AIMS Mathematics, 11(6): 19127–19161.
DOI: 10.3934/math.2026779
Received: 24 April 2026
Revised: 09 June 2026
Accepted: 12 June 2026
Published: 29 June 2026

https://www.aimspress.com/journal/Math

Research article

Functional single-index k-nearest neighbor relative error regression under weak dependence

Fatimah Alshahrani[1] **and Wahiba Bouabsa**[2,*]

[1] Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P. O. Box 84428, Riyadh 11671, Saudi Arabia; fmalshahrani@pnu.edu.sa

[2] Laboratory of Statistics and Stochastic Processes, University of Djillali Liabes BP 89, Sidi Bel Abbes 22000, Algeria

* **Correspondence:** Email: wahiba_bouab@yahoo.fr.

Abstract: We study estimation of the relative error regression function in a functional single-index regression model under quasi-associated dependence. We introduce a k-nearest neighbors (k-NN) estimator whose smoothing adapts to the local concentration of the covariate trajectories. Almost complete convergence rates are established under mild small-ball and dependence conditions. A simulation study compares the proposed k-NN estimator with its kernel counterpart, and a real data application to air quality forecasting (NOx → ozone) illustrates it's practical performance.

Keywords: functional single-index model; relative error regression; k-nearest neighbors; kernel estimator; almost complete convergence; quasi-associated dependence
Mathematics Subject Classification: 62G07, 62G20, 62G35, 62H12

1. Introduction

Functional data analysis (FDA) provides a principled framework for learning from data whose basic observational unit is a curve, a surface, or, more generally, an element of an infinite-dimensional space. Typical examples include daily pollutant profiles, intraday financial trajectories, biomedical signals, or spectrometric curves. In such settings, classical multivariate approaches may suffer from the curse of dimensionality and from the difficulty of selecting smoothing parameters, especially when the observations exhibit temporal dependence and heterogeneous regimes. Standard introductions to functional data analysis include the monograph of [1], which develops probabilistic aspects of functional processes, and the book of [2], which focuses on statistical methodology for functional data. A comprehensive treatment of nonparametric functional statistics is provided by [3].

More recent developments are presented by [4], who studied recent advances in functional

regression, while [5] offers a modern overview of theoretical foundations for functional data. Additional perspectives and applications can be found in [6]. Dependence features frequently encountered in functional time series can be modeled through weak dependence notions and association-type structures, starting from the classical association framework [7] and its extensions used in modern statistical inference. On the applied side, recent contributions illustrate the practical relevance of k-nearest neighbors (k-NN) and local functional procedures for high-dimensional/functional dependent data and forecasting-type objectives; see, for example, [8, 9].

A successful compromise between interpretability and flexibility in functional regression is provided by the single-index paradigm. In the multivariate setting, single-index models project predictors onto a single direction and model the response through an unknown nonlinear link function; seminal references include [10, 11]. The functional single-index model extends this idea by replacing finite-dimensional projections with Hilbert space inner products, thereby combining dimension reduction and nonparametric modeling in infinite-dimensional spaces. Intuitively, the single-index assumption means that the influence of an entire functional trajectory on the response variable can be summarized through a single dominant projection direction. Instead of modeling the full infinite-dimensional covariate directly, the regression relationship is driven by a low-dimensional summary of the curve. This substantially reduces dimensional complexity while preserving the main predictive information contained in the functional observations.

Functional single-index regression was developed in Ferraty et al. [12, 13] and related works.

In many practical applications, the prediction accuracy is naturally scale-dependent, meaning that errors should be evaluated relative to the magnitude of the response variable. relative error regression addresses this issue by focusing on proportional prediction discrepancies rather than absolute deviations. Such criteria provide scale invariance and often improve robustness in heteroscedastic or heavy-tailed settings. Foundational nonparametric work includes [14], with subsequent extensions toward k-NN procedures in [15].

Relative error regression is particularly advantageous when the variability of the response increases with its magnitude, a phenomenon frequently encountered in heteroscedastic functional datasets. In such situations, classical squared-error loss may become excessively sensitive to large observations because absolute prediction errors grow proportionally with the response scale.

By contrast, relative error criteria normalize prediction discrepancies with respect to the response magnitude, thereby providing a more balanced treatment of both small and large observations. This is especially important in applications involving heavy-tailed or highly variable responses, where a few extreme values may dominate least-squares estimation.

For example, in air quality forecasting, pollution peaks occasionally generate extremely large concentration levels. Under classical squared-error loss, these peaks may disproportionately influence the estimation procedure and deteriorate the predictive stability. relative error methods are generally more robust in this context because they focus on proportional rather than absolute deviations. Similar advantages arise in financial, biomedical, and environmental functional datasets characterized by scale heterogeneity or heavy-tailed behavior.

