ROYAL
STATISTICAL
SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Statistics Society
Series **B**
Statistical Methodology

B

Original Article

Ordinary differential equation models for a collection of discretized functions

Lingxuan Shao[1] and Fang Yao[2] (ID)

[1]Department of Statistics and Data Science, School of Management, Fudan University, Shanghai, China
[2]Department of Probability and Statistics, School of Mathematical Sciences, Center for Statistical Science, Peking University, Beijing, China

Address for correspondence: Fang Yao, Department of Probability and Statistics, School of Mathematical Sciences, Center for Statistical Science, Peking University, Yiheyuan Road 5, Haidian District, Beijing 100871, China. Email: fyao@math.pku.edu.cn

Abstract

The exploration of dynamic systems governed by ordinary differential equations (ODEs) holds great interest in the field of statistics. Existing research mainly focuses on a single function. This study generalizes the scope to analyse a collection of functions observed at discretized times, with sampling frequencies varying from sparse to dense designs. The range of ODE models studied caters to diverse dynamic systems, and includes the complex nonlinear and non-Lipschitz scenarios. We introduce a new concept named functional moment method, a novel approach for parameter estimation within these ODE models and facilitating the recovery of curves for the discretely observed functions. Our numerical analysis underscores the method's applicability across various application fields, including sociology, physics, and epidemiology.

Keywords: curve recovery, functional moment method, functional principal component analysis, parameter detection

1 Introduction

Ordinary differential equations (ODEs) are important in representing complex dynamic systems across a wide range of scientific and technological areas, including fields like physics (Riley et al., 1999), social sciences (Black & Scholes, 1973), biology (Leah, 2005), and engineering (Marion & Temam, 1998). While ODE models can be formulated based on expert knowledge of specific dynamic processes, their inherent parameters often remain unknown. The task of identifying these parameters and recovering dynamic systems from discretely observed and noisy data poses significant statistical challenges, owing to the noised observations and the lack of analytical solutions to the majority of ODEs.

The existing statistical literature mainly focuses on ODEs represented as $F_\beta(V) = 0$. This form is defined by a differential operator F_β with potentially an unidentified parameter β associated with a specific process V. The core objectives are estimating this unknown parameter β and reconstructing the trajectory V based on discrete observations $\{(V(t_j), t_j)\}_{1 \le j \le m}$. The seminal work (Ramsay et al., 2007) addressed parameter estimation using the *Parameter Cascading* method, which was then broadened to the first-order representable ODEs with random effects (Wang et al., 2014), where 'representable ODE' means the expressions like $V'(t) = G(V, t; \beta)$. Subsequent developments led to frameworks tackling time-varying coefficient ODEs (Xue et al., 2010), integrable ODEs (Hall & Ma, 2014; Tan et al., 2024), ODE systems (Dai & Li, 2022), and sparse additive ODEs (N. Zhang et al., 2022). A common strategy across these studies is to select a loss function that balances data fitting discrepancies with ODE conformity, which we have adopted in a similar spirit.

In contrast to existing research that focused on analysing a single function, we aim to explore a unified framework for a collection of discretized functions driven by an underlying mechanism. Specifically, the dataset comprises samples of discretely observed trajectories, accommodating sampling frequencies ranging from sparse to dense designs. This is common across multiple scientific fields. Examples include population growth counts in sociology, particle movement studies in physics (Landau & Lifshitz, 1976), and COVID-19 virus spreading patterns in epidemiology (Ma et al., 2009), among others. In the aspect of the driven mechanism, our study concentrates on the differential operator F_β, a multivariate polynomial of the function X and its derivatives X, $X^{(1)}$, $X^{(2)}$, …. This operator form not only addresses linear, first-order or integrable ODEs but also extends its applicability to a broader spectrum, including nonlinear, higher-order, non-Lipschitz, nonintegrable, and nonrepresentable ODEs. This versatility enhances its adaptability to the diverse requirements of various fields.

In our framework, the first goal is to estimate the unknown parameter β in the differential operator F_β, utilizing discretely observed functions. The second goal is to recover the trajectories of realizations conforming to the ODE model $F_\beta(X) = 0$ over the entire time interval. This recovery is based on the parameter estimates from the first objective, guided by the aim of minimizing discrepancies in data fitting and ensuring adherence to the ODE. To realize these objectives, we introduce a novel concept termed *functional moment*, which facilitates the adoption of a unified approach, the pooling strategy (Li & Hsing, 2010; Yao et al., 2005), and achieves combining observed data from the entire collection of functions to handle sampling frequencies ranging from sparse to dense designs. This approach contrasts with traditional methods of individual curve smoothing that require sufficiently dense measurements per function (Ramsay et al., 2007; Xun et al., 2013; N. Zhang et al., 2022). The functional moment also allows us to estimate the unknown parameter by solving linear equations, similar to ordinary linear regression, which avoids dealing with complicated differential operators and is thus applicable to a wide range of ODE models. However, traditional methods (Wang et al., 2014; Xue et al., 2010) are bogged down in numerically solving ODEs and are limited to a narrow scope. Additionally, our theoretical analysis reveals phase transition phenomena in the convergence rates for parameter estimation and trajectory recovery that can offer insights for the design of sampling strategies.

The structure of this paper is organized as follows. Section 2 presents the ODE models based on the underlying mechanism. Section 3 delves into the proposed functional moment method (FMM) for parameter estimation and recovery from a collection of discretized functions, with theoretical guarantees in Section 4. Sections 5 and 6 showcase numerical results from simulations and real data analyses, respectively, affirming the efficacy of our proposed method. The extensions to models infused with time-varying coefficients and mechanism derived from partial differential equations (PDEs) are presented in Section 7. The online supplementary material encompasses the additional simulations and technical proofs.

2 ODE models

The underlying population is modelled as a random process $X : \Omega \to H^{v_0}(\mathcal{T}) \subset L^2(\mathcal{T})$. Here $H^{v_0}(\mathcal{T})$ refers to the Sobolev space consisting of functions with L^2 integrable v_0th-order derivative on the compact domain $\mathcal{T}$. This random process X is supposed to be subject to the constraint imposed by a v_0th-order differential operator F_β, that is, $F_\beta(X) = 0$. This constraint can be equivalently stated as $P(X \in V_\beta) = 1$, where $V_\beta = \{f \in H^{v_0} : F_\beta(f) = 0\} \subset H^{v_0}$ is the solution space.

We focus on a general class of differential operators that act on the population X as multivariate polynomials of the population X and its derivatives, X', X'', …. This general form includes the often-encountered ODEs, as well as those examined in existing research (Dai & Li, 2022; Ramsay et al., 2007; Wang et al., 2014; Xue et al., 2010; N. Zhang et al., 2022). The structure of these operators is described by given parameters $\{g_k\}_{k=0}^p$ and $\{\alpha_k\}_{k=0}^p$, and the unknown coefficient parameters $\{\beta_k\}_{k=0}^p$. In particular, the ODE model takes the following form:

$$F_\beta(X) = \sum_{k=0}^p \beta_k X^{(\alpha_k)}(t) g_k(t) = 0, \tag{1}$$

where the multi-index $X^{(\alpha_k)}$ of a vector $\alpha_k = (\alpha_{k,[1]}, \alpha_{k,[2]}, \ldots, \alpha_{k,[N_k]})$ means $X^{(\alpha_k)} = X^{(\alpha_{k,[1]})} \times X^{(\alpha_{k,[2]})} \times \cdots \times X^{(\alpha_{k,[N_k]})}$ and $X^{(\alpha_{k,[l]})}$ denotes the $\alpha_{k,[l]}$th-order derivative of X. For each vector α_k, the norms $\|\alpha_k\|_0 = N_k$, $\|\alpha_k\|_1 = \sum_{1 \le l \le N_k} \alpha_{k,[l]}$, $\|\alpha_k\|_\infty = \max_{1 \le l \le N_k} \alpha_{k,[l]}$ represent the dimension, the total order of derivatives and the maximum order of derivatives, respectively. For example, $X^2 X' X''$ is expressed as $X^{(0,0,1,2)}$ with norms $\|(0, 0, 1, 2)\|_0 = 4$, $\|(0, 0, 1, 2)\|_1 = 3$ and $\|(0, 0, 1, 2)\|_\infty = 2$. In the sequel, we assume, without loss of generality, $\beta_0 = 1$ for identifiability and the time interval $\mathcal{T} = [0, 1]$.

The above model structure (1) encompasses a wide range of ODEs that have garnered great interest and applications in statistics. Below, we present examples of linear, nonlinear, nonrepresentable, non-Lipschitz, and higher-order ODEs.

Example 1   (POPULATION MODEL). The basic population model detailing the count of individuals within a specific region is the Malthusian population model. This model is based on the tenet that the growth of a population is linearly proportional to the existing size of the population. Such relationship is expressed through the equation $N' = rN$, in which N signifies the total number of individuals in the population, while r symbolizes the rate of population growth. Subsequently, this growth function is reformulated in the following manner:

$$X^{(1)} + \beta X^{(0)} = 0, \tag{2}$$

where the term $X^{(1)}$ denotes the first-order derivative, and β represents the unknown parameter to estimate.

