notation,

$$X_K^{\text{prox}} = \Phi^{-1}\left(\hat{F}_K(Y)\right).$$

First we observe that, by the Glivenko-Cantelli theorem and continuity of Φ^{-1}, the convergence (B.6) holds almost surely. To show the $L^p(\Omega)$ consistency, it is sufficient to show that

$$\sup_{|K|\geq 1} \mathbb{E}\left|X_K^{\text{prox}}\right|^p < \infty.$$

Indeed, this implies that, for any $q < p$, the family

$$\left|X_K^{\text{prox}}\right|^q, \qquad |K| = 1, 2, \ldots$$

is uniformly integrable. This implies that (B.6) holds in $L^q(\Omega)$. Since p is arbitrary, the result would hold.

In the sequel, we assume without loss of generality that the bijective function f is strictly increasing (the symmetric case with decreasing f can be tackled in a similar fashion). It holds, for any $j \in K$

$$\hat{F}_K(Y^j) = \frac{1}{|K|+1}\sum_{k\in K} \mathbf{1}_{Y^k \leq Y^j} = \frac{1}{|K|+1}\sum_{k\in K} \mathbf{1}_{X^k \leq X^j} = \frac{1}{|K|+1}\Big(1 + \sum_{k\neq j, k\in K} \mathbf{1}_{X^k \leq X^j}\Big).$$

By independence

$$\mathbb{E}\left|X_K^{\text{prox}}\right|^p = \int_{\mathbb{R}^{|K|}} \left|\Phi^{-1}\left(\frac{1}{|K|+1}\Big(1+\sum_{k=1,k\neq j}^{|K|}\mathbf{1}_{x_k\leq x_j}\Big)\right)\right|^p d\Phi(x_1)\dots d\Phi(x_j)\dots d\Phi(x_{|K|}).$$

By considering the cases where exactly $m = 0, 1, \dots, |K|-1$ of the variables are less than x_j, integrating first with respect to variables $x_k, k \neq j$, and using change of variable $y = \Phi(x_j)$ we get

$$\begin{aligned}
&\int_{\mathbb{R}^{|K|}} \left|\Phi^{-1}\left(\frac{1}{|K|+1}\Big(1+\sum_{k=1,k\neq j}^{|K|}\mathbf{1}_{x_k\leq x_j}\Big)\right)\right|^p d\Phi(x_1)\dots d\Phi(x_j)\dots d\Phi(x_|K|) \\
&= \sum_{m=0}^{|K|-1}\binom{|K|-1}{m}\int_{\mathbb{R}}\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p \Phi(x_j)^m(1-\Phi(x_j))^{|K|-1-m}d\Phi(x_j) \\
&= \sum_{m=0}^{|K|-1}\binom{|K|-1}{m}\int_0^1\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p y^m(1-y)^{|K|-1-m}dy \\
&= \sum_{m=0}^{|K|-1}\binom{|K|-1}{m}\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p B(m+1,|K|-m),
\end{aligned}$$

where $B(x,y)$ denotes the Beta function. Since

$$\binom{|K|-1}{m}B(m+1,|K|-m) = \frac{(|K|-1)!}{m!(|K|-m-1)!}\cdot\frac{m!(|K|-m-1)!}{|K|!} = \frac{1}{|K|},$$

we have

$$\begin{aligned}
\sum_{m=0}^{|K|-1}\binom{|K|-1}{m}\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p B(m+1,|K|-m) &= \frac{1}{|K|}\sum_{m=0}^{|K|-1}\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p \\
= \frac{1}{|K|}\sum_{m\leq\frac{|K|-1}{2}}\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p + \frac{1}{|K|}\sum_{m\geq\frac{|K|-1}{2}}\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p &= (A)+(B).
\end{aligned}$$

In (A), $m \leq \frac{|K|-1}{2}$ so that $\frac{m+1}{|K|+1} \leq \frac{1}{2}$. Using the fact that $|\Phi^{-1}(x)|^p$ is decreasing on $x \in \left(0, \frac{1}{2}\right]$ we get

$$\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p \leq \left|\Phi^{-1}\left(\frac{x}{|K|+1}\right)\right|^p$$

for all $x \in [m, m+1]$. Therefore (A) can be bounded as follows.