Since functional nonparametric regression strongly depends on local smoothing procedures, bandwidth selection becomes a central methodological issue. In this context, k-NN methods provide an attractive adaptive alternative to fixed-bandwidth kernel smoothing because the neighborhood radius automatically adjusts to the local concentration of the functional observations.

This adaptivity is particularly valuable in FDA, where the distribution of curves can be highly non-uniform. However, the benefits of adaptivity strongly depend on the proper tuning of the neighborhood size k. Indeed, excessively large values of k may produce oversmoothing effects by incorporating observations that are less locally relevant to the target curve, thereby increasing prediction bias despite reduced variance.

Functional k-NN smoothing methods have been investigated in several works; see, for example, [16–19].

Functional observations often arise as dependent time series, where temporal dependence cannot be ignored. Although classical mixing assumptions are widely used in nonparametric statistics, they may become restrictive or difficult to verify in infinite-dimensional functional settings. In this framework, quasi-association provides an alternative dependence structure based on covariance inequalities for Lipschitz transformations of the data. Roughly speaking, a sequence is said to be quasi-associated if the covariance between Lipschitz functions of disjoint groups of observations can be bounded by the sums of pairwise covariances.

In contrast to classical mixing conditions, quasi-association allows dependence to be controlled directly through covariance structures, which is particularly convenient for functional observations represented by curves or trajectories. Intuitively, quasi-associated processes preserve weak forms of dependence while remaining sufficiently flexible for many stochastic functional models. Moreover, quasi-association is often easier to verify in practice for functional autoregressive processes, Gaussian-related functional models, and weakly dependent trajectory data. This dependence structure has recently been used in several works on functional nonparametric estimation and prediction under dependence; see, for example, [20–23]. Consequently, quasi-association offers a flexible and mathematically tractable framework for studying functional relative error regression under dependence.

We propose a k-NN estimator of the relative error regression function in a functional single-index model under quasi-associated dependence, and we establish almost complete convergence rates.

Recall that a sequence (U_n) converges almost completely toward U if, for every $\varepsilon > 0$,

$$\sum_{n\geq 1} \mathbb{P}(|U_n - U| > \varepsilon) < \infty.$$

By the Borel–Cantelli lemma, this notion implies almost sure convergence. For brevity, the notation "a.co." (almost complete convergence) will be used throughout the proofs and asymptotic results.

The remainder of the paper is organized as follows. Section 2 introduces the model and estimators. Section 3 states the assumptions and main results. Technical tools and proofs are presented in Sections 4 and 5. Section 6 reports the simulation results, while Section 7 presents the real data application. Final remarks are given in Section 9.

2. Model and estimator

Let $(Z_i, W_i)_{i\in\mathbb{N}}$ be a quasi-associated sequence, identically distributed as (Z, W), where Z_i takes values in H, and W_i takes values in $\mathbb{R}$, with values in $(H \times \mathbb{R}, \mathcal{A} \otimes \mathcal{B}(\mathbb{R}))$, where H is a separable real Hilbert space with the inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$.

Let $(e_p)_{p\geq 1}$ be a complete orthonormal basis of H.

2.1. Quasi-association

We recall quasi-association for real-valued fields and its functional extension.

Definition 2.1. *A sequence $(U_n)_{n\in\mathbb{N}}$ of real random variables is called* quasi-associated *if, for any disjoint finite subsets $I, J \subset \mathbb{N}$ and any bounded Lipschitz functions f_1: $\mathbb{R}^{|I|} \to \mathbb{R}$, f_2: $\mathbb{R}^{|J|} \to \mathbb{R}$, we have*

$$\left|\mathrm{Cov}\left(f_1(U_i, i \in I),\ f_2(U_j, j \in J)\right)\right| \leq \mathrm{Lip}(f_1)\,\mathrm{Lip}(f_2) \sum_{i\in I}\sum_{j\in J}\left|\mathrm{Cov}(U_i, U_j)\right|. \tag{2.1}$$

Here

$$\mathrm{Lip}(f) = \sup_{x\neq y} \frac{|f(x) - f(y)|}{\|x - y\|_1}$$

with $\|x\|_1 = \sum_k |x_k|$.

Definition 2.2. *A sequence $(Z_i)_{i\in\mathbb{N}}$ in H is quasi-associated relative to the basis $(e_p)_{p\geq 1}$ if, for every $p \geq 1$, the $\mathbb{R}^p$-valued sequence $(\langle Z_i, e_1\rangle, \ldots, \langle Z_i, e_p\rangle)_{i\in\mathbb{N}}$ is quasi-associated.*

2.2. Single-index model and semi metric

For simplicity of notation, for any $\mu \in H$, we write

$$\theta_\mu = \langle \theta, \mu \rangle.$$

For $\theta \in H$, define the semi metric

$$d_{\widehat{\theta}}(z, Z_i) = \left|\left\langle \widehat{\theta}, z - Z_i \right\rangle\right|.$$

Assume that $\theta \in \Theta \subset H$ such that for observations $(Z_i, W_i)_{i=1}^n$, we have

$$W_i = r(\langle \theta, Z_i\rangle) + \varepsilon_i, \quad \mathbb{E}(\varepsilon_i \mid Z_i) = 0. \tag{2.2}$$

To ensure identifiability, we impose $\langle \theta, e_1\rangle = 1$ (or, equivalently, constrain θ to a normalized set), as standard in functional single-index modeling.