Example 2   (STABLE FIELD MODEL). The trajectory of the rigid particle in a stable field is depicted by the nonlinear energy conservation equation given as $m(x')^2/2 + Vx = E$ (Landau & Lifshitz, 1976), wherein x signifies the position of the particle as a function of time t, m represents the mass, E corresponds to the total energy, and Vx constitutes the linear potential energy induced by the stable field. This ODE can be reformulated as the subsequent expression:

$$X^{(1,1)} + \beta X^{(0)} + C = 0, \tag{3}$$

where the term $X^{(1,1)}$ denotes the square of the first-order derivative $(X')^2$, C is a constant indicating the total energy, and β represents the unknown parameter to estimate. This is also a non-Lipschitz and nonpresentable ODE that shares no closed expression for the derivative $X^{(1)}$.

Example 3   (DAMPED OSCILLATION MODEL). The vibrational mode in damped oscillations can be represented using the equation of motion given by $x'' + \lambda x' + kx = 0$ (Landau & Lifshitz, 1976), where x denotes the time-dependent particle position, λ represents the damping coefficient, and k stands for the elasticity constant. This ODE can be reformulated as follows:

$$X^{(2)} + \beta X^{(1)} + C X^{(0)} = 0, \tag{4}$$

where the term $X^{(2)}$ denotes the second-order derivative, C relates to the elasticity constant, and β is the unknown parameter to estimate.

Example 4   (EPIDEMIC DISEASE MODEL). The epidemic spreading pattern is captured by SIRD model (Ma et al., 2009), where the term 'SIRD' abbreviates for four categories: susceptible, infected, cured, and dead. They are described by the ODEs: $I'(t) = \beta I(t)$, $R'(t) = \gamma I(t)$, $D'(t) = \rho I(t)$, where the unknown parameters β, γ, and ρ correspond to the rates of diffusion, curing, and death, respectively. These ODEs formulated in our terms are

$$I^{(1)} = \beta I, \quad R^{(1)} = \gamma I, \quad D^{(1)} = \rho I. \tag{5}$$

Remark 1   We note that the differential operator F_β considered in this paper includes non-Lipschitz ODEs for which the initial conditions do not uniquely determine the solution. For example, consider the continuous curve satisfying the ODE in Example 2: $F(X) = (X')^2 - 4X = 0$ with the initial condition $X(0) \equiv 0$. The random process could be $X = (t - A)^2 I_{t \geq A}$ for any $0 \leq A \leq 1$. As a result, the solution space V_0 cannot be reduced to the space of initial values, and this non-Lipschitz ODE is neither integrable nor representable. To the best of our knowledge, the polynomial structure considered in this paper extends the scope of existing research by addressing this challenging case, which has previously been overlooked.

3 Estimation

In our statistical models, the random trajectories $\{X_i\}_{i=1}^n$ are drawn from the underlying population X and satisfy the ODE (1) in the sense that $F_\beta(X_i) = 0$ holds for all $1 \leq i \leq n$. These realizations are usually assumed independent and identically distributed, and are observed intermittently at random time points $\{T_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq m_i}$. Thus, the samples can be represented as $\{X_{ij} = X_i(T_{ij}) + \varepsilon_{ij}\}_{1 \leq i \leq n, 1 \leq j \leq m_i}$, where ε_{ij} denotes noise and the number of measurements, m_i, represents the sampling frequency. In the sequel, we assume $m_1 = \cdots = m_n = m$ for a clear exposition; extension to more general cases is technically straightforward (X. Zhang & Wang, 2016). The aim of this section is to establish a unified framework for estimating the scalar parameters $\{\beta_k\}_{k=1}^p$ and recovering the realizations $\{X_i\}_{i=1}^n$ from discrete observations $\{(X_{ij}, T_{ij})\}_{1 \leq i \leq n, 1 \leq j \leq m}$, accommodating sampling frequencies from sparse to dense designs.

3.1 Estimating based on FMM

To better capture the differential structure and implement the pooling strategy, we generalize the concept of moment for a scalar random variable to a random process. Specifically, we introduce the concept of the *functional moment*, which is defined as the expectation

$$\mathbb{E}\, X^{(\alpha)}(t) = \mathbb{E}\left\{ X^{(\alpha_{[1]})}(t) X^{(\alpha_{[2]})}(t) \cdots X^{(\alpha_{[N]})}(t) \right\}$$

of a multivariate monomial $X^{(\alpha)}$ and the multi-index vector $\alpha = (\alpha_{[1]}, \alpha_{[2]}, \ldots, \alpha_{[N]})$. For instance, the functional moments associated with Examples 1–3 are $\mathbb{E}\, X^{(1)}(t) = \mathbb{E}\{X'(t)\}$, $\mathbb{E}\, X^{(1,1)}(t) = \mathbb{E}\{X'(t)^2\}$, and $\mathbb{E}\, X^{(2)}(t) = \mathbb{E}\{X''(t)\}$, respectively. Furthermore, when $\alpha_{[1]} = \alpha_{[2]} = \cdots = \alpha_{[N]} = 0$, the functional moment $\mathbb{E}\, X^{(\alpha)}(t)$ simplifies to the Nth moment of $X(t)$, i.e. $\mathbb{E}\{X(t)^N\}$.

Utilizing the functional moment, the structure of the ODE model can be explored through the following *functional estimating equation* (FEE):

$$\sum_{k=0}^p \beta_k \mathbb{E}\, X^{(\alpha_k)}(t) g_k(t) = 0, \tag{6}$$

which is obtained by taking the expectation on (1). Now we propose the *FMM* to realize parameter estimation, which entails first approximating the functional moments $\mathbb{E}\, X^{(\alpha_k)}(t)$ $(0 \leq k \leq p)$ and

then deriving estimates for the unknown parameters $\{\beta_k\}_{k=1}^{p}$. It effectively bridges the gap between individual curves and population characteristics within the same ODE framework $F_\beta(X) = 0$. Consequently, this allows us to examine the ODE at the population level instead of relying on some specific realization.

We utilize the local polynomial regression technique to estimate these functional moments $\mathbb{E}\, X^{(\alpha_k)}(t)$ ($0 \leq k \leq p$), based on the observations $\{(X_{ij}, T_{ij})\}_{1 \leq i \leq n, 1 \leq j \leq m}$. By pooling data from all trajectories, these functional moments can be well estimated when there are only a few measurements from some or even all functions as long as the pooled data are sufficient. Specifically, for any vector $\alpha = (\alpha_{[1]}, \ldots, \alpha_{[N]})$, the moment $\mathbb{E}\, X^{(\alpha)}(t)$ can be equivalently expressed as

$$\mathbb{E}\, X^{(\alpha)}(t) = \partial^{\alpha} C(s_1, s_2, \ldots, s_N)\Big|_{s_1 = t, s_2 = t, \ldots, s_N = t}$$

with $C(s_1, s_2, \ldots, s_N) = \mathbb{E}\, X(s_1)\ldots X(s_N)$ and $\partial^{\alpha} = \partial_{s_1}^{\alpha_{[1]}} \cdots \partial_{s_N}^{\alpha_{[N]}}$. Then the local qth-order polynomial regression (Fan & Gijbels, 1996; Masry, 1997) is utilized to estimate the $\|\alpha\|_1$th-order derivative of $C(s_1, s_2, \ldots, s_N)$

$$\widehat{\theta} = \arg\min_{\theta} \sum_{1 \leq i \leq n, 1 \leq j_1, \ldots, j_N \leq m} K_b(T_{ij_1} - s_1) \times \cdots \times K_b(T_{ij_N} - s_N)$$

$$\times \Bigg\{ X_{ij_1} X_{ij_2} \cdots X_{ij_N} - \theta_0 - \sum_{1 \leq l \leq N} \theta_l (T_{ij_l} - s_l) - \cdots$$

$$- \frac{1}{q!} \sum_{1 \leq l_1, l_2, \ldots, l_q \leq N} \theta_{l_1, l_2, \ldots, l_q} (T_{ij_{l_1}} - s_{l_1})(T_{ij_{l_2}} - s_{l_2}) \cdots (T_{ij_{l_q}} - s_{l_q}) \Bigg\}^2,$$

where $K_b(\cdot) = \frac{1}{b} K(\frac{\cdot}{b})$ and $K(\cdot)$ is a kernel (symmetric and nonnegative density) function, and the indexes $j_1, \ldots, j_N$ differ from each other in the summation above to eliminate the influences from noise.

The estimate of partial derivative $\widehat{\partial^{\alpha} C}$ for $\alpha = (\alpha_{[1]}, \ldots, \alpha_{[N]})$ is set as

$$\widehat{\partial^{\alpha} C}(s_1, s_2, \ldots, s_N) = \widehat{\theta}_{l_1, l_2, \ldots, l_{\|\alpha\|_1}} \text{ with } l_k \text{ defined as } \alpha_{[l]} = \#\{1 \leq k \leq \|\alpha\|_1 : l_k = l\}$$

and subsequently the functional moment estimate is

$$\widehat{\mathbb{E}}\, X^{(\alpha)}(t) = \widehat{\partial^{\alpha} C}(t, t, \ldots, t). \tag{7}$$

The bandwidth for zero-order functional moment estimation is tuned by multifold cross validation. The bandwidth for nonzero-order functional moment estimation is computed from the factor method; see Liu and Müller (2009) for details.

