ON THE OPTIMAL RECONSTRUCTION OF PARTIALLY OBSERVED FUNCTIONAL DATA

By Dominik Liebl and Alois Kneip

University of Bonn

If only fragments of functions are observable and none of these fragments cover the total domain, then it is impossible to estimate the covariance function over its total support. We propose a new functional PCA based prediction procedure that allows to reconstruct functional data from their fragmental observations under this challenging setup. By contrast to classical prediction models, we can proof the optimality of our prediction model without making use of the usual, but unverifiable, orthogonality assumption on the prediction error. This allows us to identify situations under which our prediction procedure leads to a perfect reconstruction of the fragmental observations, i.e., without any prediction error. Finite sample properties are investigated through simulations and a real data application.

1. Introduction. Let $(X_t)_t$ be a stationary weakly dependent time series of random functions $t \in \{1, \dots, T\}$, where each random function X_t is square integrable, i.e., $X_t \in L^2(I_0)$ with $I_0 = [a, b]$. Though, instead of observing realizations of X_t, we only observe "small" fragments X_t^{sm}, where $X_t^{\text{sm}}(u) = X_t(u)$ with $u \in I_{\text{sm},t}$ and $I_{\text{sm},t} \subset I_0$. This situation is illustrated in the left panel of Figure 1. There the (electricity) price function of March 22, 2011, covers only a subset of the total support $I_0 = [a, b]$. See our real data application in Section 6 for more information on these price functions.

The situation becomes unpleasant, if $I_{\text{sm},t} \subset I_0$ for all $t \in \{1, \dots, T\}$, which is the case in our real data application. If the functions X_t are observable only over *proper* subsets $I_{\text{sm},t} \subset I_0$, then it is impossible to estimate the covariance function $\gamma(u, v) = \text{Cov}(X_t(u), X_t(v))$ over its total support $I_0^2 = [a, b]^2$. For instance, we cannot estimate the value $\gamma(a, b)$, since we cannot observe $X_t(a)$ and $X_t(b)$ for the same time point t. The covariance function γ is then only estimable for points $(u, v) \in \bigcup_t I_{\text{sm},t}^2$. The empirical counterpart of the feasible region $\bigcup_t I_{\text{sm},t}^2$ is illustrated by the gray area in the right panel of Figure 1. The white areas at the outer off-diagonal parts of the square $[a, b]^2$ describe the infeasible regions.

AMS 2000 subject classifications: Primary 62M20, 62H25, 62G05, 62G08, 62M10

Keywords and phrases: functional data analysis, functional linear prediction, functional principal components, time series analysis

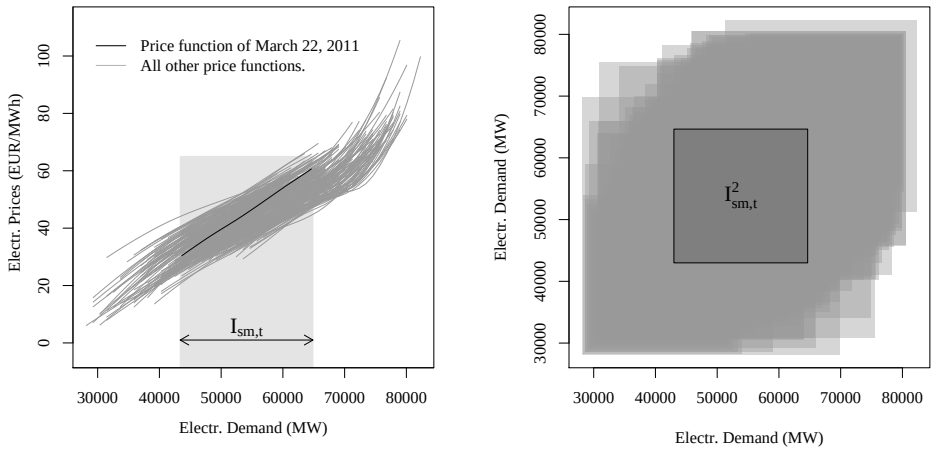

FIG 1. LEFT PANEL: *Fragmental random functions observed only over proper subset* I_{sm} *of the total support* $[a, b]$. RIGHT PANEL: *The gray area correspond to the set* $\bigcup_t I^2_{\text{sm},t}$ *over which the covariance function is estimable.*

The German electricity market, like many other electricity markets, provides purchase guarantees for renewable energy sources (RES). Therefore, the relevant variable for pricing at the energy exchange is electricity demand minus electricity infeeds from RES (Nicolosi, 2010). Correspondingly, "electricity demand" in Figure 1 actually denotes *residual* electricity demand, i.e., electricity demand minus infeeds from RES. Practitioners in energy economics are interested in comparative statics with respect to changes in residual electricity demand (cf. Weigt 2009 and Hirth 2013). As these changes can be drastic, it is necessary to predict the partially observed price functions over a large as possible part of the total support. Typically, this is done through extrapolation based on rigid parametric model assumptions.

Our main contribution is a nonparametric functional PCA (FPCA) based prediction model

$$X_t^{\text{la}}(u) = [SX_t^{\text{sm}}](u) + Z_t(u), \quad u \in I_{\text{la},t}$$

that allows to reconstruct the partially observed, i.e., "small" random functions X_t^{sm} over a "large" as possible part of the total domain $I_{\text{la},t} \subseteq [a, b]$ with $I_{\text{sm},t} \subseteq I_{\text{la},t}$, where $S : L^2(I_{\text{sm},t}) \to L^2(I_{\text{la},t})$ denotes the prediction operator and $Z_t \in L^2(I_{\text{la},t})$ the prediction error. Whether the functions can be reconstructed over the total domain, i.e., $I_{\text{la},t} = [a, b]$, eventually depends on how much of the covariance function is estimable. In Section 3.1 we propose an prediction algorithm which typically allows to reconstruct the total

trajectories.

In our theoretical part, we show that our predictor $\widetilde{X}_t^{\mathrm{la}} = SX_t^{\mathrm{sm}}$ is the best linear predictor under the L2 loss. Under similar setups, this property is already claimed by two other functional prediction models, namely that of Goldberg, Ritov and Mandelbaum (2014) and Kraus (2014). Though, these approaches do not consider the case of truly fragmental observations. I.e., they assume that the covariance function can still be estimated over its total support $[a, b]^2$, which simplifies the prediction problem and constitutes quite a restrictive assumption within this problem set. Beyond this, the work of Goldberg, Ritov and Mandelbaum (2014) considers only the case of finite dimensional functional data. Their results have well known counterparts in multivariate statistics. Kraus (2014) only informally claims the optimality property, but lacks of a formal statement which is common in the literature on the classical functional prediction model (cf. Yao, Müller and Wang (2005b), Crambes and Mas (2013), and Hörmann and Kidziński (2015)). In fact, under the common (unverifiable) assumption that the prediction error Z_t is orthogonal to the predictor X_t^{sm}, this optimality result is trivial and does not demand for a formal statement. It follows directly from arguments as in He, Müller and Wang (2003). Delaigle and Hall (2013) consider the case of truly fragmental functional observations, though, within a classification framework. The FPCA based model of James, Hastie and Sugar (2000) and James and Hastie (2001) allows for sparse functional data that are observed over proper subintervals, too. This approach makes use of rigid parametric (mixed effects) model assumptions in order to cope with the sparseness in the data. By contrast, we consider a less rigid nonparametric model setup. A further reconstruction method is the PACE estimation procedure (cf. Yao, Müller and Wang (2005a) and Yao, Müller and Wang (2005b)). Though, this method assumes too that covariance function is estimable over its total support $[a, b]^2$.

Our prediction model is fundamentally different to the above cited classical functional prediction models, since the predictor function X_t^{sm} and the function X_t^{la}, that we aim to predict, are fragments of the *same* underlying function X_t. This specific situation allows us to propose an prediction operator S for which it can be shown that the prediction error Z_t is orthogonal to the predictor X_t^{sm}. That is, in our case orthogonality is not an unverifiable model assumption, but a proven fact that applies in principle to any process X_t. This orthogonality result makes our optimality result much more substantial than the optimality results of the classical functional prediction models. Furthermore, we are able to derive an expression for the variance of the prediction error which allows to identify situations under which it is

possible to recover the partially observed functions perfectly, i.e., without any prediction error. These results go beyond that of Kraus (2014), who consider a similar prediction model as we do. Though, the author treats his model as a classical functional prediction model and therefore misses to exploit the specific prediction context.

The rest of this paper is structured as follows: The next section introduces our statistical model and assumptions. Section 2 presents our main theoretical results with respect our prediction model. Section 3 contains our estimation procedure and the asymptotic results. Section 5 presents our simulation study. The real data application can be found Section 6 and Section 7 concludes. All proofs can be found in the supplemental paper.

2. Optimal prediction of partially observed functions. Let us first fix our basic setup. Each random function $X_t \in H$ is an element of the Hilbert space $H = L^2(I_0)$, with $I_0 = [a, b] \subset \mathbb{R}$, $0 < a < b < \infty$, inner product $\langle x, y\rangle_H = \int_{I_0} x(s)y(s)ds$, and norm $||y||_H = (\int_{I_0}(y(s))^2 ds)^{1/2}$ for all $x, y \in H$. Furthermore, we need to assume that $\mathbb{E}(||X_t||_H^4) < \infty$, which assures existence (and estimability) of the mean function $\mu(u) = \mathbb{E}(X_t(u))$, with $u \in I_0$, and of the covariance function $\gamma(u, v) = \mathrm{Cov}(X_t(u), X_t(v))$, with $(u, v) \in I_0^2$. In order to simplify the notation within this theoretical section, we assume centered random functions, i.e., $\mu(u) = 0$ for all $u \in I_0$. This is, of course, without loss of generality.