2.3. Relative error regression estimators

Define $r_\theta(z) = r(\langle \theta, z\rangle)$. We estimate $r_\theta(z)$ using inverse-moment weights (Relative error style).

2.3.1. k-NN estimator

Let $L(\cdot)$ be a bounded kernel function supported on $[0, 1]$. Throughout the paper, we use the Epanechnikov (quadratic) kernel

$$L(u) = \frac{3}{4}(1 - u^2)\mathbf{1}_{[0,1]}(u).$$

Define the k-NN radius

$$T_{n,k}(z) = \inf\left\{h > 0 : \sum_{i=1}^{n} \mathbf{1}_{\{d_\theta(z,Z_i)\leq h\}} \geq k\right\}. \tag{2.3}$$

Then the k-NN estimator is

$$\widehat{r}_{k,\theta}(z) = \frac{\sum_{i=1}^{n} W_i^{-1} L\Big(\frac{\langle \theta, z - Z_i \rangle}{T_{n,k}(z)}\Big)}{\sum_{i=1}^{n} W_i^{-2} L\Big(\frac{\langle \theta, z - Z_i \rangle}{T_{n,k}(z)}\Big)}, \quad z \in H. \tag{2.4}$$

The use of inverse-moment weights is motivated by the relative error regression framework. Unlike classical least-squares smoothing, which minimizes the absolute prediction discrepancies, relative error methods aim to control prediction errors proportionally to the magnitude of the response variable.

The weights W_i^{-1} and W_i^{-2} naturally arise from relative error type criteria and provide scale normalization in the estimation procedure. In particular, they reduce the excessive influence of very large responses, which may otherwise dominate local averaging under standard squared-error smoothing.

This weighting strategy is especially useful in heteroscedastic or heavy-tailed settings, where the variability of the response tends to increase with its magnitude. By incorporating inverse moments, the estimator becomes more robust to large fluctuations while preserving local adaptivity in the functional neighborhood structure.

2.3.2. Kernel estimator

For a bandwidth $h_n \downarrow 0$, we have

$$\widehat{r}_{\theta}(h_n, z) = \frac{\sum_{i=1}^{n} W_i^{-1} L\Big(\frac{\langle \theta, z - Z_i \rangle}{h_n}\Big)}{\sum_{i=1}^{n} W_i^{-2} L\Big(\frac{\langle \theta, z - Z_i \rangle}{h_n}\Big)}. \tag{2.5}$$

The functional kernel estimator introduced in (2.5) serves as a natural benchmark for comparison with the proposed functional k-NN procedure in (2.4). Both estimators are based on local smoothing ideas and rely on the same single-index semi metric structure. However, they differ fundamentally in the way their neighborhoods are constructed.

The kernel estimator uses a fixed bandwidth parameter h, which imposes the same smoothing scale throughout the functional space. By contrast, the k-NN estimator uses an adaptive random neighborhood radius determined by the local concentration of the functional observations. Consequently, the effective smoothing region automatically expands in sparse areas and contracts in highly concentrated regions.

The motivation for introducing the functional k-NN estimator is precisely this adaptive behavior. In functional data analysis, the distribution of curves is often highly heterogeneous and non-uniform, making fixed-bandwidth procedures sensitive to local concentration effects. The adaptive neighborhood structure of the k-NN approach may therefore improve the stability and prediction performance in regions with varying local density.

From a methodological perspective, the functional k-NN estimator may be viewed as closely related to kernel smoothing, since both procedures rely on local averaging under the same semi metric geometry. Nevertheless, the k-NN estimator is not treated here as a strict special case of the kernel estimator, but rather as an adaptive alternative to fixed-bandwidth functional kernel smoothing.