Example 5 The formula for general local polynomial regression can be more clearly understood through this example with $\alpha = (1, 0)$. In this case, the functional moment is $\mathbb{E}\, X^{(\alpha)}(t) = \partial_{s_1} C(s_1, s_2)|_{s_1 = t, s_2 = t}$, with dimension $\|\alpha\|_0 = \|(1, 0)\|_0 = 2$ and total order of derivatives $\|\alpha\|_1 = \|(1, 0)\|_1 = 1$. The estimate of the first-order derivative $\partial_{s_1} C(s_1, s_2)$ obtained through the following local linear ($q = 1$) regression

$$(\widehat{\theta}_0, \widehat{\theta}_1, \widehat{\theta}_2) = \arg\min_{(\theta_0, \theta_1, \theta_2) \in \mathbb{R}^3} \sum_{1 \leq i \leq n, 1 \leq j_1, j_2 \leq m, j_1 \neq j_2} K_b(T_{ij_1} - s_1) \times K_b(T_{ij_2} - s_2)$$

$$\times \left\{ X_{ij_1} X_{ij_2} - \theta_0 - \theta_1(T_{ij_1} - s_1) - \theta_2(T_{ij_2} - s_2) \right\}^2, \tag{8}$$

is $\widehat{\partial_{s_1} C}(s_1, s_2) = \widehat{\theta}_1$.

Based on the functional moment estimates (7), the plugged-in FEE from (6) is $\sum_{k=0}^{p} \beta_k \widehat{\mathbb{E}} X^{(\alpha_k)}(t) g_k(t)$, which should be close to zero for all $t \in [0, 1]$. Consequently, the parameters $(\beta_1, \ldots, \beta_p)$ are estimated by minimizing the L^2 norm of this plugged-in FEE in the interval $[H, 1 - H]$ as follows:

$$(\widehat{\beta}_1, \ldots, \widehat{\beta}_p) = \arg \min_{(\beta_1, \ldots, \beta_p) \in \mathbb{R}^p} \int_H^{1-H} \left\{ \sum_{k=0}^{p} \beta_k \widehat{\mathbb{E}} X^{(\alpha_k)}(t) g_k(t) \right\}^2 dt. \quad (9)$$

Here, H is a boundary parameter, chosen to be a small constant approaching zero but significantly larger than the bandwidth h. The regions $[0, H] \cup [1 - H, 1]$ are excluded from the integration to guard against edge effects (Hall & Ma, 2014).

By taking the partial derivatives with respect to β_l, the minimization problem in (9) is equivalent to solving the ensuing critical point equations

$$\int_H^{1-H} \widehat{\mathbb{E}} X^{(\alpha_0)}(t) g_0(t) \widehat{\mathbb{E}} X^{(\alpha_l)}(t) g_l(t) dt + \sum_{k=1}^{p} \widehat{\beta}_k \int_H^{1-H} \widehat{\mathbb{E}} X^{(\alpha_k)}(t) g_k(t) \widehat{\mathbb{E}} X^{(\alpha_l)}(t) g_l(t) dt = 0$$

for all $1 \le l \le p$. Define the integration $\widehat{M}_{kl} = \int_H^{1-H} \widehat{\mathbb{E}} X^{(\alpha_k)}(t) g_k(t) \widehat{\mathbb{E}} X^{(\alpha_l)}(t) g_l(t) dt$, the matrix $\widehat{M} = (\widehat{M}_{kl})_{1 \le k, l \le p}$, and the vector $\widehat{M}_0 = (\widehat{M}_{0l})_{1 \le l \le p}$. The estimates of the unknown parameters $\{\beta_k\}_{k=1}^{p}$ are computed using the following explicit expression:

$$(\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^T = -\widehat{M}^{-1} \widehat{M}_0.$$

The matrix inverse $\widehat{M}^{-1}$ exists with probability tending to one, as substantiated by Assumption 1 (3) and Theorem 1 in Section 4.

We note that these parameter estimates are obtained without numerically solving ODEs; instead, the challenges of solving ODEs are incorporated within the functional moment. Therefore, the concept of the functional moment is important to the applicability of our proposed method to a wide range of ODEs. This includes nonrepresentable, non-Lipschitz and higher-order ODEs, which are seldom considered in existing research (Dai & Li, 2022; Ramsay et al., 2007; Wang et al., 2014; Xue et al., 2010; N. Zhang et al., 2022).

3.2 Trajectory recovery

The approach that balances data fitting discrepancies with ODE conformity is widely adopted to recover the trajectories (Ramsay et al., 2007; Xun et al., 2013; N. Zhang et al., 2022), i.e. consider

$$\arg \min_{f \in H^{\nu_0}(\mathcal{T})} \frac{1}{m} \sum_{j=1}^{m} \left(f(T_{ij}) - X_{ij} \right)^2 + \kappa \frac{1}{|\mathcal{T}|} \int_{\mathcal{T}} \left(F_{\widehat{\beta}}(f) \right)^2 dt, \quad (10)$$

where $\widehat{\beta} := \{\widehat{\beta}_k\}_{k=1}^{p}$ are the estimated parameters in (9) and κ serves as a tuning parameter that balances the two types of errors. The fidelity to the ODE enables us to circumvent the need for describing the solution space V_β, which may not have explicit expressions and not directly connected with the initial conditions. Instead, it focuses on quantifying the proximity of a function f to this solution space V_β.

To solve this minimization problem, we employ the functional principal component method (Hsing & Eubank, 2015). The eigen-decomposition of the covariance function

$$\Sigma(s, t)(:= \mathbb{E} \{X(s)X(t)\} - \mathbb{E} \{X(s)\} \mathbb{E} \{X(t)\}) = \sum_{1 \le l \le \infty} \lambda_l \phi_l(s) \phi_l(t)$$

gives the Karhunen–Loève decomposition $X(t) = \mathbb{E}\,X(t) + \sum_{1 \leq l \leq \infty} \xi_l \phi_l(t)$ of the random process X, where ξ_l, ϕ_l and λ_l represent the lth score, eigenfunction and eigenvalue, respectively. With aid of this decomposition, we approximate the curve X_i by projecting it onto the leading L principal components to obtain

$$X_i^L = \mathbb{E}\,X + \sum_{1 \leq l \leq L} \xi_{il}\phi_l. \tag{11}$$

This approximation X_i^L converges to X_i in the L^2 norm sense, as $\mathbb{E}\int_{\mathcal{T}}(X_i^L - X_i)^2 = \sum_{l=L+1}^{\infty}\lambda_l \to 0$ as L increases. The truncation number L is selected by the Fraction of Variance Explained method (Liu & Müller, 2009).

Subsequently, the minimization problem (10) is transformed into the following where only the finite scores need to estimate, given by

$$(\widehat{\xi}_{i1}, \ldots, \widehat{\xi}_{iL}) = \arg\min_{(\theta_{i1},\ldots,\theta_{iL})\in\mathbb{R}^L} \frac{1}{m}\sum_{j=1}^{m}\left(\widehat{Z}_i(T_{ij}) - X_{ij}\right)^2 + \frac{\kappa}{1-2H}\int_{H}^{1-H}\left(F_{\widehat{\beta}}(\widehat{Z}_i)\right)^2 \mathrm{d}t, \tag{12}$$

where $\widehat{Z}_i(\theta_{i1}, \ldots, \theta_{iL}) = \widehat{\mathbb{E}}\,X + \sum_{1 \leq l \leq L}\theta_{il}\widehat{\phi}_l$ is a truncation approximation. The tuning parameter κ is also selected via multifold cross-validation.

Finally, the recovery of a curve X_i is achieved using the first L principal components accompanied by the corresponding estimated scores

$$\widehat{X}_i = \widehat{\mathbb{E}}\,X + \sum_{1 \leq l \leq L}\widehat{\xi}_{il}\widehat{\phi}_l. \tag{13}$$

We still need to address estimates for the functional covariance/eigenfunctions and their derivatives, as studied in the literature (Liu & Müller, 2009; Yao et al., 2005). The functional covariance and its derivatives are estimated by

$$\widehat{\partial_t^v\Sigma}(s, t) = \widehat{\partial_t^v C_2}(s, t) - \widehat{\mathbb{E}}\,X(s)\widehat{\mathbb{E}}\,X^{(v)}(t) \tag{14}$$

for $0 \leq v \leq v_0$, where $\mathbb{E}\,X$ is the functional mean and $C_2(s, t) = \mathbb{E}\,X(s)X(t)$ is the functional second moment. Consequently, the eigenvalue estimates $\widehat{\lambda}_l$ and eigenfunction estimates $\widehat{\phi}_l$ are obtained from the eigen-decomposition of the covariance estimate $\widehat{\Sigma}(s, t) = \sum_{1 \leq l \leq \infty}\widehat{\lambda}_l\widehat{\phi}_l(s)\widehat{\phi}_l(t)$. Moreover, the derivative of the eigenfunction, $\phi_l^{(v)}$, also denoted by $\partial_v\phi_l$, is estimated by

$$\widehat{\partial_v\phi_l}(t) = \widehat{\lambda}_l^{-1}\int\widehat{\partial_t^v\Sigma}(s, t)\widehat{\phi}_l(s)\mathrm{d}s. \tag{15}$$

Note that the two potential orientations of eigenfunction estimates, $\widehat{\phi}_l$ and $-\widehat{\phi}_l$, would lead to identical recovery estimates $\widehat{X}_i$ as in (13). Consequently, there is no imperative to select a specific direction within our framework.