As outlined in the introduction, we do not observe the random functions X_t, but only "small" fragments X_t^{sm}, where $X_t^{\mathrm{sm}}(u) = X_t(u)$ with $u \in I_{\mathrm{sm},t}$ and $I_{\mathrm{sm},t} \subset I_0$. The set $I_{\mathrm{sm},t}$ denotes a random set and can be thought of as some censoring process. In principle our theory applies to very general cases of random subsets $I_{\mathrm{sm},t}$, i.e., containing, e.g., "wholes". Though, for simplicity we treat $I_{\mathrm{sm},t}$ as a random subinterval, i.e., $I_{\mathrm{sm},t} = [A_t, B_t] \subset I_0 = [a, b]$ with $a \leq A_t < B_t \leq b$. In the following we list our regularity assumptions with respect to the random subsets $I_{\mathrm{sm},t}$:

RS Let $(I_{\mathrm{sm},t})_t$ denote an strictly stationary, weakly dependent series of random sets with $I_{\mathrm{sm},t} \subset I_0 = [a, b]$, where $I_{\mathrm{sm},t}$ is independent from X_s and ε_{jk} for all $t, s, k \in \{1, \ldots, T\}$ and all $j \in \{1, \ldots, n\}$. Further, let for all t, (i) $\mathbb{P}(I_{\mathrm{sm},t} = \emptyset) = 0$, (ii) $\mathbb{P}(I_0 \setminus I_{\mathrm{sm},t} = \emptyset) = 0$, and (iii) $\mathbb{P}(u \in I_{\mathrm{sm},t}) > 0$ for all $u \in I_0$.

Remarks on Assumption RS. Property (i) excludes the degenerated cases and property (ii) assures that $I_{\mathrm{sm},t}$ is a *proper* subset, i.e., that $I_{\mathrm{sm},t} \subset I_0$. The latter implies that the composed set $\bigcup_t I_{\mathrm{sm},t}^2$ does not cover the total square $I_0^2 = [a, b]^2$. This has far-reaching consequences, since then it is impossible to estimate the covariance $\gamma(u, v)$ for all $(u, v) \in I_0^2$; see also

the right panel of Figure 1. Property (iii) assures that every point $u \in I_0$ is covered by the random subset $I_{\mathrm{sm},t}$ with a strictly positive probability. Under the latter property we can estimate the mean function $\mu(u)$ for all $u \in I_0$.

Now we can consider the prediction problem from a general theoretical point of view, where we seek to reconstruct a partially observed random function X_t^{sm} over a large as possible interval $I_{\mathrm{la},t}$, given the information with respect to the observed "small" fraction X_t^{sm}. The large interval $I_{\mathrm{la},t}$ must be chosen sufficiently small, such that $\mathbb{P}(I_{\mathrm{la},t} \subseteq I_{\mathrm{sm},s}) > 0$ for $s \neq t$. That is, with positive probability there are random subintervals $I_{\mathrm{sm},s}$ that contain the whole large interval $I_{\mathrm{la},t}$. Otherwise, we cannot consistently estimate the underlying covariance function γ_t^{la} over its total support, where $\gamma_t^{\mathrm{la}}(u,v) := \gamma(u,v)$ for all $(u,v) \in I_{\mathrm{la},t}^2$. In order to simplify the notation, we suppress in the following the index t for covariances and intervals.

Let us first assume the ideal situation that we know γ^{la}. But if $\gamma^{\mathrm{la}}(u,v)$ is known over I_{la}^2, we also know $\gamma^{\mathrm{sm}}(u,v)$ over $I_{\mathrm{sm}}^2 \subset I_{\mathrm{la}}^2$ and therefore γ^{la} defines a unique covariance operator Γ^{sm} of $X_t^{\mathrm{sm}} \in L^2(I_{\mathrm{sm}})$ as

$$\Gamma^{\mathrm{sm}}(x)(u) = \int_{I_{\mathrm{sm}}} \gamma^{\mathrm{sm}}(u,v)x(v)dv, \quad x \in L^2(I_{\mathrm{sm}}),$$

where $\lambda_1^{\mathrm{sm}} \geq \lambda_2^{\mathrm{sm}} \geq \ldots$ and $\phi_1^{\mathrm{sm}}, \phi_2^{\mathrm{sm}}, \cdots \in L^2(I_{\mathrm{sm}})$ denote eigenvalues and eigenfunctions of Γ^{sm}, and where $(\phi_k^{\mathrm{sm}})_{k\geq 1}$ forms an orthonormal basis system that spans the space $L^2(I_{\mathrm{sm}})$. The covariance operators Γ and Γ^{la} of the total random function X_t and the large random fraction X_t^{la}, as well as their eigenvalues and eigenfunctions, are defined correspondingly.

Any centered fragment X_t^{sm} adopts then the well known Karhunen-Loéve (KL) representation

$$X_t^{\mathrm{sm}}(u) = \sum_{k=1}^{\infty} \xi_{tk}^{\mathrm{sm}} \phi_k^{\mathrm{sm}}(u), \quad u \in I_{\mathrm{sm}}, \tag{1}$$

where the Functional Principal Component (FPC) scores are defined by $\xi_{tk}^{\mathrm{sm}} = \langle X_t^{\mathrm{sm}}, \phi_k^{\mathrm{sm}} \rangle$, with $\langle X_t^{\mathrm{sm}}, \phi_k^{\mathrm{sm}} \rangle = \int_{I_{\mathrm{sm}}} X_t^{\mathrm{sm}}(x)\phi_k^{\mathrm{sm}}(x)dx$, and where $\mathbb{E}(\xi_{tk}^{\mathrm{sm}}) = 0$ and $\mathbb{E}(\xi_{tk}^{\mathrm{sm}} \cdot \xi_{tl}^{\mathrm{sm}}) = \lambda_k^{\mathrm{sm}}$ for all $k = l$ and zero else. The KL representations of the total random function X_t and the large random fragment X_t^{la} are defined correspondingly.

The properties of the KL representation imply that if $\lambda_k^{\mathrm{sm}} > 0$, the value $\phi_k^{\mathrm{sm}}(u)$ can be obtained as the slope-coefficient of a simple linear regression of the small fragment $X_t^{\mathrm{sm}}(u)$ on the small FPC score ξ_{tk}^{sm}:

$$\phi_k^{\mathrm{sm}}(u) = \frac{\mathbb{E}(\xi_{tk}^{\mathrm{sm}}\, X_t^{\mathrm{sm}}(u))}{\lambda_k^{\mathrm{sm}}} = \frac{1}{\lambda_k^{\mathrm{sm}}} \int_{I_{\mathrm{sm}}} \phi_k^{\mathrm{sm}}(v)\gamma^{\mathrm{sm}}(u,v)dv,$$

for $u \in I_{\text{sm}}$, where the latter equation is a direct result form the definition of the kth FPC score ξ_{tk}^{sm}.

This can obviously be generalized to an enlarged set of points $u \in I_{\text{la}}$ by regressing the large fragment $X_t^{\text{la}}(u)$ on the small FPC score ξ_{tk}^{sm}, which leads to our definition of the predictive kth eigenfunction:

$$\tilde{\phi}_k^{\text{la}}(u) \quad := \quad \frac{\mathbb{E}(\xi_{tk}^{\text{sm}}\, X_t^{\text{la}}(u))}{\lambda_k^{\text{sm}}} = \frac{1}{\lambda_k^{\text{sm}}} \int_{I_{\text{sm}}} \phi_k^{\text{sm}}(v)\gamma^{\text{la}}(u,v)dv \tag{2}$$

for $u \in I_{\text{la}}$ if $\lambda_k^{\text{sm}} > 0$, while $\tilde{\phi}_k^{\text{la}}(x) := 0$ if $\lambda_k^{\text{sm}} = 0$. The latter equality in Eq. (2) follows from the definition of the kth FPC score ξ_{tk}^{sm} and the KL representation of X_t^{la}. In fact, $\tilde{\phi}_k^{\text{la}}(u)$ is predictive only with respect to the remainder fragments $u \in I_{\text{la}} \setminus I_{\text{sm}}$, but for $v \in I_{\text{sm}}$ we have that $\tilde{\phi}_k^{\text{la}}(v) = \phi_k^{\text{sm}}(v)$. Furthermore, continuity of γ^{la} implies continuity of $\tilde{\phi}^{\text{la}}$.

Given the observable centered small fragment X_t^{sm} we now propose to predict the remainder fragments by applying Eq. (1) with respect to the predictive eigenfunctions $\tilde{\phi}_k^{\text{la}}$:

$$\widetilde{X}_t^{\text{la}}(u) \quad := \quad \sum_{k=1}^{\infty} \xi_{tk}^{\text{sm}}\, \tilde{\phi}_k^{\text{la}}(u), \quad u \in I_{\text{la}}. \tag{3}$$

Note that $\tilde{X}_t^{\text{la}}$ is continuous and $\tilde{X}_t^{\text{la}}(v) = X_t^{\text{sm}}(v)$ for all $v \in I_{\text{sm}}$, due to the above described properties of $\tilde{\phi}_k^{\text{la}}(u)$.