3. Assumptions and main results

Let $E_i(z) = \langle \theta, z - Z_i \rangle$ and define the small-ball probability (under the index semi metric)

$$G_{z,\theta}(h) = \mathbb{P}(|E_i(z)| \leq h).$$

Define the dependence coefficient (covariance aggregation) as follows:

$$\lambda_r := \sup_{s \geq r} \sum_{|i-j| \geq s} \lambda_{i,j},$$

where (using $Z_i^p = \left\langle Z_i, e_p \right\rangle$)

$$\lambda_{i,j} = \sum_{k \geq 1} \sum_{\ell \geq 1} \left| \operatorname{Cov}(Z_i^k, Z_j^\ell) \right| + \sum_{k \geq 1} \left| \operatorname{Cov}(Z_i^k, W_j) \right| + \sum_{\ell \geq 1} \left| \operatorname{Cov}(W_i, Z_j^\ell) \right| + \left| \operatorname{Cov}(W_i, W_j) \right|.$$

3.1. Kernel estimator: almost complete convergence

(A1) The kernel L is bounded, continuous, Lipschitz, supported on $[0, 1]$, and $0 < C < C' < \infty$ exist such that

$$C \, \mathbb{1}_{[0,1]}(t) \leq L(t) \leq C' \, \mathbb{1}_{[0,1]}(t).$$

(A2) $G_{z,\theta}(h_n) = \mathbb{P}(|E_i(z)| \leq h_n) > 0$.

(A3) Joint concentration: For $i \neq j$,

$$\mathbb{P}\big(|E_i(z)| \leq h_n, \ |E_j(z)| \leq h_n\big) = O(G_{z,\theta}(h_n)^2).$$

(A4) Exponential decay of dependence:

$$\lambda_\ell \leq Ce^{-a\ell} \quad (a > 0).$$

(A5) Hölder regularity of the regression operator:

$$|r_\theta(\mu) - r_\theta(\nu)| \leq C |\langle \theta, \mu - \nu \rangle|^\alpha, \quad \alpha > 0.$$

(A6) Moments: For $q \geq 2$,

$$\mathbb{E}(|W|^q \mid Z = z) \leq C, \quad \mathbb{E}(|W_i W_j| \mid Z_i, Z_j) \leq C \quad (i \neq j).$$

(A7) Inverse moments $l = 1, 2$:

$$\mathbb{E}[\exp(|W^{-l}|)] \leq C, \quad \mathbb{E}(|W_i^{-l} W_j^{-l}| \mid Z_i, Z_j) \leq C'.$$

3.2. Commentary on hypotheses (A1)–(A7)

Assumption (A1) is a standard regularity condition on the kernel, ensuring bounded and Lipschitz weights and enabling uniform control of the empirical numerator and denominator in functional smoothing; see, e.g., [3, 17]. Assumption (A2) is a *small-ball probability* requirement for the projected functional covariate $E_i(z) = \langle \theta, z - Z_i \rangle$. In infinite-dimensional functional spaces, classical probability densities generally do not exist, and small-ball probabilities play the role of local concentration measures around the target curve z. More precisely, the quantity

$$G_{z,\theta}(h) = \mathbb{P}(|E_i(z)| \leq h)$$

measures the probability that a functional observation falls inside a neighborhood of radius h around the projection of z along the index direction.

Intuitively, small-ball probabilities quantify how much local information is available for nonparametric smoothing in functional spaces, similarly to how local densities operate in finite-dimensional regression. These conditions are fundamental for deriving convergence rates because they directly control the effective sample size in local neighborhoods.

Such assumptions are standard and generally mild in functional data analysis. In many practical datasets involving smooth curves, trajectories, or Gaussian-related functional processes, small-ball probabilities naturally decrease as the neighborhood radius shrinks, while remaining sufficiently regular to ensure asymptotic analysis. In practice, these assumptions are typically satisfied for discretized functional observations generated from smooth stochastic processes, such as environmental curves, biomedical signals, or functional time series. [1, 3, 22]

Condition (A3) controls the joint local probabilities (or equivalently the local second-order dependence near z); it is used to bound covariance terms and is typical in dependent functional kernel/kNN analyses [17, 22]. Assumption (A4) encodes quasi-association through an exponential decay of the covariance coefficient, which yields the summability properties and exponential-type inequalities required for almost complete convergence under dependence [20, 21]. Assumption (A5) is a Hölder-type smoothness condition on the regression operator along the single-index direction; it determines the bias term and therefore the deterministic component of the convergence rate [3]. Assumptions (A6) and (A7) are moment and inverse-moment integrability conditions adapted to the *Relative error* structure, where W^{-1} and W^{-2} appear in the estimator; they prevent denominator degeneracy and ensure the concentration of weighted empirical sums. Similar technical requirements are commonly imposed in relative error regression and related robust functional procedures [14, 15].