4 Theory

We begin by examining the functional moment estimate $\widehat{\mathbb{E}}\,X^{(\alpha)}$ of (7) and impose the standard regularity conditions, summarized as follows.

Assumption 1   Assume that

1. The kernel K_b is a smooth probability function and compactly supported on $[-b, b]$.

2. The observed time points $\{T_{ij}\}_{1 \le i \le n, 1 \le j \le m}$ are independent and identically distributed (i.i.d.) with a positive density function f_T on the domain $\mathcal{T}$. The random process X has a finite fourth moment $\mathbb{E} X^4(t) < \infty$ for every $t \in \mathcal{T}$.

3. The $p \times p$ matrix $M = (M_{kl})_{1 \le k, l \le p}$ with entry
$$M_{kl} = \int_H^{1-H} \mathbb{E} X^{(\alpha_k)}(t) g_k(t) \mathbb{E} X^{(\alpha_l)}(t) g_l(t) \mathrm{d}t \text{ is nonsingular (rank } M = p).$$

The Conditions (1) and (2) align with standard practices in local polynomial regression and functional data analysis (Yao et al., 2005; X. Zhang & Wang, 2016). Condition (3) guarantees the identifiability of the unknown parameters $\{\beta_k\}_{k=1}^p$. We subsequently delineate the convergence rate of the functional moment estimate integration in the following theorem, which facilitates the analysis of the parameter estimate (9). The regularity condition $\mathbb{E} X^{(2\alpha)} < \infty$ for a vector $\alpha = (\alpha_{[1]}, \alpha_{[2]}, \ldots, \alpha_{[N]})$, signifies that

$$\mathbb{E} X^{(2\alpha)}(t) = \mathbb{E}\left\{ X^{(2\alpha_{[1]})}(t) X^{(2\alpha_{[2]})}(t) \cdots X^{(2\alpha_{[N]})}(t) \right\} < \infty.$$

This condition essentially requires that the product of certain derivatives of $X(t)$ is integrable so that the FEE admits reasonable solutions, similar to the moment conditions required by standard least-square problems.

Theorem 1 Suppose that Assumption 1 (1)(2) hold and the random process $X : \Omega \to H^{v_0}(\mathcal{T})$ exhibits the regularity property $\mathbb{E} X^{(2\alpha)} < \infty$ for a vector $\alpha = (\alpha_{[1]}, \alpha_{[2]}, \ldots, \alpha_{[N]})$ and $v_0 \ge \|\alpha\|_\infty$. Then, the weighted integration $\int_H^{1-H} \widehat{\mathbb{E}} X^{(\alpha)}(t) c(t) \mathrm{d}t$ of the moment estimate (7) with a smooth function $c(x)$ yields the following convergence rate

$$\int_H^{1-H} \left(\widehat{\mathbb{E}} X^{(\alpha)}(t) - X^{(\alpha)}(t) \right) c(t) \mathrm{d}t$$
$$= O_p \left(h^{1+q-\|\alpha\|_1} + \sqrt{\frac{1}{nh^{2\|\alpha\|_1}} \left(1 + \frac{1}{m^{\|\alpha\|_0} h^{\|\alpha\|_0 - 1}} \right)} \right)$$

that depends on q (the degree of the local polynomial regression), the bandwidth h, the dimension $\|\alpha\|_0 = N$, and the total order of derivatives $\|\alpha\|_1 = \sum_{1 \le l \le N} \alpha_{[l]}$.

As the degree q increases, this convergence rate improves; however, it worsens with larger dimension $\|\alpha\|_0$ and total order of derivatives $\|\alpha\|_1$. This rate is more precise compared with the conventional nonparametric pointwise rate of $|\widehat{\mathbb{E}} X^{(\alpha)}(t) - X^{(\alpha)}(t)|$ for any fixed $t \in \mathcal{T}$ (Fan & Gijbels, 1996; Liu & Müller, 2009; Masry, 1997; X. Zhang & Wang, 2016)

$$O_p \left(h^{1+q-\|\alpha\|_1} + \sqrt{\frac{1}{nh^{2\|\alpha\|_1}} \left(1 + \frac{1}{m^{\|\alpha\|_0} h^{\|\alpha\|_0}} \right)} \right).$$

This conventional rate is improved in our context because the integration operation $\int_H^{1-H} (\cdot) \mathrm{d}t$ compresses an N-dimensional space into an $(N-1)$-dimensional space, and thus the exponential index over h that depends on the dimension is reduced by 1. We note that the existing theory in Hall and Ma (2014) is a special case of our results: the convergence rate of local mean estimate integration on a one-dimensional interval, meaning $q = 0$ and $\alpha = (0)$, is of the parametric order $n^{-1/2}$ that is faster than the nonparametric pointwise rate $n^{-1/3}$.

The bandwidth h in Theorem 1 is selected according to the sample size n and the sampling frequency m, as stated in the following corollary. A phase transition phenomenon is observed in the

magnitude order $m \asymp n^{(\|a\|_0-1)/(2\|a\|_0 q+2\|a\|_0)}$: when m is less than this order, the convergence rate is jointly determined by n and m; when m grows faster than this order, it can no longer improve the convergence rate.

Corollary 1   Assume the conditions of Theorem 1. The notation $a_n \lesssim b_n$ implies $a_n/b_n = O(1)$ as n grows large, and $a_n \asymp b_n$ connotes both $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

- When $m \lesssim n^{(\|a\|_0-1)/(2\|a\|_0 q+2\|a\|_0)}$, the optimal bandwidth is $h \asymp n^{-1/(1+2q+\|a\|_0)} m^{-\|a\|_0/(1+2q+\|a\|_0)}$ and the rate is

$$\int_H^{1-H} \left(\widehat{\mathbb{E}} X^{(a)}(t) - X^{(a)}(t) \right) c(t) dt = O_p\left(n^{-\frac{1+q-\|a\|_1}{1+2q+\|a\|_0}} m^{-\frac{\|a\|_0(1+q-\|a\|_1)}{1+2q+\|a\|_0}} \right).$$

- When $m \gtrsim n^{(\|a\|_0-1)/(2\|a\|_0 q+2\|a\|_0)}$, the optimal bandwidth is $h \asymp n^{-1/(2+2q)}$ and the rate is

$$\int_H^{1-H} \left(\widehat{\mathbb{E}} X^{(a)}(t) - X^{(a)}(t) \right) c(t) dt = O_p\left(n^{-\frac{1+q-\|a\|_1}{2+2q}} \right).$$

By utilizing the rate of functional moment estimate integration stated in Theorem 1 and the principles of bandwidth selection presented in Corollary 1, we can specify the convergence rate of the parameter estimates $\{\widehat{\beta}_k\}_{k=1}^p$ in (9) and the corresponding bandwidths as follows.

Theorem 2   Suppose that Assumption 1 holds, and let the random process $X : \Omega \to H^{v_0}(\mathcal{T})$ exhibit the regularity $\mathbb{E} X^{(2\alpha_k)} < \infty$ for any vector α_k, where $v_0 = \max_{0 \le k \le p} \|\alpha_k\|_\infty$. Then, the parameter estimates $\{\widehat{\beta}_k\}_{k=1}^p$ in (9) achieve the convergence rate of $\max_{1 \le k \le p} |\widehat{\beta}_k - \beta_k| = O_p(r_\beta)$ with

$$r_\beta = \max_{0 \le k \le p} \left\{ h_k + \sqrt{\frac{1}{nh_k^{2\|\alpha_k\|_1}} \left(1 + \frac{1}{m^{\|\alpha_k\|_0} h_k^{\|\alpha_k\|_0-1}} \right)} \right\},$$

when employing the local polynomial method of $\|\alpha_k\|_1$-degree to estimate the functional moment $\mathbb{E} X^{(\alpha_k)}$ with bandwidth h_k.

The bandwidths $\{h_k\}_{k=1}^p$ are chosen accordingly.

- When $m \lesssim \min_{1 \le k \le p} n^{(\|\alpha_k\|_0-1)/(2\|\alpha_k\|_0\|\alpha_k\|_1+2\|\alpha_k\|_0)}$, the optimal bandwidths are $h_k \asymp n^{-1/(1+2\|\alpha_k\|_1+\|\alpha_k\|_0)} m^{-\|\alpha_k\|_0/(1+2\|\alpha_k\|_1+\|\alpha_k\|_0)}$ and the rate is

$$r_\beta = \max_{0 \le k \le p} \left\{ n^{-\frac{1}{1+2\|\alpha_k\|_1+\|\alpha_k\|_0}} m^{-\frac{\|\alpha_k\|_0}{1+2\|\alpha_k\|_1+\|\alpha_k\|_0}} \right\}.$$

- When $m \gtrsim \max_{1 \le k \le p} n^{(\|\alpha_k\|_0-1)/(2\|\alpha_k\|_0\|\alpha_k\|_1+2\|\alpha_k\|_0)}$, the optimal bandwidths are $h_k \asymp n^{-1/(2+2\|\alpha_k\|_1)}$ and the rate is

$$r_\beta = \max_{0 \le k \le p} \left\{ n^{-\frac{1}{2+2\|\alpha_k\|_1}} \right\}.$$

Now we investigate the convergence of recovered curve $\widehat{X}_i$ in (13) and state the standard regularity conditions collected as follows.