2.1. *Theoretical properties.* First of all, we need to address the question whether Eq. (3) actually determines a well-defined random process $\tilde{X}_t^{\text{la}}$ on $L^2(I_{\text{la}})$, i.e., whether the infinite sum Eq. (3) converges in the L2 sense and whether the random functions $\tilde{X}_t^{\text{la}}$ are continuous. Both is in fact true and for ease of reference we state this in the following theorem:

THEOREM 2.1. *The random function $\tilde{X}_t^{\text{la}}$ defined in Eq. (3) has a continuous and finite variance function $\mathbb{V}(\tilde{X}_t^{\text{la}}(u)) < \infty$ for all $u \in I_{\text{la}}$.*
The proof can be found in Appendix B of the supplemental paper.

In the following we analyze the theoretical properties of the prediction error

$$Z_t(u) = X_t^{\text{la}}(u) - \widetilde{X}_t^{\text{la}}(u), \quad \text{for all} \quad u \in I_{\text{la}}, \tag{4}$$

but with a particular interest for all $u \in I_{\text{la}} \setminus I_{\text{sm}}$, since anyways $Z_t(v) = 0$ for all $v \in I_{\text{sm}}$; see our discussion of Eq. (3). An obvious consequence of

Eq. (3) is that $Z_t(u)$ has mean zero for all $u \in I_{\mathrm{la}}$, i.e., the prediction procedure is unbiased. Result (a) in the following theorem shows that the prediction error $Z_t(u)$ is orthogonal to the predictor $X_t^{\mathrm{sm}}(u)$ which serves as an auxiliary result for result (b). Result (b) shows that $\widetilde{X}_t^{\mathrm{la}}(u)$ is the optimal linear predictor of the true value $X_t^{\mathrm{la}}(u)$ for all $u \in I_{\mathrm{la}}$, i.e., having the lowest prediction error variance among all linear predictors. Finally, result (c) allows us to identify specifications of random functions X_t that can be predicted without any prediction error.

THEOREM 2.2 (Optimal linear prediction).

(a) Under our setup we obtain that for every $v \in I_{\mathrm{sm}}$ and $u \in I_{\mathrm{la}} \setminus I_{\mathrm{sm}}$

$$\mathbb{E}\left(X_t^{\mathrm{sm}}(v) Z_t(u)\right) = 0 \quad \textit{as well as} \tag{5}$$

$$s_Z^2(u) := \mathbb{E}\left(Z_t(u)^2\right) = \gamma^{\mathrm{la}}(u,u) - \sum_{k=1}^{\infty} \lambda_k^{\mathrm{sm}} \tilde{\phi}_k^{\mathrm{la}}(u)^2. \tag{6}$$

(b) For any continuous linear functional $\ell_u : L^2(I_{\mathrm{sm}}) \to \mathbb{R}$ we have that

$$\mathbb{E}\left((X_t^{\mathrm{la}}(u) - \ell_u(X_t^{\mathrm{sm}}))^2\right) \geq \mathbb{E}\left((X_t^{\mathrm{la}}(u) - \widetilde{X}_t^{\mathrm{la}}(u))^2\right) = s_Z^2(u)$$

for all $u \in I_{\mathrm{la}}$.

(c) Let X_t be a Gaussian process, i.e., all FPC scores are Gaussian random variables. Then the variance of the prediction error can be written as

$$s_Z^2(u) = \frac{1}{2}\,\mathbb{E}\left(\mathbb{E}\left(\left(X_t^{\mathrm{la}}(u) - X_s^{\mathrm{la}}(u)\right)^2 \big| X_t^{\mathrm{sm}} = X_s^{\mathrm{sm}}\right)\right) + \mathcal{O}(r^{|t-s|}), \tag{7}$$

where $X_t^{\mathrm{sm}} = X_s^{\mathrm{sm}}$ denotes point-wise equality, i.e., $X_t^{\mathrm{sm}}(v) = X_s^{\mathrm{sm}}(v)$ for all $v \in I_{\mathrm{sm}}$.

The proofs can be found in Appendix B of the supplemental paper.

In order to explain result (c) of Theorem 2.2 consider first the case of i.i.d. random functions X_t, i.e., with $r = 0$ in Assumption A2. Then the variance of the prediction error can be written as

$$s_Z^2(u) = \frac{1}{2}\;\mathbb{E}\left(\mathbb{E}\left(\left(X_t^{\mathrm{la}}(u) - X_s^{\mathrm{la}}(u)\right)^2 \big| X_t^{\mathrm{sm}} = X_s^{\mathrm{sm}}\right)\right), \quad t \neq s.$$

This result tells us that there is no prediction error, i.e., $s_Z^2(u) = 0$ for all $u \in I_{gr}$, if the structure of X_t is such that the event $X_t^{\mathrm{sm}}(v) = X_s^{\mathrm{sm}}(v)$ for all $v \in I_{\mathrm{sm}}$ and $t \neq s$ implies that also $X_t^{\mathrm{la}}(u) = X_s^{\mathrm{la}}(u)$ for all $u \in I_{\mathrm{la}} \setminus I_{\mathrm{sm}}$.

The same interpretation applies to weakly dependent functional time series $(X_t)_t$, but with the additional requirement that the distance between the time points t and s needs to be large enough such that $\mathcal{O}(r^{|t-s|})$ is negligible. This additional requirement assures that the implication, "$X_t^{\text{sm}} = X_s^{\text{sm}} \Rightarrow X_t^{\text{la}} = X_s^{\text{la}}$", is not just a mere consequence of a (positive) auto correlation when $|t-s|$ is small.

To give an example, assume that for some integer $K < \infty$ we have $X_t^{\text{la}}(u) = \sum_{r=1}^{K} \theta_{tr}^{\text{la}} f_r^{\text{la}}(u)$ for all $u \in I_{\text{la}}$, where $f_1^{\text{la}}, \ldots, f_K^{\text{la}}$ are continuous orthonormal functions on I_{la}, and where $\theta_{tr}^{\text{la}} \in \mathbb{R}$ are Gaussian random variables with mean zero and $\mathbb{E}(\theta_{tk}^{\text{la}}\, \theta_{tl}^{\text{la}}) = v_k < \infty$ for all $k = l \in \{1, \ldots, K\}$ and zero else. Now consider the smaller interval I_{sm}, and suppose that also the K truncated functions $f_1^{\text{sm}}, \ldots, f_K^{\text{sm}}$, with $f_k^{\text{sm}}(u) := f_k^{\text{la}}(u)$ for all $u \in I_{\text{sm}}$ and all $k \in \{1, \ldots, K\}$, are linearly independent. This is equivalent to require that

$$\int_{I_{\text{sm}}} \left(\sum_{k=1}^{K} (\theta_{tk}^{\text{la}} - \alpha_k) f_k^{\text{la}}(u) \right)^2 du = 0, \text{ if and only if, } \theta_{t1}^{\text{la}} = \alpha_1, \ldots, \theta_{tK}^{\text{la}} = \alpha_K.$$

In other words, in this situation the knowledge of the structure of X_t^{la} on I_{sm} is sufficient to determine the corresponding coefficients $\theta_{t1}^{\text{la}}, \ldots, \theta_{tK}^{\text{la}}$, and thus uniquely determines the function X_t^{la} on the larger set I_{la}. By (7) we can then conclude that there is no prediction error, i.e., $s_Z^2(u) = 0$ for all $u \in I_{\text{la}}$. This can be made more precisely using change of basis arguments. If the orthonormal functions $f_1^{\text{la}}, \ldots, f_K^{\text{la}}$ are linear independent over I_{sm}, we have that $\sum_{k=1}^{K} \theta_{tk}^{\text{la}} f_k^{\text{la}}(u) = \sum_{k=1}^{K} \xi_{tk}^{\text{sm}} \phi_k^{\text{sm}}(u)$ for all $u \in I_{\text{sm}}$, where $\lambda_K^{\text{sm}} > 0$ and $\lambda_{K+1}^{\text{sm}} = 0$. Furthermore, there exist unique coefficients $\beta_1, \ldots, \beta_K$ such that $\phi_k^{\text{sm}}(u) = \sum_{r=1}^{K} \beta_r f_r^{\text{la}}(u)$ for all $u \in I_{\text{sm}}$. This then implies that $\tilde{\phi}_k^{\text{la}}(u) = \sum_{r=1}^{K} \beta_r f_r^{\text{la}}(u)$ as well as $X_t^{\text{la}}(u) = \sum_{r=1}^{K} \theta_{tk}^{\text{la}} f_k^{\text{la}}(u) = \sum_{k=1}^{K} \xi_{tk}^{\text{sm}} \tilde{\phi}_k^{\text{la}}(u) = \widetilde{X}_t^{\text{la}}(u)$ for all $u \in I_{\text{la}}$, i.e., $s_Z^2(u) = 0$ for all $u \in I_{\text{la}}$.

Note that the definition of the predictor in Eq. (3) as well as $s_Z^2(u) = \mathbb{V}(Z_t(u))$ are uniquely determined by the structure of the covariance function γ, and do not depend on whether the process is Gaussian or not. Hence, result (c) of Theorem 2.2 tells us that $s_Z^2(u) = 0$ if the structure of X_t is such that for two independent realizations X_t and X_s the event $X_t^{\text{sm}} = X_s^{\text{sm}}$ implies that also $X_t^{\text{la}}(u) = X_s^{\text{la}}(u)$ for all $u \in I_{\text{la}} \setminus I_{\text{sm}}$. This may be fulfilled for simple structured functional data.