$$(A) = \frac{1}{|K|}\sum_{m\leq\frac{|K|-1}{2}}\int_m^{m+1}\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p dx \leq \frac{1}{|K|}\sum_{m\leq\frac{|K|-1}{2}}\int_m^{m+1}\left|\Phi^{-1}\left(\frac{x}{|K|+1}\right)\right|^p dx$$

$$= \frac{1}{|K|}\int_0^{\frac{|K|+1}{2}}\left|\Phi^{-1}\left(\frac{x}{|K|+1}\right)\right|^p dx \leq \frac{|K|+1}{|K|}\int_0^1\left|\Phi^{-1}(y)\right|^p dy \leq 2\int_{-\infty}^{\infty}|x|^p d\Phi(x) < \infty.$$

A similar argument holds for (B), where we use that $|\Phi^{-1}(x)|^p$ is increasing on $x \in \left[\frac{1}{2}, 1\right)$ and that, for all $x \in [m+1, m+2]$,

$$\left|\Phi^{-1}\left(\frac{m+1}{|K|+1}\right)\right|^p \leq \left|\Phi^{-1}\left(\frac{x}{|K|+1}\right)\right|^p.$$

Consequently, for every $p \geq 1$ the family $X_K^{\text{prox}}, |K| = 1, 2, \ldots$ is uniformly integrable. This concludes the proof. □

COROLLARY B.3. *Let $t, s \in \pi_N$ be fixed and let $\hat{R}_J(t,s)$ be given by (3.5). Then, for any $p \geq 1$ and as $|J| \to \infty$, we have $\hat{R}_J(t,s) \to R(t,s)$ in $L^p(\Omega)$.*

PROOF. The claim follows directly from Proposition B.1 together with the fact that $|J| \leq \min(|K(t)|, |K(s)|)$ and that X is Gaussian. □

COROLLARY B.4. *Let $t \in \Delta \cap \pi_N$ be fixed and let the approximation $\hat{X}_{t,N}^j$ of the predictor be given by (3.6). Then, as $\min_{(t,s)\in\pi_N} |J(t,s)| \to \infty$,*

$$\hat{X}_{t,N}^j \to \hat{X}_{t,N}$$

in probability, where $\hat{X}_{t,N}$ is given by (2.16).

PROOF. The claim follows directly from Proposition B.1 and Corollary B.3. □

REMARK B.2. Note that the result above does not necessarily hold for the stronger $L^p(\Omega)$ convergence. Indeed, while Proposition B.1 ensures that $X_{t,K}^{\text{prox}}$ converges in $L^p(\Omega)$, this is not necessarily true for the precision matrix $\hat{\mathbf{R}}_{N,\Delta}^{-1}$.

LEMMA B.1. *Let $\hat{\rho}_M(t,t|\Delta \cap \pi_N)$ be given by (3.7) and let Z_N be a centered Gaussian vector with covariance $\hat{\boldsymbol{R}}_N$. Let $\hat{\rho}(t,t|\Delta \cap \pi_N)$ denote the conditional variance associated to Z_N and let $\rho(t,t|\Delta \cap \pi_N)$ denote the conditional variance associated to X, given by (2.16). Then, as $M \to \infty$, we have*

$$\hat{\rho}_M(t,t|\Delta \cap \pi_N) \to \hat{\rho}(t,t|\Delta \cap \pi_N)$$

in probability. Moreover, $\hat{\rho}(t,t|\Delta \cap \pi_N)$ is deterministic and satisfies, as $\min_{(t,s)\in\pi_N} |J(t,s)| \to \infty$,

$$\hat{\rho}(t,t|\Delta \cap \pi_N) \to \rho(t,t|\Delta \cap \pi_N).$$

PROOF. The first claim is a simple application of the law of large numbers. For the second claim, it suffices to observe that $\hat{\rho}(t,t|\Delta \cap \pi_N)$ depends only on the covariance $\hat{\mathbf{R}}_N$ given by (3.5). Thus the claim follows directly from Corollary B.3. □

LEMMA B.2. *Let $L > 0$ be fixed and let $\phi(z)$ be the density function of the standard normal distribution. Then for every bounded function f the mapping $\mathbb{R}^2 \mapsto \mathbb{R}$ defined by*

$$(x, c) \mapsto \int_{-L}^{L} f(x + zc)\phi(z)dz$$

is continuous.