Assumption 2 Assume that

1. The underlying eigenvalues exhibit a polynomial decay of the form $\lambda_l \asymp l^{-2a}$, where the spacing between them is $\lambda_l - \lambda_{l+1} \asymp l^{-2a-1}$ for a positive constant $a > 2$. Furthermore, the derivatives of the eigenfunction are uniformly bounded $|\phi_l^{(v)}| \lesssim l^{vb}$ for $1 \le l < \infty$. The decay rate satisfies $a > v_0 b + 1$, with $v_0 = \max_{0 \le k \le p} \|\alpha_k\|_\infty$.

2. The population X possesses the integrable conditions

$$\mathbb{E} \|dF_\beta(X)\|^2 < C_{(1)}, \quad \mathbb{E} \|d^2(F_\beta(X))^2\| < C_{(2)}$$

for some positive constants $C_{(1)}$, $C_{(2)}$. Here $dF_\beta : H^{v_0} \to \mathcal{L}(H^{v_0}, L^2)$ and $d^2 F_\beta : H^{v_0} \to \mathcal{L}(H^{v_0} \times H^{v_0}, L^2)$ are the Fréchet derivatives of F_β, where $\mathcal{L}(\mathcal{A}, \mathcal{B})$ denotes the space of linear functions from $\mathcal{A}$ to $\mathcal{B}$.

The condition (1), which is often employed in functional principal component analysis (Dou et al., 2012), implies that the selected principal components X_i^L in (11) can provide desired approximation to X_i as well as their corresponding vth-order derivatives ($v \le v_0$) in the uniform sense

$$\mathbb{E} \sup_{t \in \mathcal{T}} |(X_i^L)^{(v)}(t) - (X_i)^{(v)}(t)| \lesssim L^{-a+vb+1} \to 0, \text{ for } \forall \ 0 \le v \le v_0.$$

The integrability requirement in condition (2) describes the regularity of the underlying population X under the functional moments $(dF_\beta(X))^2$ and $d^2(F_\beta(X))^2$ that are comprised of $X, X', \ldots, X^{(v_0)}$, where the Fréchet-differentiability property is naturally satisfied due to the polynomial structure (1).

It is noteworthy that the second-order Fréchet derivative $d^2(F_\beta)^2$ generally does not have a lower bound (elliptical); this implies that there may not exist a positive bound $K_{F_\beta} > 0$ such that

$$\|\langle d^2(F_\beta)^2, \varphi, \varphi \rangle\|_{L^2} \ge K_{F_\beta} \|\varphi\|_{H^{v_0}}^2, \quad \text{for } \forall \ \varphi \in H^{v_0},$$

as the solution of $(F_\beta)^2 = 0$ might not be unique. This observation is also mentioned in N. Zhang et al. (2022). Consequently, in order to eliminate this nonconvexity when establishing the consistency of the recovery $\widehat{X}_i$ in (13), the tuning parameter κ in (12) should be sufficiently small, as demonstrated in the subsequent theorem.

Theorem 3 Suppose that Assumptions 1 and 2 hold and the random process $X : \Omega \to H^{v_0}(\mathcal{T})$ possesses regularity $\mathbb{E} X^{(2\alpha_k)} < \infty$ for any vector α_k and $v_0 = \max_{0 \le k \le p} \|\alpha_k\|_\infty$. Then the recovery $\widehat{X}_i = \widehat{\mathbb{E}} X + \sum_{1 \le l \le L} \widehat{\xi}_{il} \widehat{\phi}_l$ with cut-off L in (13) has the following L^2 convergence rate

$$\|\widehat{X}_i - X_i\|_{L^2} = O_p(L^{-a+2} + Lm^{-1/2} + L^{a+5/2} r_{\Sigma,0})$$

where the tuning parameter κ satisfies $\kappa L^{2v_0 b + 1} = o(1)$ and $\kappa(r_\beta + L^{a+1} r_{\Sigma,v_0} + L^{3a+2} r_{\Sigma,0}) = O_p(L^{-a+3/2} + L^{1/2} m^{-1/2} + L^{a+2} r_{\Sigma,0})$, r_β denotes the rate of parameter estimates in Theorem 2, and $r_{\Sigma,v} = h_\Sigma + n^{-1/2} h_\Sigma^{-v}(1 + m^{-1} h_\Sigma^{-1})$ refers to the L^2 convergence rate of covariance derivative estimate in (14) with bandwidth h_Σ.

The three components of this rate, L^{-a+2}, $Lm^{-1/2}$, and $L^{a+5/2} r_{\Sigma,0}$, arise from truncation approximation, discrete observations, and functional covariance/eigenfunction estimates, respectively. To

appreciate this new recovery rate and offer insights into the design of sampling principles, we choose the truncation parameter L and bandwidth h_Σ by optimizing the convergence rate of $\|\widehat{X}_i - X_i\|_{L^2}$, and clarify the relationship between sample size n and sampling frequency m in the following corollary.

Corollary 2 Assume the conditions of Theorem 3 and the bandwidth $h_\Sigma \asymp m^{-1/2}n^{-1/4}$.

- When $n \lesssim m^2$, the truncation level $L \asymp n^{1/(4a+1)}$ and

$$\|\widehat{X}_i - X_i\|_{L^2} = O_p(n^{-(a-2)/(4a+1)}).$$

- When $m^2 \lesssim n \lesssim m^{2+5/(a-1)}$, the truncation level $L \asymp m^{1/(4a+1)}n^{1/(8a+2)}$ and

$$\|\widehat{X}_i - X_i\|_{L^2} = O_p(m^{-(a-2)/(4a+1)}n^{-(a-2)/(8a+2)}).$$

- When $n \gtrsim m^{2+5/(a-1)}$, the truncation level $L \asymp m^{1/(2a-2)}$ and

$$\|\widehat{X}_i - X_i\|_{L^2} = O_p(m^{-(a-2)/(2a-2)}).$$

This corollary delineates notable phase transition phenomena in the recovery convergence rate, wherein two orders of magnitude $n \asymp m^2$ and $n \asymp m^{2+5/(a-1)}$ arise for the sample size n with respect to the sampling frequency m. The first transition order $n \asymp m^2$ comes from the covariance estimate, with its rate $r_{\Sigma,0}$ at a parametric order of $n^{-1/2}$ when $n \lesssim m^2$. Consequently, the truncation selection and recovery rate depend solely on the sample size and cannot benefit from an increasing sampling frequency m. When n grows faster than m^2, the rate $r_{\Sigma,0} = m^{-1/2}n^{-1/4}$ is jointly determined by the sample size n and sampling frequency m. As a result, both the truncation selection and recovery rate are influenced by the sampling frequency m. If n continues to increase up to the second transition order, given by $m^{2+5/(a-1)}$, it can no longer enhance the final rate $m^{-(a-2)/(2a-2)}$, which remains solely dependent on the sampling frequency m. It is important to note that this final rate cannot reach the parametric rate $m^{-1/2}$ for any decay rate a and large sample size n. This is already evident in classical smoothing spline analysis (Cai & Yuan, 2011).

5 Simulation

We illustrate our proposed parameter estimation and recovery method using the Population Model, Stable Field Model, and Damped Oscillation Model outlined in Examples 1–3. The settings for these models are as follows.

(i) In the population model (Example 1) with parameter $\beta = -1$, the random process $X(t)$ is given by $X(t) = Ze^t$.
(ii) In the stable field model (Example 2) with constant $C = 8$ and parameter $\beta = -4$, the random process $X(t)$ is defined as $X(t) = t^2 + 2\sqrt{Z+1}t + Z + 3$.
(iii) In the damped oscillation model (Example 3) with constant $C = 1$ and parameter $\beta = 2$, the random process $X(t)$ is expressed as $X(t) = (2Z + Zt)e^{-t}$.

The discrete observations are $\{X_{ij} = X_i(T_{ij}) + \varepsilon_{ij}\}_{1 \le i \le n, 1 \le j \le m}$, where $\{T_{ij}\}$ are uniformly distributed on $[0, 1]$ and the noise ε_{ij} follows Gaussian distribution Normal$(0, 0.1^2)$. The score Z that determines the random process is set to follow Uniform$(1, 2)$, Normal$(1, 0.1^2)$ or LogNormal$(0.5, 0.1^2)$ distributions.