2.2. *The common features and differences regarding the classical functional prediction model.* Eq. (3), together with Eq. (2), defines a linear

operator mapping $S : X_t^{\mathrm{sm}} \mapsto \widetilde{X}_t^{\mathrm{la}}$, such that

$$\widetilde{X}_t^{\mathrm{la}}(u) = \left[SX_t^{\mathrm{sm}}\right](u) \;\Leftrightarrow\; X_t^{\mathrm{la}}(u) = \left[SX_t^{\mathrm{sm}}\right](u) + Z_t(u),\ u \in I_{\mathrm{la}}, \tag{8}$$

where the prediction error Z_t is as defined in Eq. (4), and where the linear operator $S : H_{\mathrm{sm}} \to H_{\mathrm{la}}$ is defined as $S = \sum_{k=1}^{\infty} \tilde{\phi}_k^{\mathrm{la}} \otimes \phi_k^{\mathrm{sm}}$ by using the tensor product notation $x^{\mathrm{la}} \otimes y^{\mathrm{sm}}$ to denote $(x^{\mathrm{la}} \otimes y^{\mathrm{sm}})(z^{\mathrm{sm}}) = \langle y^{\mathrm{sm}}, z^{\mathrm{sm}} \rangle_{\mathrm{sm}} x^{\mathrm{la}}$, for all $y^{\mathrm{sm}}, z^{\mathrm{sm}} \in H_{\mathrm{sm}}$ and all $x^{\mathrm{la}} \in H_{\mathrm{la}}$.

Identification Restriction: As in the case of classical functional prediction models, we need to address the identifiability issue as dealing with infinite dimensional functional data generally involves always the issue of (in-principle infinitely many) zero-valued eigenvalues. I.e., we need to address the problem that $SX_t^{\mathrm{sm}} = S(X_t^{\mathrm{sm}} + W)$, where $W \in \ker(\Gamma^{\mathrm{sm}})$. We follow the typical approach and focus only on that part of SX_t^{sm} which is identifiable, i.e., for which the full rank condition $\ker(\Gamma^{\mathrm{sm}}) = \{0\}$ holds, which is equivalent to the condition that $\lambda_k^{\mathrm{sm}} > 0$ for all $k \in \{1, 2, \dots\}$. The latter explains our definition in (2), where we set $\tilde{\phi}_k^{\mathrm{la}}(x) := 0$ if $\lambda_k^{\mathrm{sm}} = 0$.

By plugging the definition of $\tilde{\phi}_k^{\mathrm{la}}$ into the definition of S and replacing X_t^{la} by its KL representation we can write Eq. (8) as

$$\begin{aligned} X_t^{\mathrm{la}}(u) &= \int_{I_{\mathrm{sm}}} \left[\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sigma_{km}}{\lambda_k^{\mathrm{sm}}} \phi_m^{\mathrm{la}}(u) \phi_k^{\mathrm{sm}}(v) \right] X_t^{\mathrm{sm}}(v)\, dv + Z_t(u) \\ \Leftrightarrow \widetilde{X}_t^{\mathrm{la}}(u) &= \int_{I_{\mathrm{sm}}} \left[\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sigma_{km}}{\lambda_k^{\mathrm{sm}}} \phi_m^{\mathrm{la}}(u) \phi_k^{\mathrm{sm}}(v) \right] X_t^{\mathrm{sm}}(v)\, dv, \end{aligned} \tag{9}$$

where $\sigma_{km} = \mathbb{E}(\xi_{tk}^{\mathrm{sm}} \xi_{tm}^{\mathrm{la}})$. For general dependent variables Y_t, instead of X_t^{la}, and under the additional (restrictive) assumption that the limit of the infinite double sum within the square brackets in Eq. (9) exits, our functional prediction model becomes equivalent to the classical functional prediction model as considered by Bosq (2000), Yao, Müller and Wang (2005b), Crambes and Mas (2013), and Kraus (2014). These classical functional prediction models use a methodology that is motivated from generalizing the multivariate linear prediction model with its typical focus on parameter estimation to an infinite dimensional functional version (see, e.g., He, Müller and Wang, 2003). In fact, the summability assumption essentially means a particular, however restrictive, focus on the so-called integral kernel parameter $\beta(u, v) = \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sigma_{km}}{\lambda_k^{\mathrm{sm}}} \phi_m^{\mathrm{la}}(u) \phi_k^{\mathrm{sm}}(v)$ and assuming the existence of $\beta(u, v)$ is, under the classical setup with $X_t \in L^2([a, b])$, equivalent to the assumption that S is a Hilbert-Schmidt operator.

By contrast to the above cited classical prediction approaches, we do not require the restrictive existence of $\beta(u, v)$, since we are anyways only

interested in estimating the integral value $\widetilde{X}_t^{\text{la}}$. Indeed, a Hilbert-Schmidt assumption on S is particularly inappropriate in our rather specific prediction context, since it rules out the identity operator. I.e., it would rule out the in (9) necessary functional identity mapping of an observed fragment X_t^{sm} to itself (remember that $I_{\text{sm}} \subset I_{\text{la}}$). Our Theorem 2.1 assures that the integral value $\widetilde{X}_t^{\text{la}}$ is well defined even though the infinite double sum within the square brackets in Eq. (9) does not necessarily converge.

Essentially, Theorems 2.1 and 2.2 follow from the very specific nature of our prediction problem where we use the observed parts of X_t to predict the missing ones. This is fundamentally, different to the classical prediction models which aim to predict some general function Y. Within the classical prediction setup, one has to postulate a model like $Y = SX + \epsilon$ and then aims to identify and estimate the operator S, or the integral value SX, which requires the unverifiable model assumption of orthogonality between the predictor X and the prediction error ϵ (see, e.g., Crambes and Mas 2013, Section 1.2, or Hörmann and Kidziński 2015, Section 2.3).

By contrast, we derive our prediction model in Eq. (8) on basis of the heuristic considerations that lead to Eq. (2). Through this approach we get the explicit expression for the prediction error Z_t in Eq. (4). The orthogonality between the predictor X_t^{sm} and the prediction error Z_t is assured by result (a) of Theorem 2.2, i.e., it is not just an unverifiable model assumption, but a direct consequence of our specific prediction context. The orthogonality result (a) of Theorem 2.2 implies the optimal linear prediction result (b) of Theorem 2.2 which is much stronger than the non-formal optimality statement in Kraus (2014) that is based on the usual (unverifiable) orthogonality assumption on the prediction error and the restrictive Hilbert-Schmidt assumption.

3. Estimation and prediction. We rarely observe the functional trajectories directly, but typically only their noisy discretization points. In fact, the left panel of Figure 1 shows the *pre-smoothed* functions, however, the actual raw data is shown in Figure 2. In our estimation theory we take into account these additional discretization and measurement errors.

Let (Y_{it}, U_{it}) denote the observed noisy discretization pairs of a random functions X_t with

$$Y_{it} = X_t(U_{it}) + \varepsilon_{it}, \quad t \in \{1, \dots, T\}, \quad i \in \{1, \dots, n\}, \tag{10}$$

where $U_{it} \in \mathbb{R}$ is a stationary weakly dependent time series with finite second moments and ε_{it} a real i.i.d. random error term with mean zero, finite variance $\mathbb{V}(\varepsilon_{it}) = \sigma_\varepsilon^2$, and finite fourth moment. The random variables U_{it}

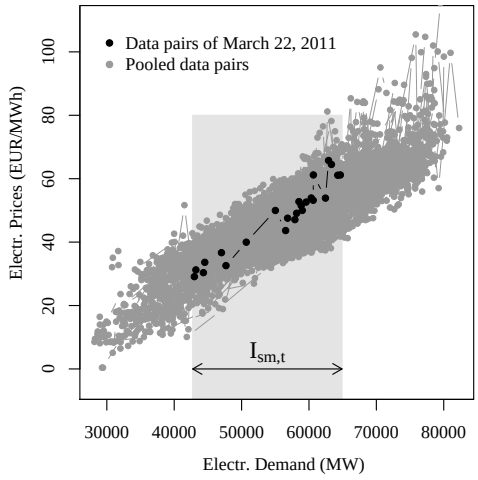

FIG 2. *Scatter plot of data pairs* (Y_{it}, U_{it}) *plus a visualization of the proper (random) subdomain* $[A_t, B_t] \subset [a, b]$ *of one randomly chosen day.*

and ε_{js} are assumed to be independent from each other and from the random functions X_k for all $i, j \in \{1, \ldots, n\}$ and all $s, t, k \in \{1, \ldots, T\}$.

In classical functional data scenarios it is assumed that the random function argument U_{it} is distributed as $U_0 \sim f_{U_0}$, where $f_{U_0}(u) > 0$ for all $u \in I_0$ and zero else. This applies to the case of sparse functional data (see, e.g., Yao, Müller and Wang, 2005a), but essentially also includes the classical case where each function can be pre-smoothed in a pre-processing step (see Ramsay and Silverman, 2005, Ch. 3). In order to model the case of partially observed functional data, we assume that U_{it} is distributed as $U_t \sim f_{U_t}$, where $f_{U_t}(u) > 0$ for all $u \in I_{\mathrm{sm},t}$ with $I_{\mathrm{sm},t} \subset I_0$ as specified in Assumption RS; see also Assumption A3 below.