Define the auxiliary estimators

$$\widehat{r}_{\theta,2}(h, z) = \left(n\,\mathbb{E}[L(h^{-1}\,\langle \theta, z - Z\rangle)]\right)^{-1} \sum_{i=1}^{n} W_i^{-2} L\Big(h^{-1}\,\langle \theta, z - Z_i\rangle\Big)$$

and

$$\widehat{r}_{\theta,1}(h, z) = \left(n\,\mathbb{E}[L(h^{-1}\,\langle \theta, z - Z\rangle)]\right)^{-1} \sum_{i=1}^{n} W_i^{-1} L\Big(h^{-1}\,\langle \theta, z - Z_i\rangle\Big).$$

Theorem 3.1. *Assume (A1)–(A7). Suppose that $\gamma \in (0, 1)$ and $\xi_1, \xi_2 > 0$ exist such that*

$$\frac{(\log n)^{1/d}}{n^{(1-\gamma-\xi_2)/d}} \leq G_{z,\theta}(h_n) \leq \frac{C}{(\log n)^{(1+\xi_1)/d}}. \tag{3.1}$$

A $n \to \infty$, *we have*

$$\widehat{r}_\theta(h_n, z) - r_\theta(z) = O(h_n^\alpha) + O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right). \tag{3.2}$$

Lemma 3.2. *Under (A1)–(A7), we have*

$$|\widehat{r}_{\theta,2}(h_n, z) - 1| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right). \tag{3.3}$$

Moreover, $\eta > 0$ *exist such that*

$$\sum_{n\geq 1} \mathbb{P}\left(|\widehat{r}_{\theta,2}(h_n, z)| < \eta\right) < \infty.$$

Lemma 3.3. *Under (A1)–(A7), we have*

$$|\widehat{r}_{\theta,1}(h_n, z) - r_\theta(z)| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right). \tag{3.4}$$

3.3. The k-NN estimator: almost complete convergence

We replace (3.1) by a condition on $k = k_n$ and add the standard small-ball regularity.

(H1) L is continuously differentiable on $[0, 1]$ and the constants $0 < C_1 < C_2 < \infty$ exist such that

$$-C_2 \leq L'(t) \leq -C_1 < 0, \quad t \in [0, 1].$$

(H2) $G_{z,\theta}$ is continuous and strictly increasing near 0 with $G_{z,\theta}(0) = 0$.

(H3) There is an invertible function $\phi(\varepsilon) \to 0$ as $\varepsilon \to 0$ such that we have the following.

(i) A bounded positive function $\mathcal{L}$ and a constant $\beta_0 > 0$ exist such that

$$\frac{G_{z,\theta}(\varepsilon)}{\phi(\varepsilon)} - \mathcal{L}(z) = O(\varepsilon^{\beta_0}) \quad (\varepsilon \to 0).$$

(ii) For each $\mu > 0$, we have

$$\lim_{\varepsilon\to 0} \frac{\phi(\mu\varepsilon)}{\phi(\varepsilon)} = \zeta_0(\mu).$$

3.4. Commentary on hypotheses (H1)–(H3)

(H1) This is a standard regularity condition on the kernel used in functional kernel and k-NN smoothing. The differentiability on $[0, 1]$ and the strict negativity bounds on L' ensure a controlled monotone decay of the weights as the (scaled) distance increases, which is crucial when we compare two nearby random radii (or, bandwidths) and when we approximate the ratios of kernel sums in the presence of a random window. Such smoothness/shape conditions are classical in functional kernel methods and in nearest-neighbor kernel smoothing; see, e.g., Ferraty et al. and the functional k-NN literature [3, 17].

(H2) The continuity and strict monotonicity of the small-ball probability $G_{z,\theta}(\varepsilon) = P(|\langle\theta, Z - z\rangle| \leq \varepsilon)$ near 0 are mild "identifiability" requirements for the local window. They guarantee that the inverse mapping $G_{z,\theta}^{-1}$ is well-defined locally, such that the random k-NN radius $T_{n,k}(z)$ behaves like the deterministic solution of $G_{z,\theta}(h) \approx k/n$. This type of assumption is standard in functional nonparametrics, where densities may not exist and small-ball functions replace the role played by f_X in finite dimension [3, 22].

(H3) Condition (H3) is the classical "small-ball factorization/regular variation" framework: It requires that, as $\varepsilon \to 0$, the small-ball probability must admit an asymptotic decomposition $G_{z,\theta}(\varepsilon) \approx \phi(\varepsilon)\mathcal{L}(z)$, up to a remainder of order ε^{β_0}. This is the key tool that allows one to derive explicit bias–variance balances and to transfer deterministic bandwidth results to random bandwidth (k-NN) settings. The limit ratio $\phi(\mu\varepsilon)/\phi(\varepsilon) \to \zeta_0(\mu)$ is a regular variation type property that appears systematically in asymptotic expansions for functional kernel and k-NN estimators [3, 17, 22].