We calculate the relative mean square errors

$$\{\mathbb{E}|\widehat{\beta} - \beta|^2\}/|\beta|^2$$

Table 1. Relative mean square errors $\mathbb{E}\,|\widehat{\beta} - \beta|^2/|\beta|^2$ of our method, assessed by 100 Monte Carlo runs with standard errors in the parentheses

Model (i)	n	$m = 3$	$m = 5$	$m = 7$	$m = 9$
Uniform	200	6.4e−03 (1.6e−03)	3.9e−03 (9.3e−04)	2.5e−03 (5.2e−04)	2.5e−03 (6.4-04)
	400	2.6e−03 (5.1e−04)	1.9e−03 (5.0e−04)	1.0e−03 (2.4e−04)	1.3e−03 (2.8e−04)
Normal	200	1.6e−03 (4.2e−04)	8.3e−04 (3.2e−04)	1.0e−03 (2.1e−04)	2.8e−04 (1.1e−04)
	400	6.3e−04 (1.1e−04)	5.2e−04 (1.0e−04)	4.3e−04 (8.8e−05)	1.7e−04 (7.6e−05)
LogNormal	200	1.8e−03 (3.2e−04)	9.1e−04 (1.1e−04)	5.3e−04 (9.2e−05)	5.9e−04 (8.7e−05)
	400	6.3e−04 (1.4e−04)	3.2e−04 (1.0e−04)	3.1e−04 (7.8e−05)	2.6e−04 (7.2e−05)

Model (ii)	n	$m = 3$	$m = 5$	$m = 7$	$m = 9$
Uniform	200	3.2e−02 (6.1e−03)	6.9e−03 (1.5e−03)	3.5e−03 (6.2e−04)	2.7e−03 (5.3e−04)
	400	9.2e−03 (1.4e−03)	3.9e−03 (6.8e−04)	2.3e−03 (7.2e−04)	1.0e−03 (1.3e−04)
Normal	200	3.3e−03 (6.3e−04)	1.1e−03 (2.1e−04)	6.3e−04 (1.2e−04)	4.2e−04 (8.4e−05)
	400	1.9e−03 (4.3e−04)	7.1e−04 (2.2e−04)	3.1e−04 (1.1e−04)	2.6e−04 (9.8e−05)
LogNormal	200	9.5e−03 (1.4e−03)	3.0e−03 (5.8e−04)	1.2e−03 (2.1e−04)	1.1e−03 (1.9e−04)
	400	4.0e−03 (7.4e−04)	1.4e−03 (3.2e−04)	5.8e−04 (1.2e−04)	4.3e−04 (1.0e−04)

Model (iii)	n	$m = 3$	$m = 5$	$m = 7$	$m = 9$
Uniform	200	8.4e−02 (1.5e−02)	5.7e−02 (8.9e−03)	3.9e−02 (6.5e−03)	2.9e−02 (4.4e−03)
	400	4.3e−02 (6.9e−03)	3.4e−02 (5.4e−03)	2.2e−02 (3.6e−03)	2.2e−02 (3.1e−03)
Normal	200	3.8e−02 (6.1e−03)	1.9e−02 (2.6e−03)	1.4e−02 (1.9e−03)	9.0e−03 (1.4e−03)
	400	1.9e−02 (3.3e−03)	1.1e−02 (1.7e−03)	6.6e−03 (9.1e−04)	5.3e−03 (1.0e−03)
LogNormal	200	2.4e−02 (3.4e−03)	1.8e−02 (3.6e−03)	1.2e−02 (2.0e−03)	8.6e−03 (1.8e−03)
	400	1.5e−02 (3.4e−03)	7.3e−03 (1.2e−03)	6.7e−03 (1.1e−03)	3.6e−03 (6.8e−04)

for the parameter estimates at $n = 200$, 400 and $m = 3$, 5, 7, 9, employing models (i), (ii), and (iii), as presented in Table 1. These results show that our proposed parameter estimation method provides good approximations across various ODE models, distribution types, and sampling frequencies. The empirical error diminishes as either the sample size n or the sampling frequency m increases. Furthermore, these performances vary with the complexity of the ODE; specifically, parameter estimation performance declines as the dimension or the orders of the corresponding functional moments increase. Additional simulations of more (n, m) pairs with denser sampling frequencies and varying sampling frequency (random m) are provided in the online supplementary material.

Moreover, we compare our method with the Parameter Cascading method (Wang et al., 2014) using a random-effect ODE framework, which employs B-spline techniques in parameter estimation. It is important to note that the Parameter Cascading method is suitable only for the first-order representable models, such as the population model (i) in our study. Table 2 provides the relative mean square errors for an additional sampling frequency of $m = 1$ under a sample size of $n = 200$. These results demonstrate that our method consistently outperforms the Parameter Cascading method in different scenarios, particularly in extremely sparse situations (e.g. $m = 1$), where the Parameter Cascading method fails to provide a stable estimate.

We also assess the performance of our proposed recovery method and compare it with the PACE method (Yao et al., 2005) and the Parameter Cascading method (Wang et al., 2014). The PACE method that is particularly suitable for Gaussian error achieves curve recovery based solely on observations, without integrating the dynamic mechanism. The integral mean square errors

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \big(\widehat{X}_i(t) - X_i(t)\big)^2 \mathrm{d}t$$

Table 2. Relative mean square errors $\mathbb{E}|\widehat{\beta} - \beta|^2/|\beta|^2$ in Model (i) of our method and Parameter Cascading method (Wang et al., 2014), assessed by 100 Monte Carlo runs with standard errors in the parentheses

Our method	$m = 1$	$m = 3$	$m = 5$	$m = 7$	$m = 9$
Uniform	1.5e−02 (4.0e−03)	6.4e−03 (1.6e−03)	3.9e−03 (9.1e−04)	2.5e−03 (5.2e−04)	2.5e−03 (6.2e−04)
Normal	5.6e−03 (1.2e−03)	1.6e−03 (3.7e−04)	8.1e−04 (2.6e−04)	1.0e−03 (2.1e−04)	3.3e−04 (1.4e−04)
LogNormal	4.8e−03 (1.0e−03)	1.8e−03 (2.9e−04)	9.1e−04 (1.3e−04)	5.0e−04 (1.1e−04)	6.0e−04 (8.2e−05)

Cascading	$m = 1$	$m = 3$	$m = 5$	$m = 7$	$m = 9$
Uniform	fail	6.2e−03 (3.2e−04)	6.0e−03 (1.2e−04)	5.7e−03 (9.8e−05)	4.8e−03 (8.4e−05)
Normal	fail	5.8e−03 (3.1e−04)	5.4e−03 (1.4e−04)	5.2e−03 (9.3e−05)	4.3e−03 (8.6e−05)
LogNormal	fail	5.4e−03 (1.8e−04)	5.2e−03 (1.1e−04)	5.0e−03 (7.4e−05)	4.6e−03 (6.2e−05)

is used to evaluate the curve recovery performance, as it captures the difference between $\widehat{X}_i(t)$ and $X_i(t)$ along all time interval. Table 3 presented the MSE results for $n = 200$. These discrepancy results generally decrease with increasing sampling frequency m across various models and distribution types, indicating that higher sampling frequencies enhance curve recovery accuracy. The results demonstrate that our recovery method outperforms the Parameter Cascading method and exceeds the PACE method in handling non-Gaussian scores across different models and distribution types, particularly in sparse designs (lower sampling frequencies). Additionally, our method exhibits stable performance across these distribution types; and also demonstrates consistency in different Monte Carlo runs, as evidenced by the relatively lower standard errors in parentheses.

6 Real data examples

6.1 COVID-19 data

We evaluate a real dataset that records COVID-19 information from 33 Chinese provinces ($n = 33$), excluding Hubei, the epicenter of the outbreak. The dataset compiles counts of infected, cured, and deceased individuals over a 40-day period ($m = 40$) from February 10th to March 20th, 2020, following the lockdown of Wuhan, Hubei's provincial capital. The epidemic trends across these provinces, as illustrated in Figure 1, exhibit similar trajectories and with random variation among them. In epidemiology, the ODE-based SIRD model in Example 4 is suitable for COVID-19 data due to its ability to capture the continuous-time transmission dynamics and compartmental structure of diseases. Unlike traditional functional data analysis models that only treat data as static curves, the ODE model provides unique insights by considering and quantifying specific epidemiological parameters like the diffusion rate (new daily infected individuals relative to current cases) that do not appear in traditional functional data analysis.

We apply the proposed Functional Moment Model (FMM) to estimate the unknown parameters β, γ, and ρ in this SIRD ODE system, representing the rates of diffusion, curing, and death, respectively. The estimated diffusion rate, $\widehat{\beta} = -4.530$, the recovery rate, $\widehat{\gamma} = 4.522$, and the mortality rate, $\widehat{\rho} = 0.0227$, are presented in the second column of Table 4. Specifically, the diffusion and recovery rates indicate that, on average, the number of infected/recovered individuals decreases/increases by $|\widehat{\beta}|/m \times 100\% = 11.33\%$ or $|\widehat{\gamma}|/m \times 100\% = 11.31\%$ of the current number of infected individuals per day. In contrast, the estimated mortality rate, $\widehat{\rho} = 0.0227$, implies a smaller but notable daily increase, constituting $|\widehat{\rho}|/m \times 100\% = 0.06\%$ of the current number of infected individuals. These parameter estimates not only underscore the effectiveness of the proposed FMM in capturing key epidemiological parameters but also enhance our understanding of the disease's transmission and progression.