Our empirical predictor is the truncated, empirical version of the definition of $\widetilde{X}_t^{\mathrm{la}}$ in Eq. (3):

$$\hat{\widetilde{X}}_{t,K}^{\mathrm{la}}(u) = \hat{\mu}(u; h_\mu) + \sum_{k=1}^{K} \frac{\hat{\xi}_{tk}^{\mathrm{sm}}}{\hat{\lambda}_k^{\mathrm{sm}}} \int_{I_{\mathrm{sm}}} \hat{\phi}_k^{\mathrm{sm}}(v) \hat{\gamma}^{\mathrm{la}}(u, v) dv, \tag{11}$$

where $\hat{\xi}_{kt}^{\mathrm{sm}}$ can be estimated, e.g., by the following integral approximation:

$$\hat{\xi}_{kt}^{\mathrm{sm}} = \sum_{i=1}^{n} \hat{\phi}_k^{\mathrm{sm}}(U_{it}^{\mathrm{sm}})(Y_{it} - \hat{\mu}(U_{it}^{\mathrm{sm}}; h_\mu))(U_{it}^{\mathrm{sm}} - U_{i-1,t}^{\mathrm{sm}}) \tag{12}$$

with $U_{0,t}^{\mathrm{sm}} = A_t$. The parameter K can be chosen, e.g., by the usual fraction-of-explained-variance approach.

This integral approximation, however, may have bad finite sample properties. The discretization of the empirical eigenfunctions at the design points U_{it} may lead to a lack of orthogonality between the empirical eigenfunctions. More stable and asymptotically equivalent to (12) is to estimate the FPC scores $(\xi_{1t}^{\mathrm{sm}}, \ldots, \xi_{Kt}^{\mathrm{sm}})^\top$ through a multiple linear regression of $Y_{it}^{\mathrm{sm}} - \hat{\mu}(U_{it}^{\mathrm{sm}}; h_\mu)$ on the K regressors $\hat{\phi}_1^{\mathrm{sm}}(U_{it}^{\mathrm{sm}}), \ldots, \hat{\phi}_K^{\mathrm{sm}}(U_{it}^{\mathrm{sm}})$, with $(Y_{it}^{\mathrm{sm}}, U_{it}^{\mathrm{sm}})$ denoting the noisy observations from the partially observed function X_t^{sm}; note that this regression must not contain a constant. We use this regression approach in our implementations.

For estimating μ and γ we use classical local linear estimators, where the below definitions are based on matrix notation as proposed, e.g., in Ruppert and Wand (1994). For given data (Y_{it}, U_{it}), with $t \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, n\}$, we estimate the mean function $\mu(u)$ by

$$\hat{\mu}(u; h_\mu) = e_1^\top \left([\mathbf{1}, \mathbf{U}_u]^\top \mathbf{W}_{\mu,u} [\mathbf{1}, \mathbf{U}_u] \right)^{-1} [\mathbf{1}, \mathbf{U}_u]^\top \mathbf{W}_{\mu,u} \mathbf{Y}, \tag{13}$$

where $e_1 = (1, 0)^\top$, $[\mathbf{1}, \mathbf{U}_u]$ is a $nT \times 2$ dimensional data matrix with typical rows $(1, U_{it} - u)$, the $nT \times nT$ dimensional diagonal weighting matrix $\mathbf{W}_{\mu,u}$ holds the kernel weights $K_{\mu,h}(U_{it} - u) = h_\mu^{-1} \kappa(h_\mu^{-1}(U_{it} - u))$. The kernel function κ is assumed to be a univariate, symmetric, pdf with compact support $\mathrm{supp}(\kappa) = [-1, 1]$, such as, e.g., the univariate Epanechnikov kernel (Assumption A7). The usual kernel constants are given by $\nu_2(\kappa) = \int v^2 \kappa(v) dv$, and $R(\kappa) = \int \kappa(v)^2 dv$. The objects $\mathbf{U}_u$ and $\mathbf{W}_{\mu,u}$ are filled in correspondence with the nT dimensional vector $\mathbf{Y} = (Y_{11}, Y_{21}, \ldots, Y_{n-1,T}, Y_{n,T})^\top$.

For given finite data $(\hat{C}_{ijt}, U_{it}, U_{jt})$, with $t \in \{1, \ldots, T\}$ and $i \neq j \in \{1, \ldots, n\}$, we estimate the covariance function $\gamma(u, v)$ by $\hat{\gamma}(u, v; h_\gamma) =$

$$= e_1^\top \left([\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v]^\top \mathbf{W}_{\gamma,u,v} [\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v] \right)^{-1} [\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v]^\top \mathbf{W}_{\gamma,u,v} \hat{\mathbf{C}}, \tag{14}$$

where $e_1 = (1, 0, 0)^\top$, $[\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v]$ is a $NT \times 3$ dimensional data matrix with typical rows $(1, U_{it} - u, U_{jt} - v)$, $N = (n^2 - n)$, the $NT \times NT$ dimensional diagonal weighting matrix $\mathbf{W}_{\gamma,u,v}$ holds the bivariate kernel weights $K_{\gamma,h}(U_{it} - u, U_{jt} - v)$. For the bivariate kernel weights $K_{\gamma,h}(z_1, z_2) = h_\gamma^{-2} \kappa_\gamma(z_1, z_2)$ we use a multiplicative kernel function $\kappa_\gamma(z_1, z_2) = \kappa(z_1)\kappa(z_2)$ with κ as defined above. The usual kernel constants are then $\nu_2(\kappa_\gamma) = (\nu_2(\kappa))^2$ and $R(\kappa_\gamma) = R(\kappa)^2$. The objects $[\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v]$ and $\mathbf{W}_{\gamma,u,v}$ are filled in correspon-

dence with the NT dimensional vector of empirical raw covariances with

$$\hat{\mathbf{C}} = (\hat{C}_{211}, \hat{C}_{311}, \ldots, \hat{C}_{n-2,n,T}, \hat{C}_{n-1,n,T})^\top$$
$$\hat{C}_{ijt} = (Y_{it} - \hat{\mu}(U_{it}))(Y_{jt} - \hat{\mu}(U_{jt})) \tag{15}$$

for all t and all $i \neq j$. Raw covariances $\hat{C}_{ijt}$ with $i = j$ need to be removed as these would introduce an estimation bias at the diagonal of γ through taking squares of the error term ε_{it} that is contained in Y_{it}. This removal explains the number of raw covariance points $(n^2 - n)T = NT$.

Estimates of the eigenvalues λ_k^{sm} and the eigenfunctions ϕ_k^{sm} are defined by the corresponding solutions of the empirical eigenequations:

$$\int_{I_{\mathrm{sm}}} \hat{\gamma}(u, v; h_\gamma)\hat{\phi}_k^{\mathrm{sm}}(v)\, dv = \hat{\lambda}_k^{\mathrm{sm}}\, \hat{\phi}_k^{\mathrm{sm}}(u), \text{ for } u \in I_{\mathrm{sm}}. \tag{16}$$

3.1. *Prediction algorithm.* In practice our prediction algorithm usually needs to be repeated. In the following we describe the first run:

1st Run: For each $t \in \{1, \ldots, T\}$, compute $\overset{\approx}{X}{}_{t,K}^{\mathrm{la}}$ according to Eq. (11) with $I_{\mathrm{sm},t} = [A_t, B_t]$, where $A_t = \min_{1 \leq i \leq n}(U_{it})$ and $B_t = \max_{1 \leq i \leq n}(U_{it})$.

An exemplary first run is visualized in Figure 3. The right panel of Figure 3 demonstrates how the resulting large intervals $I_{\mathrm{la},t}$ depend on $I_{\mathrm{sm},t}$ and the non-missing parts of the covariance function $\hat{\gamma}$; an uncovered plot of $\hat{\gamma}$ can be seen in the right panel of Figure 5.