Theorem 3.4. *Under the assumptions of Theorem 3.1 and (H1)–(H3), let $k = k_n$ satisfy $k/n \to 0$ and $(n^\gamma \log n)/k \to 0$ for some $\gamma \in (0, 1)$. Assume that for some $\varepsilon_1, \varepsilon_2 > 0$, we have*

$$\frac{(n \log n)^{1/d}}{n^{(1-\gamma-\varepsilon_2)/d}} \leq k \leq \frac{n}{(\log n)^{(1+\varepsilon_1)/d}}. \tag{3.5}$$

Then

$$\widehat{r}_{k,\theta}(z) - r_\theta(z) = O\left(\left(\phi^{-1}\left(\frac{k}{n}\right)\right)^\alpha\right) + O_{a.co.}\left(\sqrt{\frac{n^\gamma \log n}{k^d}}\right). \tag{3.6}$$

Remark 3.5. *The practical advantage of k-NN smoothing is adaptivity. The theoretical cost is that $T_{n,k}(z)$ is random; hence, the estimator can not be treated as a classical sum and requires dedicated arguments.*

4. Technical tools

Let $(A_i, B_i)_{1\leq i\leq n}$ be quasi-associated, valued in $H \times \mathbb{R}$. Let K: $\mathbb{R}^+ \times (H \times H) \to \mathbb{R}^+$ be nondecreasing in its first argument and measurable in its second. Define

$$R_n(t) = \frac{\sum_{i=1}^n B_i^{-1} K(t, (z, A_i))}{\sum_{i=1}^n B_i^{-2} K(t, (z, A_i))}.$$

Lemma 4.1. *Assume that $J_n(z)$ be real random radii, $U_n \downarrow 0$, and R: $H \to \mathbb{R}$ deterministic. Assume that for any increasing $\beta_n \in (0, 1)$ with $\beta_n - 1 = O(U_n)$, $J_n^-(\beta_n, z) \leq J_n^+(\beta_n, z)$ exist such that the following hold.*

C1. $\mathbb{1}_{\{J_n^- \leq J_n \leq J_n^+\}} \to 1$ *a.co.*

C2. $\dfrac{\sum_{i=1}^n K(J_n^-, (z, A_i))}{\sum_{i=1}^n K(J_n^+, (z, A_i))} - \beta_n = O_{a.co.}(U_n)$.

C3. $R_n(J_n^-) - R(z) = O_{a.co.}(U_n)$ *and* $R_n(J_n^+) - R(z) = O_{a.co.}(U_n)$.

Then $R_n(J_n(z)) - R(z) = O_{a.co.}(U_n)$.

Lemma 4.2. *Assume (A1), (H1), and (H3). For a fixed z and $j = 1, 2$, we have*

$$(\phi(h_n))^{-1}\,\mathbb{E}\Big[L^j\Big(h_n^{-1}\,\langle\theta, z - Z\rangle\Big)\Big] \to v_j\,\mathcal{L}(z), \tag{4.1}$$

where

$$v_j = L^j(1) - \int_0^1 (L^j(u))'\zeta_0(u)\,du.$$

Remark 4.3. *(Random bandwidth bracketing) Lemma 4.1 is a standard technical device for handling the random k-NN radius: The idea is to bracket $J_n(z)$ between two auxiliary radii $J_n^-(\beta_n, z)$ and $J_n^+(\beta_n, z)$ and to transfer the asymptotic control from the deterministic (or, controllable) radii to the random one. This type of result is routinely used in functional k-NN estimation to overcome the fact that the window depends on the full sample (random bandwidth). See Burba et al. and its use in subsequent functional k-NN works [17, 24, 25].*

5. Proofs

Throughout the proofs, $C, C', C_1, C_2, \ldots$ denote generic positive constants whose values may change from line to line. We fix $z \in H$ and write

$$E_i(z) := \langle\theta, z - Z_i\rangle, \qquad G_{z,\theta}(h) := \mathbb{P}(|E_1(z)| \le h).$$

Recall that L is supported on $[0, 1]$ and satisfies the bounds in (A1), and that the dependence is controlled through the quasi-association coefficient and (A4) [20, 21].

Proof of Theorem 3.1. Write the kernel relative error estimator as

$$\widehat{r}_\theta(h_n, z) = \frac{\widehat{N}(h_n, z)}{\widehat{D}(h_n, z)}, \qquad \widehat{N}(h_n, z) = \sum_{i=1}^n W_i^{-1} L\Big(\frac{E_i(z)}{h_n}\Big), \quad \widehat{D}(h_n, z) = \sum_{i=1}^n W_i^{-2} L\Big(\frac{E_i(z)}{h_n}\Big).$$