Table 3. Mean square errors $\frac{1}{n}\sum_{i=1}^{n}\int_0^1(\widehat{X}_i(t) - X_i(t))^2 \mathrm{d}t$ of our method, PACE method (Yao et al., 2005) and Parameter Cascading method (Wang et al., 2014), assessed by 100 Monte Carlo runs with standard errors in the parentheses

(i)	Our method	$m = 3$	$m = 5$	$m = 7$	$m = 9$
	Uniform	8.4e−03 (4.2e−04)	4.5e−03 (2.1e−04)	3.5e−03 (1.2e−04)	2.8e−03 (9.5e−05)
	Normal	4.6e−03 (9.2e−05)	2.6e−03 (8.7e−05)	1.9e−03 (6.6e−05)	1.4e−03 (5.1e−05)
	LogNormal	6.2e−03 (2.2e−04)	3.5e−03 (1.3e−04)	2.4e−03 (1.2e−04)	1.9e−03 (8.8e−05)
	PACE	$m = 3$	$m = 5$	$m = 7$	$m = 9$
	Uniform	1.2e−01 (3.0e−02)	9.1e−02 (3.1e−02)	2.7e−02 (1.1e−02)	5.4e−03 (1.3e−03)
	Normal	5.9e−03 (2.1e−04)	3.2e−03 (1.3e−04)	2.2e−03 (9.1e−05)	1.7e−03 (7.8e−05)
	LogNormal	8.2e−02 (4.0e−02)	1.2e−02 (4.8e−03)	7.3e−03 (2.6e−03)	2.7e−03 (1.1e−04)
	Cascading	$m = 3$	$m = 5$	$m = 7$	$m = 9$
	Uniform	3.8e−02 (1.1e−02)	2.0e−02 (1.3e−04)	1.1e−02 (1.0e−04)	9.6e−03 (1.2e−04)
	Normal	2.9e−02 (7.4e−03)	9.6e−03 (1.2e−04)	7.2e−03 (1.1e−04)	5.8e−03 (8.8e−05)
	LogNormal	3.6e−02 (8.9e−03)	2.3e−02 (1.4e−04)	1.3e−02 (1.2e−04)	1.1e−02 (1.3e−04)
(ii)	Our method	$m = 3$	$m = 5$	$m = 7$	$m = 9$
	Uniform	8.9e−03 (2.1e−04)	6.3e−03 (2.2e−04)	5.2e−03 (1.4e−04)	3.7e−03 (1.0e−04)
	Normal	6.6e−03 (2.3e−04)	4.9e−03 (2.0e−04)	3.9e−03 (1.1e−04)	3.5e−03 (1.0e−04)
	LogNormal	7.8e−03 (2.2e−04)	4.7e−03 (1.1e−04)	4.0e−03 (2.0e−04)	3.3e−03 (1.1e−04)
	PACE	$m = 3$	$m = 5$	$m = 7$	$m = 9$
	Uniform	7.1e−02 (1.8e−02)	2.0e−02 (7.4e−03)	7.3e−03 (2.5e−03)	4.0e−03 (2.0e−04)
	Normal	4.3e−03 (1.4e−04)	2.6e−03 (1.2e−04)	1.9e−03 (1.0e−04)	1.4e−03 (1.3e−04)
	LogNormal	1.6e−02 (4.8e−03)	5.3e−03 (1.6e−03)	3.6e−03 (1.1e−04)	2.9e−03 (1.0e−04)
(iii)	Our method	$m = 3$	$m = 5$	$m = 7$	$m = 9$
	Uniform	5.9e−03 (2.1e−04)	3.3e−03 (1.2e−04)	2.5e−03 (8.7e−05)	1.8e−03 (6.8e−05)
	Normal	3.9e−03 (1.2e−04)	2.3e−03 (1.0e−04)	1.6e−03 (1.3e−04)	1.3e−03 (7.7e−05)
	LogNormal	4.7e−03 (1.0e−04)	2.8e−03 (1.0e−04)	2.0e−03 (1.1e−04)	1.5e−03 (1.4e−04)
	PACE	$m = 3$	$m = 5$	$m = 7$	$m = 9$
	Uniform	9.5e−02 (2.2e−02)	3.6e−02 (1.1e−02)	1.5e−02 (6.0e−03)	4.4e−03 (1.0e−03)
	Normal	7.2e−03 (2.5e−03)	2.9e−03 (1.3e−04)	2.0e−03 (7.2e−05)	1.5e−03 (1.3e−04)
	LogNormal	3.5e−02 (8.8e−03)	1.1e−02 (3.9e−03)	2.9e−03 (1.2e−04)	2.2e−03 (1.1e−04)

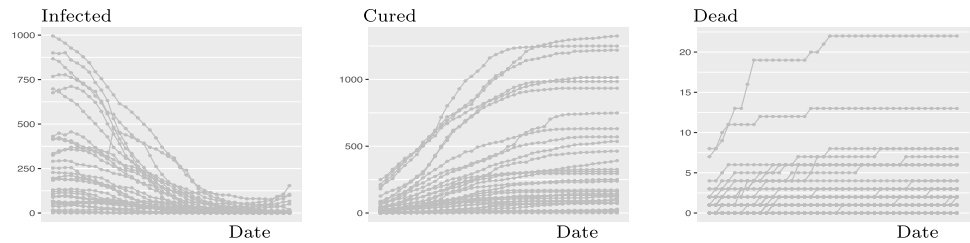*Note.* Here (i), (ii), and (iii) refer to Models (i), (ii), and (iii), respectively.

Figure 1. The plots from left to right represent the number of infected, cured, and dead across 33 Chinese provinces from 10 February to 20 March, respectively.

Table 4. The estimates $\widehat{\beta}, \widehat{\gamma}, \widehat{\rho}$ $(\widehat{\beta}', \widehat{\gamma}', \widehat{\rho}')$ and the RRE, evaluated across 100 randomly sparsified repetitions

	no sparsity	$m' = 14$	$m' = 15$	$m' = 16$	$m' = 17$	$m' = 18$	$m' = 19$	$m' = 20$	$m' = 21$
$-\widehat{\beta}$	4.530	4.408	4.457	4.514	4.481	4.439	4.537	4.493	4.528
$\widehat{\gamma}$	4.522	4.545	4.621	4.530	4.546	4.620	4.478	4.494	4.500
$\widehat{\rho}$	.0227	.0209	.0201	.0211	.0207	.0213	.0214	.0194	.0218
RRE	0	.0392	.0387	.0386	.0370	.0370	.0361	.0358	.0350

	$m' = 22$	$m' = 23$	$m' = 24$	$m' = 25$	$m' = 26$	$m' = 27$	$m' = 28$	$m' = 29$	$m' = 30$
$-\widehat{\beta}$	4.545	4.480	4.542	4.507	4.534	4.506	4.509	4.526	4.537
$\widehat{\gamma}$	4.515	4.534	4.509	4.516	4.556	4.519	4.543	4.533	4.573
$\widehat{\rho}$	.0203	.0233	.0214	.0220	.0221	.0221	.0217	.0220	.0225
RRE	.0351	.0353	.0351	.0345	.0345	.0343	.0340	.0337	.0341

	$m' = 31$	$m' = 32$	$m' = 33$	$m' = 34$	$m' = 35$	$m' = 36$	$m' = 37$	$m' = 38$	$m' = 39$
$-\widehat{\beta}$	4.551	4.542	4.540	4.523	4.535	4.538	4.517	4.534	4.541
$\widehat{\gamma}$	4.533	4.486	4.516	4.540	4.536	4.502	4.534	4.504	4.509
$\widehat{\rho}$	.0221	.0223	.0224	.0226	.0227	.0225	.0224	.0225	.0225
RRE	.0338	.0335	.0331	.0332	.0329	.0326	.0325	.0324	.0325

RRE, relative recovery error.

To assess the effectiveness of our method under sparse sampling conditions, we randomly select $m' = 14, 15, \ldots, 38, 39$ time points in each trajectory, meaning uniformly sampling m' observations without replacement from $\{(T_{ij}, X_{ij})\}_{1 \leq j \leq m}$ for each curve X_i. We then compare the parameter estimates $\widehat{\beta}', \widehat{\gamma}', \widehat{\rho}'$ obtained from the sparsified dataset with the corresponding estimates $\widehat{\beta}, \widehat{\gamma}, \widehat{\rho}$ derived from the complete dataset. Additionally, we compare the recovered infection counts I' with I from the full dataset. Table 4 presents the estimates $\widehat{\beta}', \widehat{\gamma}', \widehat{\rho}'$ and the Relative Recovery Error (RRE), defined as

$$\text{RRE} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \left(\widehat{I_i}'(t_j) - I_i(t_j)\right)^2}{\sum_{i=1}^{n} \sum_{j=1}^{m} \left(I_i(t_j)\right)^2}.$$

The trajectories of recovery, depicted across all provinces, are detailed in the online supplementary material. These results indicate that the estimates derived from the sparsified data converge to those from the complete dataset, with the RRE diminishes. Furthermore, even with smaller m' values, such as $m' = 14$ or 15, the estimates closely approximate those obtained from the complete dataset. Consequently, our proposed method demonstrates desirable performance under conditions of sparse sampling frequencies, which is owing to the proposed FMM that can effectively pool measurements together from different functions. Given the resource-intensive nature of COVID-19 data collection and the associated risks of reinfection during nucleic acid testing, employing a sparser sampling frequency with our method can conserve resources and reduce reinfection risks while still delivering reliable estimates.

6.2 AIDS data

We analyse the Acquired Immune Deficiency Syndrome (AIDS) dataset (Whiteside, 2016), provided by the China Center for Public Health Data Science at https://www.phsciencedata.cn. This dataset includes the counts I of infected individuals across 31 provinces in mainland China in 2020. The AIDS dataset is sparse with only $m = 12$ observed time points in each curve, compared with totally 366 days in 2020, due to the objective conditions of data collection. The AIDS detection depends on individuals voluntarily undergoing testing, and there are no daily mandatory tests. Besides, the

government reports AIDS data monthly rather than daily. Therefore, the data available for analysis are limited to 12 time points at the end of each month. This sparsity is not a choice but a reflection of the actual data-collection practices and the nature of AIDS reporting.