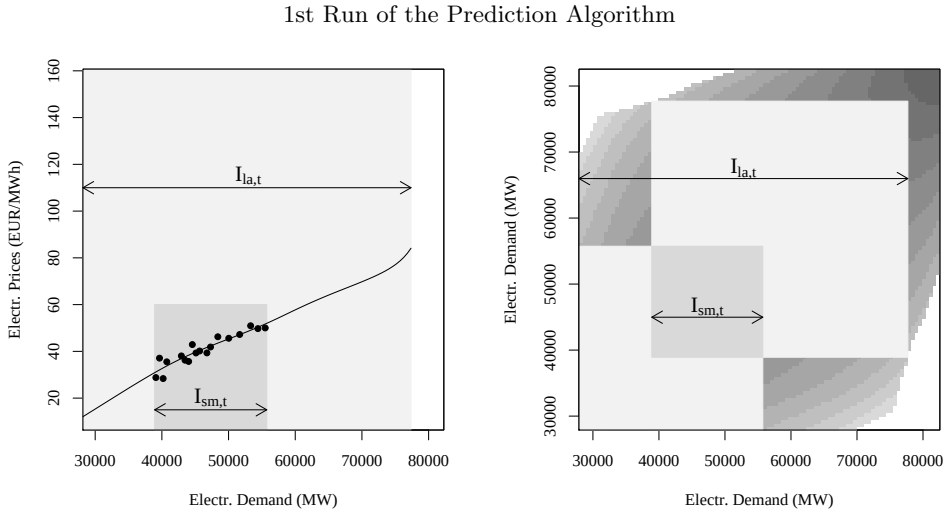

FIG 3. *Explanatory plots for the first run of the prediction algorithm.*

The first run of the prediction algorithm my not allow to predict all the missing fragments. For instance, the predicted price function in the left panel of Figure 3 still lacks the upper fragment for values of electricity demand $u \in [77362\ (\text{MW}), 82282\ (\text{MW})]$. In order to predict also the further missing fragments, we suggest to repeat the prediction algorithm. In the following we describe the ℓth, $\ell \geq 2$, run of the algorithm:

ℓth Run: For each $t \in \{1, \ldots, T\}$, compute $\hat{\widetilde{X}}_{t,K}^{\text{la}}$ according to Eq. (11), but with $I_{\text{sm},t} \subseteq \bigcup_{m<\ell} I_{\text{la},t}^{m\text{th Run}}$, such that:

1. A missing fragment can be predicted.
2. The width of $I_{\text{sm},t}$ is maximized.

That is, the positioning of $I_{\text{sm},t}$ needs to enable the prediction of missing fragments, but otherwise we need to maximize the width of $I_{\text{sm},t}$ in order to maximize the information contained in the "small" fragment.

An exemplary second run is visualized in Figure 4. There the new small interval $I_{\text{sm},t} \subseteq I_{\text{la},t}^{1\text{st Run}}$ is chosen such that the missing upper fragment can be predicted, but otherwise such that its width is maximized. The new large interval $I_{\text{la},t}$ contains the missing upper fragment, but also the redundant region $I_{\text{la},t} \cap I_{\text{la},t}^{1\text{st Run}}$. Therefore, we keep only the newly predicted fragment with respect to values $u \in [77362\ (\text{MW}), 82282\ (\text{MW})]$ and join it with the prediction from the first run.

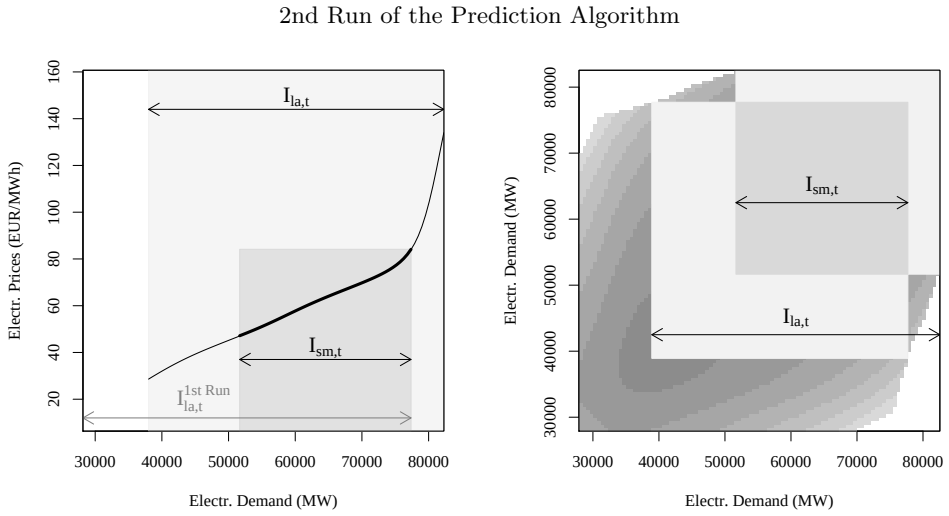

FIG 4. *Explanatory plots for the second run of the prediction algorithm.*

Two remarks on the ℓth run of the prediction algorithm. First, after each repetition, the newly predicted fragments need to be joined with the predicted fragments from the preceding runs. As long as the parameter K remains fix, the single fragments will be perfectly aligned.

Second, in the ℓth, $\ell \geq 2$, run of the prediction algorithm, we can make use of the predicted smooth function that is constructed from the preceding $(\ell - 1)$ runs, say, $\hat{\widetilde{X}}_t^{(\ell-1)}$. In particular, we can estimate the new "small" FPC scores by the classical integral method, i.e., by

$$\hat{\xi}_{kt}^{\mathrm{sm}} = \int_{I_{\mathrm{sm},t}} (\hat{\widetilde{X}}_t^{(\ell-1)}(u) - \hat{\mu}(u))\hat{\phi}_k^{\mathrm{sm}}(u)du.$$

3.2. *Checking for perfect predictability.* Result (c) of Theorem 2.2 tells us that the prediction error can be zero if the structure of the random function X_t is simple enough. It is straight forward to check empirically for perfect predictability, i.e., whether $X_t^{\mathrm{la}}(v) = \widetilde{X}_t^{\mathrm{la}}(v)$ for all $v \in I_{\mathrm{la}} \setminus I_{\mathrm{sm}}$.

As we do not observe (or cannot estimate) X_t^{la} we need to use the observable (or estimable) fragment X_t^{sm} as a "new large" fragment, i.e., $X_t^{\mathrm{new.la}} = X_t^{\mathrm{sm}}$ with $I_{\mathrm{new.la}} = I_{\mathrm{sm}}$. From this we can construct a "new small" fragment $X_t^{\mathrm{new.sm}}(u) = X_t^{\mathrm{new.la}}(u)$ with $u \in I_{\mathrm{new.sm}}$, where $I_{\mathrm{new.sm}} \subset I_{\mathrm{new.la}} = I_{\mathrm{sm}}$. This allows us then to compare the observable (or estimable) $X_t^{\mathrm{new.la}}$ with its prediction $\widetilde{X}_t^{\mathrm{new.la}} = S[X_t^{\mathrm{new.sm}}]$. If $X_t^{\mathrm{new.la}}(v) = \widetilde{X}_t^{\mathrm{new.la}}(v)$ for all $v \in I_{\mathrm{new.la}} \setminus I_{\mathrm{new.sm}}$, then there is no prediction error.

Of course, in practice, one has to compare the corresponding empirical versions $\hat{X}_t^{\mathrm{new.la}}$ and $\hat{\widetilde{X}}_t^{\mathrm{new.la}}$ with $I_{\mathrm{new.la}} = [A_t, B_t]$, and a reasonable amount of different "new small" intervals $I_{\mathrm{new.sm}} \subset [A_t, B_t]$. For this let $\hat{X}_t^{\mathrm{new.la}} = \hat{X}_t^{\mathrm{sm}}$, where

$$\hat{X}_t^{\mathrm{sm}}(u) = \hat{\mu}(u) + \sum_{k=1}^{K} \hat{\xi}_{kt}^{\mathrm{sm}} \hat{\phi}_k^{\mathrm{sm}}(u), \quad u \in [A_t, B_t],$$

is the "pre-smoothing" estimate which is a special case of Eq. (11) with $I_{\mathrm{sm}} = I_{\mathrm{la}} = [A_t, B_t]$ and $\hat{\xi}_{kt}^{\mathrm{sm}}$ computed by the multiple linear regression as explained in the first paragraph of Section 3.1. Furthermore, let

$$\hat{\widetilde{X}}_{t,K}^{\mathrm{new.la}}(u) = \hat{\mu}(u) + \sum_{k=1}^{K} \frac{\hat{\xi}_{tk}^{\mathrm{new.sm}}}{\hat{\lambda}_k^{\mathrm{new.sm}}} \int_{I_{\mathrm{new.sm}}} \hat{\phi}_k^{\mathrm{new.sm}}(v)\hat{\gamma}^{\mathrm{new.la}}(u, v)dv,$$

where $\hat{\lambda}_k^{\mathrm{new.sm}}$ and $\hat{\phi}_k^{\mathrm{new.sm}}$ are the kth eigenelements of the covariance op-

erator $\hat{\Gamma}^{\text{new.sm}}$ with covariance function $\hat{\gamma}^{\text{new.sm}}$, and where

$$\hat{\xi}_{kt}^{\text{new.sm}} = \int_{I_{\text{new.sm}}} (\hat{X}_t^{\text{new.la}}(u) - \hat{\mu}(u))\hat{\phi}_k^{\text{new.sm}}(u)du.$$

This way, $\hat{\tilde{X}}_t^{\text{new.la}}$ and $\hat{X}_t^{\text{new.la}}$ are affected by the estimation errors in exactly the same. That is, we can compare $\hat{\tilde{X}}_t^{\text{new.la}}$ and $\hat{X}_t^{\text{new.la}}$ with each other, conditionally on the common stochastics from the estimation errors, which are asymptotically negligible anyways. If $\hat{X}_t^{\text{new.la}}(v) = \hat{\tilde{X}}_t^{\text{new.la}}(v)$ for all $v \in I_{\text{new.la}} \setminus I_{\text{new.sm}}$, then there is no prediction error.

4. Asymptotic results. We refer to our model assumptions, as specified in the first paragraph of Section 2 and in the second paragraph of Section 3, as Assumption **A1**. Due to space limitations we list our further Assumptions A2-A9 in Appendix A of the supplemental paper. Our assumptions are generally close to those in Yao, Müller and Wang (2005b) and Hall, Müller and Wang (2006). Though, we additionally allow for weakly dependent time series of random function $(X_t)_t$ (Assumption A2), consider the case of proper (random) subsets (Assumption A3), and a more general asymptotic setup as specified by the following assumption:

A4 (Asymptotic Scenario) $Tn \to \infty$, where $n = n(T) \geq 2$ such that $n(T) \sim T^\theta$ with $0 \leq \theta < \infty$. Here "$a_T \sim b_T$" denotes that two sequences a_T and b_T are asymptotically equivalent up to some positive constant $0 < c < \infty$, i.e., that $\lim_{T\to\infty}(a_T/b_T) = c$.