PROOF. Let $x_n \to x$ and $c_n \to c$. Performing a change of variable $z = \frac{x-x_n}{c_n} + \frac{cy}{c_n}$ we see that

$$\int_{-L}^{L} f(x_n + zc_n)\phi(z)dz = \frac{c}{\sqrt{2\pi}c_n} \int_{A_n}^{B_n} f(x + cy) \exp\left(-\frac{1}{2}\left(\frac{x - x_n}{c_n} + \frac{cy}{c_n}\right)^2\right) dy,$$

where

$$A_n = \frac{x_n - x}{c} - \frac{c_n L}{c} \qquad \text{and} \qquad B_n = \frac{x_n - x}{c} + \frac{c_n L}{c}.$$

Since $A_n \to -L$ and $B_n \to L$, the claim follows immediately by continuity of the exponential function and the fact that boundedness of f allows us to apply dominated convergence theorem. □

PROOF OF THEOREM 3.2. Let t and j be fixed. We proceed by studying the limits one by one and divide the proof accordingly into four steps.

Step 1 (limit $M \to \infty$):

Since the function $\hat{f}_K$ is bounded, Lemma B.1 and Lemma B.2 give

$$\lim_{M\to\infty} \hat{Y}_{t,L,J^*,N,M} = \int_{-L}^{L} \hat{f}_K\left(t, \hat{X}^j_{t,N} + z\sqrt{\hat{\rho}(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz.$$

Step 2 (limit $J^* \to \infty$):

We write, using in the sequel the notation $\hat{X}_{t,N,j}$ for the discrete approximation of $\mathbb{E}[X^j_t|\mathcal{F}_\Delta]$ obtained via Theorem 2.4,

$$\begin{aligned}
&\int_{-L}^{L} \hat{f}_K\left(t, \hat{X}^j_{t,N} + z\sqrt{\hat{\rho}(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz - \int_{-L}^{L} f\left(t, \hat{X}_{t,N,j} + z\sqrt{\rho(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz \\
&= \int_{-L}^{L} \hat{f}_K\left(t, \hat{X}^j_{t,N} + z\sqrt{\hat{\rho}(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz - \int_{-L}^{L} f\left(t, \hat{X}^j_{t,N} + z\sqrt{\hat{\rho}(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz \\
&+ \int_{-L}^{L} f\left(t, \hat{X}^j_{t,N} + z\sqrt{\hat{\rho}(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz - \int_{-L}^{L} f\left(t, \hat{X}_{t,N,j} + z\sqrt{\rho(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz.
\end{aligned}$$

For the second difference, note that $\hat{X}^j_{t,N} \to \hat{X}_{t,N,j}$ by Corollary B.4 and $\hat{\rho}(t, t|\Delta \cap \pi_N) \to \rho(t, t|\Delta \cap \pi_N)$ by Lemma B.1. Since f is continuous, we may apply Lemma B.2 and the continuous mapping theorem to conclude

$$\int_{-L}^{L} f\left(t, \hat{X}^j_{t,N} + z\sqrt{\hat{\rho}(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz - \int_{-L}^{L} f\left(t, \hat{X}_{t,N,j} + z\sqrt{\rho(t, t|\Delta \cap \pi_N)}\right)\phi(z)dz \to 0.$$

The first difference converges to 0 as well. To see this, we apply the fact that the empirical quantile function $\hat{Q}_{t,K}$ converges uniformly (and almost surely) to Q_t on compact intervals. Hence, $\hat{f}_K$ converges uniformly to f as well.

Thus, we have

$$\lim_{J^*\to\infty}\int_{-L}^{L}\hat{f}_K\left(t,\hat{X}^j_{t,N}+z\sqrt{\hat{\rho}(t,t|\Delta\cap\pi_N)}\right)\phi(z)dz=\int_{-L}^{L}f\left(t,\hat{X}_{t,N,j}+z\sqrt{\rho(t,t|\Delta\cap\pi_N)}\right)\phi(z)dz.$$

Step 3 (limit $N\to\infty$):
By Theorem 2.4 we have $\hat{X}_{t,N,j}\to\hat{X}^j_t=\mathbb{E}[X^j_t|\mathcal{F}_\Delta]$ in $L^2(\Omega)$ as $N\to\infty$. Consequently, we also have $\rho(t,t|\Delta\cap\pi_N)\to\rho(t,t|\Delta)$. As in the previous step, we may apply continuity of f, Lemma B.2, and the continuous mapping theorem to get

$$\lim_{N\to\infty}\int_{-L}^{L}f\left(t,\hat{X}_{t,N,j}+z\sqrt{\rho(t,t|\Delta\cap\pi_N)}\right)\phi(z)dz=\int_{-L}^{L}f\left(t,\hat{X}^j_t+z\sqrt{\rho(t,t|\Delta)}\right)\phi(z)dz.$$

Step 4 (limit $L\to\infty$):
Suppose first that $f\geq 0$. In this case the monotone convergence theorem applies and we obtain

$$\lim_{L\to\infty}\int_{-L}^{L}f\left(t,\hat{X}^j_t+z\sqrt{\rho(t,t|\Delta)}\right)\phi(z)dz=\int_{-\infty}^{\infty}f\left(t,\hat{X}^j_t+z\sqrt{\rho(t,t|\Delta)}\right)\phi(z)dz.$$

In the general case, bijectivity of $x\mapsto f(t,x)$ implies that f changes sign only once. Considering integrals separately before and after this change point and applying the monotone convergence theorem concludes step 4 and the whole proof.