This AIDS dataset is also analysed using the SIRD model as described in Example 4. We apply the proposed Functional Moment Model (FMM) to estimate the unknown parameter β, which represents the rate of diffusion. The estimated diffusion rate, $\widehat{\beta} = 0.1258$, suggests that, on average, the number of infected individuals increases by $|\widehat{\beta}|/366 \times 1000\text{‰} = 0.3435\text{‰}$ of the current infected population per day. Additionally, we recover the trajectories from the sparsely recorded AIDS dataset, with plots for all provinces included in the online supplementary material. These recovered trajectories fill the gaps for the remaining $366 - 12 = 354$ days. The trajectories exhibit a clear increasing trend that aligns with the SIRD ODE model and closely match the sparsely observed data points. These findings indicate that our proposed method is effective in identifying the key parameter in the spreading process and in reconstructing the spread over the entire time interval.

7 Discussion

The model presented in Section 2 considers an **ordinary** differential operator constraint $F_\beta(X) = 0$, wherein $\{\beta_k\}_{k=0}^p$ are **scalar** parameters. The goal in this section is to generalize the model in order to cater for diverse requirements that emerge in an extensive range of practical physical scenarios. To this end, we propose enhancements in two areas: (1) modification of scalar parameters to functional parameters $\{\beta_k(t)\}_{k=0}^p$ and (2) extensions to PDEs.

7.1 Functional parameter

The functional parameter setting considers varying-coefficient ODE $F_\beta(X) = 0$ with functional (time-varying) parameters $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$ and $\beta_k : \mathcal{T} \to \mathbb{R}$. The recovery part for this case remains the same as those elaborated in Section 3.2, while the parameter estimation part in Section 3.1 needs some modifications. In fact, the minimization method of the plugged-in FEE in (9) yields

$$(\widehat{\beta}_1, \ldots, \widehat{\beta}_p) = \arg \min_{\beta_1, \ldots, \beta_p : \mathcal{T} \to \mathbb{R}} \int_H^{1-H} \left(\sum_{k=0}^p \beta_k(t) \widehat{\mathbb{E}} X^{(\alpha_k)}(t) g_k(t) \right)^2 dt,$$

which leads to $\sum_{k=0}^p \beta_k(t) \widehat{\mathbb{E}} X^{(\alpha_k)}(t) g_k(t) = 0$ for all $t \in \mathcal{T}$. However, this equation cannot determine the estimates $\{\widehat{\beta}_k(t)\}_{k=1}^p$ when $p > 1$. This is because the functional parameters lie in an infinitely dimensional space while that of scalar parameters is finitely dimensional. This gap between infinite and finite dimensions motivate us to create more FEEs to eliminate such nonidentifiability. We propose to multiply X^r ($0 \leq r \leq p - 1$) on both sides of ODE and take expectation to obtain

$$\sum_{k=0}^p \beta_k(t) g_k(t) \mathbb{E} \left(X^{(\alpha_k)}(t) X^r(t) \right) = \sum_{k=0}^p \beta_k(t) g_k(t) \mathbb{E} X^{(\alpha_k+r)}(t) = 0, \quad \forall t,$$

where $\alpha_k + r$ here refers to adding r zeros before α_k, i.e. $\alpha_k + r = (0, \ldots, 0, \alpha_{k,[1]}, \alpha_{k,[2]}, \ldots)$. Each levelled-up functional moment $\mathbb{E} X^{(\alpha_k+r)}$ is estimated from local polynomial regression and pooling strategy described in Section 3.1. Then $\beta_k(t)$ is estimated from the multiplied FEEs

$$\widehat{\mathbb{E}} g_0(t) X^{(\alpha_0+r)}(t) + \sum_{k=1}^p \widehat{\beta}_k(t) g_k(t) \widehat{\mathbb{E}} X^{(\alpha_k+r)}(t) = 0, \quad r = 0, \ldots, p - 1.$$

This levelling-up technique overcome this nonidentifiability gap and contributes to extend our model to functional parameter case.

7.2 PDE model

In PDE model, the underlying population $X : \Omega \to H^{v_0}(\mathcal{T}^D)$ is a multi-determinant random function and the partial differential operator F_β is

$$F_\beta(X) = \sum_{k=0}^{p} \beta_k X^{(\alpha_k)}(t) g_k(t) = 0$$

where the multi-index $X^{(\alpha_k)}$ of a $D \times N$ matrix $\alpha_k = (\alpha_{k,[1]}, \alpha_{k,[2]}, \ldots, \alpha_{k,[N]})$ with $\alpha_{k,[i]}$ a D-dimensional vector is

$$X^{(\alpha_k)} = \partial^{(\alpha_{k,[1]})} X \times \partial^{(\alpha_{k,[2]})} X \times \cdots \times \partial^{(\alpha_{k,[N]})} X.$$

The parameter estimation and recovery methods for PDE are the same as those for ODE in Section 3, but suffer more heavy computation in the estimation of functional moments, i.e. $\widehat{\mathbb{E}} X^{(\alpha_k)}$. In fact, $\mathbb{E} X^{(\alpha_k)}$ include totally DN dimensions and $\|\alpha_k\|_1 := \|\alpha_{k,[1]}\|_1 + \|\alpha_{k,[2]}\|_1 + \cdots + \|\alpha_{k,[N]}\|_1$ order derivatives, which consume more computing resources and require great quantities of data to converge in the local polynomial regression method.

Conflicts of interest: No competing interest is declared.

Data availability

All the programs and data used in our numerical study are available at https://github.com/LingxuanShao/ODE-CDF.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series B*.

References

Black F., & Scholes M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654. https://doi.org/10.1086/260062

Cai T. T., & Yuan M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Annals of Statistics*, *39*(5), 2330–2355. https://doi.org/10.1214/11-AOS898

Dai X., & Li L. (2022). Kernel ordinary differential equations. *Journal of the American Statistical Association*, *117*(540), 1711–1725. https://doi.org/10.1080/01621459.2021.1882466

Dou W. W., Pollard D., & Zhou H. (2012). Estimation in functional regression for general exponential families. *Annals of Statistics*, *40*(5), 2421–2451. https://doi.org/10.1214/12-AOS1027

Fan J., & Gijbels I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall Ltd.

Hall P., & Ma Y. (2014). Quick and easy one-step parameter estimation in differential equations. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *76*(4), 735–748. https://doi.org/10.1111/rssb.12040

Hsing T., & Eubank R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley and Sons.

Landau L. D., & Lifshitz E. M. (1976). *Mechanics*. Butterworth-Heinemann.

Leah E. K. (2005). *Mathematical models in biology*. Society for Industrial and Applied Mathematics.

Li Y., & Hsing T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics*, *38*(6), 3321–3351. https://doi.org/10.1214/10-AOS813

Liu B., & Müller H. G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association*, *104*(486), 704–717. https://doi.org/10.1198/jasa.2009.0115

Ma Z., Zhou Y., & Wu J. (2009). *Modeling and dynamics of infectious diseases*. Higher Education Press.

Marion M., & Temam R. (1998). Navier-stokes equations: Theory and approximation. In *Numerical methods for solids (part 3) and numerical methods for fluids (part 1)* (Vol. 6, pp. 503–689). Elsevier.

Masry E. (1997). Multivariate regression estimation: Local polynomial fitting for time series. *Nonlinear Analysis: Theory, Methods and Applications*, *30*(6), 3575–3581. https://doi.org/10.1016/S0362-546X(97)00415-X

Ramsay J. O., Hooker G., Campbell D., & Cao J. (2007). Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *69*(5), 741–796. https://doi.org/10.1111/j.1467-9868.2007.00610.x

Riley K. F., Hobson M. P., & Bence S. J. (1999). *Mathematical methods for physics and engineering*. Cambridge University Press.

Tan J., Zhang G., Wang X., Huang H., & Yao F. (2024). Green's matching: An efficient approach to parameter estimation in complex dynamic systems. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *86*(5), 1266–1285. https://doi.org/10.1093/jrsssb/qkae031

Wang L., Cao J., Ramsay J. O., Burger D. M., Laporte C. J. L., & Rockstroh J. K. (2014). Estimating mixed-effects differential equation models. *Statistics and Computing*, *24*(1), 111–121. https://doi.org/10.1007/s11222-012-9357-1

Whiteside A. (2016). *HIV and AIDS: A very short introduction*. Oxford University Press.

Xue H., Miac H., & Wu H. (2010). Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Annals of Statistics*, *38*(4), 2351–2387. https://doi.org/10.1214/09-AOS784

Xun X., Cao J., Mallick B., Maity A., & Carroll R. J. (2013). Parameter estimation of partial differential equation models. *Journal of the American Statistical Association*, *108*(503), 1009–1020. https://doi.org/10.1080/01621459.2013.794730

Yao F., Müller H. G., & Wang J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*(470), 577–590. https://doi.org/10.1198/016214504000001745

Zhang N., Nanshan M., & Cao J. (2022). A joint estimation approach to sparse additive ordinary differential equations. *Statistics and Computing*, *32*(5), 69. https://doi.org/10.1007/s11222-022-10117-y

Zhang X., & Wang J. L. (2016). From sparse to dense functional data and beyond. *Annals of Statistics*, *44*, 2281–2321. https://doi.org/10.1214/16-AOS1446