Assumptions A5-A7 contain typical smoothness, bandwidth and kernel assumptions. In order to make use of the eigenvalue and eigenfunction expansion of Hall and Hosseini-Nasab (2006), we impose the regularity Assumptions A8 and A9.

Remarks on Assumption A4. For the estimation of the mean and covariance functions, the eigenvalues, and eigenfunctions we allow also for the case of very sparsely sampled prediction points U_{it}, i.e., with $\theta = 0$ corresponding to, e.g., $n \leq 5$. However, for estimating the pc-scores, we only focus on the case of $\theta > 0$, since otherwise the pc-scores cannot be consistently estimated. Under truly sparse sampling schemes one should use the PACE procedure of Yao, Müller and Wang (2005a) and Yao, Müller and Wang (2005b) to get approximations for the pc-scores. We follow Hall, Müller and Wang (2006) and consider n as deterministic, though, under some minor modifications n

can be assumed to be random. For instance, our results directly apply to the case with n being replaced by N_T, where N_T is an independent random variable defined as $N_T = n(T) + W_T$, where $n(T)$ is deterministic function of T with $n(T) \sim T^\theta$ and the random variable W_T has realizations in some appropriate subset of $\mathbb{N}$, such that $W_T \geq 2$ wp1 and $\mathbb{E}(N_T) = n(T)$ for all T as $T \to \infty$.

THEOREM 4.1 (Preliminary uniform consistency results).
Under Assumptions A1-A7 we have the following weak consistency results:

(a) Estimator of the mean function:

$$\sup_{u\in I_0} |\hat{\mu}(u; h_\mu) - \mu(u)| = \mathcal{O}_p\left(h_\mu^2 + \frac{1}{\sqrt{Tn\, h_\mu}} + \frac{1}{\sqrt{T}}\right). \tag{17}$$

(b) Estimator of the covariance function:

$$\sup_{(u,v)\in I_0^2} |\hat{\gamma}(u, v; h_\gamma) - \gamma(u, v)| = \mathcal{O}_p\left(h_\gamma^2 + \frac{1}{\sqrt{TN\, h_\gamma^2}} + \frac{1}{\sqrt{T}}\right). \tag{18}$$

The following results (c) and (d) are valid under the additional Assumption A9, our identification restriction in Section 2.2, i.e., that $\lambda_k > 0$ for all $k \geq 1$, and under the assumption that all eigenvalues $\{\lambda_k\}_{k\geq 1}$ are of multiplicity one. Furthermore, we assume that $\hat{\phi}_k$ is chosen such that $\langle\hat{\phi}_k, \phi_k\rangle_H > 0$.

(c) Estimators of the eigenvalues:

$$\sup_{k\geq 1} |\hat{\lambda}_k - \lambda_k| = \mathcal{O}_p\left(h_\gamma^2 + \frac{1}{\sqrt{TN\, h_\gamma^2}} + \frac{1}{\sqrt{T}}\right). \tag{19}$$

(d) Estimators of the eigenfunctions:

$$\sup_{u\in I_0} |\hat{\phi}_k(u) - \phi_k(u)| = \mathcal{O}_p\left(\frac{1}{\delta_k}\left(h_\gamma^2 + \frac{1}{\sqrt{TN\, h_\gamma^2}} + \frac{1}{\sqrt{T}}\right)\right), \tag{20}$$

for all $1 \leq k \leq \bar{K}_{TN} - 1$ with $\bar{K}_{TN} = \inf\{k \geq 1 : \lambda_k - \lambda_{k+1} \leq 2\hat{\Delta}_{TN}\}$, where $\hat{\Delta}_{TN} = (\int_{(u,v)\in I_0^2}(\hat{\gamma}(u, v) - \gamma(u, v))^2 d(u, v))^{1/2}$, with $\hat{\Delta}_{TN} = \mathcal{O}_p(h_\gamma^2 + (TN\, h_\gamma^2)^{-1/2} + T^{-1/2})$, and where $\delta_k = \min_{1\leq i\leq k}\{\lambda_i - \lambda_{i+1}\}$.

The proofs can be found in Appendix B of the supplemental paper.

Remarks on Theorem 4.1. By contrast to Yao, Müller and Wang (2005b) our rates of the estimators $\hat{\mu}$ and $\hat{\gamma}$ are optimal. Note that it is generally advisable use an under-smoothed mean function to construct the raw covariance points $\hat{C}_{ijt}$ for estimating the covariance function.

THEOREM 4.2 (Main uniform consistency results).
Under Assumptions A1-A9 and our identification restriction in Section 2.2 the following results (a)-(c) hold for $1 \leq K \leq \bar{K}_{TN} - 1$, $a > 1$, *and for optimal bandwidth choices* $h_\mu \sim (Tn)^{-1/5}$ *and* $h_\gamma \sim (TN)^{-1/6}$*:*

(a) For $0 < \theta < 1/4$ *(*$\approx T/n$ *large)*

$$\sup_{u \in I_0} \left| \widetilde{X}_t^{\text{la}}(u) - \hat{\widetilde{X}}_{t,K}^{\text{la}}(u) \right| = \mathcal{O}_p \left(K\, n^{-1/2} + (TN)^{-1/3} K^{a+2} + \sum_{k=K+1}^{\infty} k^{-a/2} \right).$$

(b) For $1/4 \leq \theta < 1$ *(*$\approx T/n$ *moderate)*

$$\sup_{u \in I_0} \left| \widetilde{X}_t^{\text{la}}(u) - \hat{\widetilde{X}}_{t,K}^{\text{la}}(u) \right| = \mathcal{O}_p \left(K\, n^{-1/2} + T^{-1/2} K^{a+2} + \sum_{k=K+1}^{\infty} k^{-a/2} \right).$$

(c) For $1 \leq \theta < \infty$ *(*$\approx T/n$ *small)*

$$\sup_{u \in I_0} \left| \widetilde{X}_t^{\text{la}}(u) - \hat{\widetilde{X}}_{t,K}^{\text{la}}(u) \right| = \mathcal{O}_p \left(T^{-1/2} K^{a+2} + \sum_{k=K+1}^{\infty} k^{-a/2} \right).$$

The proofs can be found in Appendix B of the supplemental paper.

Remarks on Theorem 4.2. The $\mathcal{O}_p(\sum_{k=K+1}^{\infty} k^{-a/2})$ terms quantify the cut-off regularization error, while the other terms quantify the different estimation errors. Eq. (50) in Appendix B of the supplemental paper contains a more general consistency result without imposing the eigenvalue Assumption A8 and without assuming optimal bandwidth choices.

5. Simulation study. For our simulation study we generate $T = 200$ many iid normal and $T = 200$ many iid exponential random functions $X_t(u) = \mu(u) + \xi_{1,t}\phi_1(u) + \xi_{2,t}\phi_2(u)$, where $\mu(u) = u + \sin(u)$, $\phi_1(u) = -\cos((\pi u)/(b-a))/\sqrt{5}$, $\phi_2(u) = -\cos(2(\pi u)/(b-a))/\sqrt{5}$, $a = 1$, and $b = 10$. For the normal case we let $\xi_{k,t} \sim N(0, \lambda_k)$ with $\lambda_1 = 4$ and $\lambda_2 = 3$. For the exponential case we let $\xi_{k,t}$ be a centered exponential random variable with rate λ_k and centered by λ_k^{-1}. Furthermore, $U_{1t}, \ldots U_{nt} \overset{\text{iid}}{\sim} \text{Unif}[A_t, B_t]$, where $A_1, \ldots, A_t \overset{\text{iid}}{\sim} \text{Unif}[a, a + (b-a) \cdot 0.25]$ and $B_t = A_t + (b-a) \cdot 0.75$. That is,

the discretization points of a function X_t cover at most three-quarter of the total domain $[a, b]$. Finally, the observations Y_{it} are generated according to $Y_{it} = X_t(U_{it}) + \varepsilon_{it}$ with $\varepsilon_{it} \sim N(0, 0.2)$.

The prediction procedure is implemented using the ambitious strategy as described in Section 3.1, with

$$I^{\text{up}}_{\text{sm},t} = [U_{(\lfloor n/2 \rfloor)t}, U_{(n)t}] \quad \text{and} \quad I^{\text{lo}}_{\text{sm},t} = [U_{(1)t}, U_{(\lfloor n/2 \rfloor+1)t}].$$

The parameter K is selected by the Fraction of Variance Explained (FVE) criterion with FVE $= 0.95$, i.e., the value of K is the highest K such that $(\sum_{l=1}^{K} \hat{\lambda}^{\text{la}}_l)/(\sum_{j\geq 1} \hat{\lambda}^{\text{la}}_j) \leq \text{FVE}$. In each of in total 100 simulation runs we compute the "inner" mean absolute error as

$$\text{MAE}_{\text{inner}} = T^{-1} \sum_{t=1}^{T} \max_{u \in I_{\text{sm},t}} |\hat{\tilde{X}}_{t,K}(u) - X_t(u)|$$

with $I_{\text{sm},t} = [A_t, B_t]$, $A_t = \min(U_{1t}, \ldots, U_{nt})$, $B_t = \max(U_{1t}, \ldots, U_{nt})$, and the "outer" MAE as

$$\text{MAE}_{\text{outer}} = T^{-1} \sum_{t=1}^{T} \max_{u \in I_{\text{la},t} \setminus I_{\text{sm},t}} |\hat{\tilde{X}}_{t,K}(u) - X_t(u)|,$$

where $I_{\text{la},t}$ denotes the interval over which $\hat{\tilde{X}}_{t,K}$ is recovered. The averages of the 150 MAE values for the different cases are reported in Table 1. In each run the percentage of functions that are recovered over the total domain, i.e., for which $I_{\text{la},t} = [a, b]$, is at least 99%.