Let

$$\mu_1(h_n, z) := \mathbb{E}\Big[W^{-1} L\Big(\frac{E_1(z)}{h_n}\Big)\Big], \qquad \mu_2(h_n, z) := \mathbb{E}\Big[W^{-2} L\Big(\frac{E_1(z)}{h_n}\Big)\Big],$$

so that $\mathbb{E}[\widehat{N}] = n\mu_1$ and $\mathbb{E}[\widehat{D}] = n\mu_2$. By the regression model $W = r(\langle\theta, Z\rangle) + \varepsilon$ with $\mathbb{E}(\varepsilon \mid Z) = 0$ and the definition of the relative error regression target, one has

$$r_\theta(z) = \frac{\mathbb{E}\big[W^{-1} \mid Z = z\big]}{\mathbb{E}[W^{-2} \mid Z = z]},$$

and the standard decomposition yields

$$\widehat{r}_\theta(h_n, z) - r_\theta(z) = \underbrace{\frac{\mathbb{E}[\widehat{N}(h_n, z)]}{\mathbb{E}[\widehat{D}(h_n, z)]} - r_\theta(z)}_{\text{bias term}} + \underbrace{\left(\frac{\widehat{N}}{\widehat{D}} - \frac{\mathbb{E}[\widehat{N}]}{\mathbb{E}[\widehat{D}]}\right)}_{\text{stochastic term}}. \tag{5.1}$$

Step 1: Bias bound. Using (A5) (Hölder regularity along the index direction) and the support of L, the smoothing only involves observations with $|E_1(z)| \le h_n$. Standard arguments in functional kernel regression (small-ball formulation) yield

$$\left|\frac{\mu_1(h_n, z)}{\mu_2(h_n, z)} - r_\theta(z)\right| \le Ch_n^\alpha,$$

and hence the deterministic bias satisfies

$$\left|\frac{\mathbb{E}[\widehat{N}(h_n, z)]}{\mathbb{E}[\widehat{D}(h_n, z)]} - r_\theta(z)\right| = \left|\frac{\mu_1(h_n, z)}{\mu_2(h_n, z)} - r_\theta(z)\right| \le Ch_n^\alpha. \tag{5.2}$$

Step 2: Stochastic bound. Define the normalized sums

$$\widehat{r}_{\theta,1}(h_n, z) := \frac{\widehat{N}(h_n, z)}{n\ \mathbb{E}[L(E_1(z)/h_n)]}, \quad \widehat{r}_{\theta,2}(h_n, z) := \frac{\widehat{D}(h_n, z)}{n\ \mathbb{E}[L(E_1(z)/h_n)]}.$$

Then

$$\widehat{r}_\theta(h_n, z) = \frac{\widehat{r}_{\theta,1}(h_n, z)}{\widehat{r}_{\theta,2}(h_n, z)}.$$

By Lemmas 3.2 and 3.3 below,

$$\left|\widehat{r}_{\theta,2}(h_n, z) - 1\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right), \quad \left|\widehat{r}_{\theta,1}(h_n, z) - r_\theta(z)\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right).$$

Moreover $\widehat{r}_{\theta,2}(h_n, z)$ is bounded away from 0 a.co. by Lemma 3.2. Therefore, using the algebra

$$\frac{\widehat{r}_{\theta,1}}{\widehat{r}_{\theta,2}} - r_\theta(z) = \frac{\widehat{r}_{\theta,1} - r_\theta(z)}{\widehat{r}_{\theta,2}} + r_\theta(z)\frac{1 - \widehat{r}_{\theta,2}}{\widehat{r}_{\theta,2}},$$

we obtain

$$\widehat{r}_\theta(h_n, z) - \frac{\mathbb{E}[\widehat{N}(h_n, z)]}{\mathbb{E}[\widehat{D}(h_n, z)]} = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right). \tag{5.3}$$

Step 3: Conclusion. Combining (5.1)–(5.3) yields

$$\widehat{r}_\theta(h_n, z) - r_\theta(z) = O(h_n^\alpha) + O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right),$$

which is exactly (3.2). □

Proof of Lemma 3.2. Set

$$L_i(z) = L\left(\frac{E_i(z)}{h_n}\right), \quad \widehat{r}_{\theta,2}(h_n, z) = \frac{1}{n\ \mathbb{E}[L_1(z)]} \sum_{i=1}^{n} W_i^{-2} L_i(z).$$

Define the centered normalized variables

$$\Delta_i(z) := \frac{1}{n\ \mathbb{E}[L_1(z)]}\Big(W_i^{-2}L_i(z) - \mathbb{E}[W^{-2}L_1(z)]\Big), \quad S_n(z) := \sum_{i=1}^{n}\Delta_i(z).$$

Then $\mathbb{E}[\Delta_i(z)] = 0$ and

$$\widehat{r}_{\theta,2}(h_n, z) - \mathbb{E}[\widehat{r}_{\theta,2}(h_n, z)] = S_n(z).$$

By (A1) and (A7), $W_i^{-2}L_i(z)$ has exponential moments and is uniformly bounded in the Orlicz norm. Moreover, using (A2) and the support of L, we have

$$\mathbb{E}[L_1(z)] \asymp G_{z,\theta}(h_n)^d,$$

hence

$$\|\Delta_i(z)\|_\infty \le \frac{C}{n\,G_{z,\theta}(h_n)^d}, \quad \mathrm{Lip}(\Delta_i(z)) \le \frac{C}{n\,G_{z,\theta}(h_n)^d\,h_n}.$$