□

REFERENCES

Biagini, F., Hu, Y., Øksendal, B. and Zhang, T. (2008). *Stochastic calculus for fractional Brownian motion and applications. Probability and its Applications (New York)*. Springer-Verlag London, Ltd., London. MR2387368 (2010a:60145)

Coeurjolly, J. F. (2000). Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *J. Stat. Software* **5** 1-53.

Delaigle, A. and Hall, P. (2013). Classification using censored functional data. *J. Amer. Stat. Assoc.* **108** 1269-1283.

Delaigle, A. and Hall, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika* **103** 779-799.

Delaigle, A., Hall, P., Huang, W. and Kneip, A. (2020). Estimating the Covariance of Fragmented and Other Related Types of Functional Data. *J. Amer. Stat. Assoc.*

Descary, M. H. and Panaretos, V. (2019). Recovering covariance from functional fragments. *Biometrika* **106** 145-160.

Ferraty, F., ed. (2011). *Recent advances in Functional Data Analysis and related topics*. Springer.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics*. Springer-Verlag, New York.

Goldberg, Y., Ritov, Y. and Mandelbaum, A. (2014). Predicting the continuation of a function with applications to call center data. *J. Statist. Plann. Inf.* **147** 53-65.

James, G., Hastie, T. J. and Sugar, C. (2000). Principal Component Models for Sparse Functional Data. *Biometrika* **87** 587-602.

Janson, S. (1997). *Gaussian Hilbert spaces. Cambridge Tracts in Mathematics* **129**. Cambridge University Press, Cambridge. MR1474726

Kneip, A. and Liebl, D. (2020). On the optimal reconstruction of partially observed functional data. *Ann. Statist.* **48** 1692-1717.

Kraus, D. (2015). Components and completion of partially observed functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **77** 777-801.

Liebl, D. (2013). Modeling and forecasting electricity spot prices: a functional data perspective. *Ann. Appl. Statist.* **7** 1562-1592.

Liebl, D. and Rameseder, S. (2019). Partially Observed Functional Data: The Case of Systematically Missing Parts. *Comput. Statist. and Data Anal.* **131** 104-115.

Mishura, Y. S. (2008). *Stochastic calculus for fractional Brownian motion and related processes. Lecture Notes in Mathematics* **1929**. Springer-Verlag, Berlin. MR2378138 (2008m:60064)

Nualart, D. (2006). *The Malliavin calculus and related topics*, second ed. *Probability and its Applications (New York)*. Springer-Verlag, Berlin. MR2200233 (2006j:60004)

Pipiras, V. and Taqqu, M. S. (2001). Are classes of deterministic integrands for fractional Brownian motion on an interval complete? *Bernoulli* **7** 873–897. MR1873833

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer.

Sottinen, T. and Viitasaari, L. (2016). Stochastic Analysis of Gaussian Processes via Fredholm Representation. *Int. J. Stoch. Anal.* Art. ID 8694365, 15. MR3536393

Sottinen, T. and Viitasaari, L. (2017). Prediction law of fractional Brownian motion. *Statist. Probab. Lett.* **129** 155–166. MR3688528

Sottinen, T. and Yazigi, A. (2014). Generalized Gaussian bridges. *Stochastic Process. Appl.* **124** 3084–3105. MR3217434

Tudor, C. A. (2013). *Analysis of Variations for Self-similar Processes*. Springer International Publishing.

Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology* **13**.

Viitasaari, L. and Ilmonen, P. (2020). On modeling a class of weakly stationary processes. *Front. Appl. Math. Stat.* **15**.

Wang, J., Chiou, J. and Müller, H. G. (2016). Functional data analysis. *Annual Review of Statistics and Its Applications* **3** 257-295.

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

Yazigi, A. (2015). Representation of self-similar Gaussian processes. *Statist. Probab. Lett.* **99** 94–100. MR3321501

Université de Namur
Département de Mathématique
Rue de Bruxelles, 61
B-5000 Namur Belgium
E-mail: germain.vanbever@unamur.be
URL: https://sites.google.com/site/germainvanbever