		$n = 15$	$n = 25$	$n = 50$	$n = 75$
Normal	Avg.-$\text{MAE}_{\text{inner}}$	0.34	0.31	0.27	0.23
	Avg. $\text{MAE}_{\text{outer}}$	0.65	0.57	0.50	0.44
Exponential	Avg.-$\text{MAE}_{\text{inner}}$	0.35	0.31	0.28	0.24
	Avg.-$\text{MAE}_{\text{outer}}$	0.71	0.65	0.55	0.46

TABLE 1
Averages over 100 Monte Carlo simulated MAE values.

Under the normal data generating process, we do not expect any prediction errors – only estimation errors. This follows from our result (c) of Theorem 2.2. In fact, if $Y_{it} = X_t(U_{it})$ and if we use the true mean and the true covariance functions, i.e., no measurement errors and no sampling error, then all figures in Table 1 drop to zero. In order to check for this prefect predictability property under the influence of measurement and sampling errors we use the procedure described in Section 3.2.

6. Application.

Data Sources. The data for our analysis come from four different sources. Hourly spot prices of the German electricity market are provided by the European Energy Power Exchange (EPEX) (www.epexspot.com), hourly values of Germany's gross electricity demand and electricity exchanges with other countries are provided by the European Network of Transmission System Operators for Electricity (www.entsoe.eu), German wind and solar power infeed data are provided by the transparency platform of the European energy exchange (www.eex-transparency.com), and German air temperature data are available from the German Weather Service (www.dwd.de). The data dimensions are given by $n = 24$ hours and $T = 241$ working days between March 15, 2012 and March 14, 2013. Very few (0.26%) of the data pairs (Y_{it}, U_{it}) with prices $Y_{it} > 120$ EUR/MWh and $U_{it} > 82000$ MW are considered as outliers and therefore removed.

The estimated mean function and the estimated covariance function are shown in Figure 5. The outer off-diagonal parts of the covariance function γ cannot be estimated, since these parts of the domain are not covered by data pairs (U_{it}, U_{jt}); see also the right panel in Figure 2.

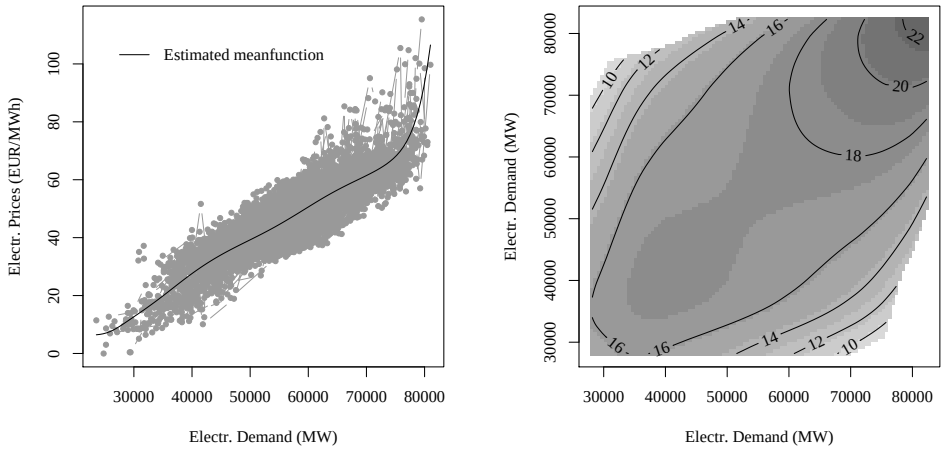

FIG 5. LEFT PANEL: *Estimated mean function plus scatter plot of the data pairs* (Y_{it}, U_{it}). RIGHT PANEL: *Contour plot of the estimated covariance function. The white regions reflect the outer off-diagonal parts which are infeasible to estimate.*

For predicting the missing parts of each function X_t we use the estimation strategy as described in Section 3.1. In a first run, we predict for each random function the lower and the upper fragments by using the information with respect to the actually observed fragments as visualized in Figure 3. In a

second run, we use the enlarged function (or a part of it) as the new observed fragment in order to recover further missing fragments as visualized in Figure 4. After this second run, over 90% of the price functions are recovered over their total support; see Figure 6. The recovered functions could now be used to re-estimate the covariance function over its total support. Given the re-estimated covariance function, we could then to recover also the rest of the functions over their total support.

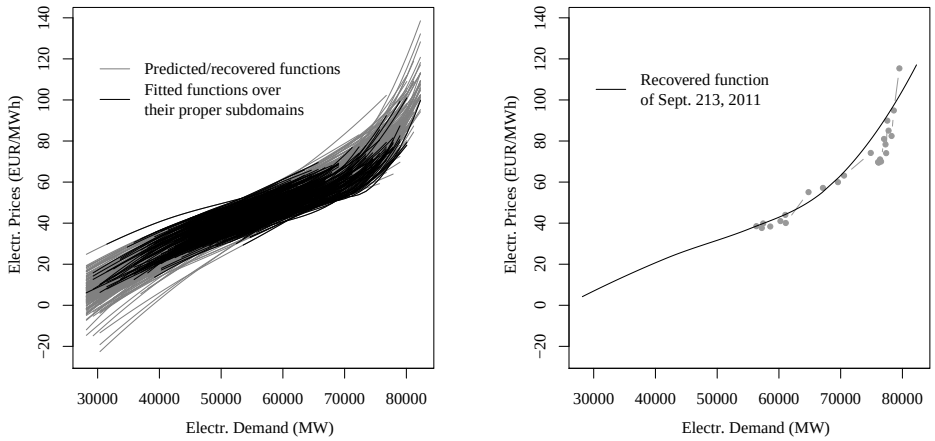

FIG 6. LEFT PANEL: RIGHT PANEL:.

7. Conclusion. We propose a new prediction model that allows to predict the missing fragments of partially observed functional data from the observed fragments. By contrast to the alternative approaches in the literature, we consider the most challenging case of partially observed functional data for which it is impossible to estimate the covariance function over its total support. Our estimation algorithm allows to predict the missing parts of the random functions under this very challenging setup.

We show that our predictor is the best linear predictor – a property that is also claimed by the alternative functional predictors in the literature. Though, by contrast to the results in the literature, our optimality result is not based on the usual, but unverifiable, model assumption of orthogonality between the functional predictor and the functional prediction error. We can do without this assumption, by making use of the very specific nature of our prediction problem: predicting the missing parts of a random functions from its own observed parts.

In our estimation theory, we focus on the relevant practical situation in

which we do not directly observe the fragments, but only their noisy discretization points. This situation involve the use of nonparametric estimation procedures and we derive the uniform rates of consistency for our nonparametric estimation procedures. The finite sample properties are investigated through simulations and a real data application.

References.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications* **149**. Springer Verlag.

Crambes, C. and Mas, A. (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli* **19** 2627–2651.

Delaigle, A. and Hall, P. (2013). Classification using censored functional data. *Journal of the American Statistical Association* **108** 1269–1283.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, 1. ed. *Springer Series in Statistics*. Springer.

Goldberg, Y., Ritov, Y. and Mandelbaum, A. (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference* **147** 53–65.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 109–126.

Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34** 1493–1517.

He, G., Müller, H. G. and Wang, J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis* **85** 54–77.

Hirth, L. (2013). The market value of variable renewables: The effect of solar wind power variability on their relative price. *Energy economics* **38** 218–236.

Hörmann, S. and Kidziński, Ł. (2015). A note on estimation in Hilbertian linear models. *Scandinavian journal of statistics* **42** 43–62.

James, G. M., Hastie, T. J. and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602.

James, G. M. and Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63** 533–550.

Kraus, D. (2014). Partially observed functional data. *Contributions in infinite-dimensional statistics and related topics* 179–184.

Nicolosi, M. (2010). Wind power integration and power system flexibility – An empirical analysis of extreme events in Germany under the new negative price regime. *Energy Policy* **38** 7257–7268.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2. ed. *Springer Series in Statistics*. Springer.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of statistics* 1346–1370.

Tsybakov, A. (2008). *Introduction to Nonparametric Estimation*, 1. ed. *Springer Series in Statistics*. Springer.

Weigt, H. (2009). Germany's wind energy: The potential for fossil capacity replacement and cost saving. *Applied Energy* **86** 1857–1863.

Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100** 577–590.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33** 2873–2903.

Dominik Liebl and Alois Kneip

Statistische Abteilung

University of Bonn

Adenauerallee 24-26

53113 Bonn, Germany

E-mail: dliebl@uni-bonn.de