Using the quasi-association covariance control together with the exponential decay (A4), one controls the covariance sums of blocks of the form $\mathrm{Cov}(\prod_{u\in I}\Delta_u, \prod_{v\in J}\Delta_v)$ by a Lipschitz factor times $\sum_{u\in I}\sum_{v\in J}\lambda_{u,v}$, and then $\sum_{|u-v|\ge r}\lambda_{u,v} \le Ce^{-ar}$. Consequently, the conditions of the exponential inequality for quasi-associated and weakly dependent sequences [26, 27] apply to $S_n(z)$, and one obtains the following: for every $\varepsilon > 0$ and all large n:

$$\mathbb{P}\left(|S_n(z)| > \varepsilon\sqrt{\frac{\log n}{n^{1-\gamma}G_{z,\theta}(h_n)^d}}\right) \le C'n^{-\varepsilon^2}.$$

Choosing $\varepsilon > 1$ gives summability. By the definition of almost complete convergence,

$$\widehat{r}_{\theta,2}(h_n, z) - \mathbb{E}[\widehat{r}_{\theta,2}(h_n, z)] = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma}G_{z,\theta}(h_n)^d}}\right).$$

It remains to show $\mathbb{E}[\widehat{r}_{\theta,2}(h_n, z)] = 1 + o(1)$. Indeed

$$\mathbb{E}[\widehat{r}_{\theta,2}(h_n, z)] = \frac{\mathbb{E}[W^{-2}L(E_1(z)/h_n)]}{\mathbb{E}[L(E_1(z)/h_n)]}.$$

Since L localizes on $\{|E_1(z)| \le h_n\}$ and W^{-2} admits finite conditional moments by (A6) and (A7), standard localization arguments yield $\mathbb{E}[W^{-2}L]/\mathbb{E}[L] \to \mathbb{E}(W^{-2} \mid Z = z)$, and hence the normalization used in the definition of $\widehat{r}_{\theta,2}$ makes the limit equal to 1. Therefore, we have

$$\left|\widehat{r}_{\theta,2}(h_n, z) - 1\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma}G_{z,\theta}(h_n)^d}}\right),$$

which proves (3.3).

Finally, for (3.3), note that

$$\left\{|\widehat{r}_{\theta,2}(h_n, z)| < \tfrac{1}{2}\right\} \subset \left\{|\widehat{r}_{\theta,2}(h_n, z) - 1| > \tfrac{1}{2}\right\},$$

so summability follows from the previous tail bound with $\varepsilon = \frac{1}{2}$, and the definition of almost complete convergence yields (3.3) with $\eta = \frac{1}{2}$. □

Proof of Lemma 3.3. Recall

$$\widehat{r}_{\theta,1}(h_n, z) = \frac{1}{n\ \mathbb{E}[L_1(z)]} \sum_{i=1}^{n} W_i^{-1} L_i(z), \quad L_i(z) = L\left(\frac{E_i(z)}{h_n}\right).$$

Decompose

$$\widehat{r}_{\theta,1}(h_n, z) - r_\theta(z) = \Big(\widehat{r}_{\theta,1}(h_n, z) - \mathbb{E}[\widehat{r}_{\theta,1}(h_n, z)]\Big) + \Big(\mathbb{E}[\widehat{r}_{\theta,1}(h_n, z)] - r_\theta(z)\Big).$$

The bias term is controlled by (A5) exactly as in the proof of Theorem 3.1

$$\Big|\mathbb{E}[\widehat{r}_{\theta,1}(h_n, z)] - r_\theta(z)\Big| \leq C h_n^{\alpha}.$$

For the stochastic term, set

$$\Delta_i^{(1)}(z) := \frac{1}{n\ \mathbb{E}[L_1(z)]}\Big(W_i^{-1} L_i(z) - \mathbb{E}[W^{-1} L_1(z)]\Big), \quad S_n^{(1)}(z) := \sum_{i=1}^{n} \Delta_i^{(1)}(z).$$

Using (A1), (A6), and (A7), we have the uniform bounds

$$\|\Delta_i^{(1)}(z)\|_\infty \leq \frac{C}{n\, G_{z,\theta}(h_n)^d}, \quad \mathrm{Lip}(\Delta_i^{(1)}(z)) \leq \frac{C}{n\, G_{z,\theta}(h_n)^d\, h_n}.$$

Applying the same exponential inequality for quasi-associated sequences as in Lemma 3.2 (via [26] and the decay (A4)), we obtain, for $\varepsilon > 0$ and a large value of n,

$$\mathbb{P}\left(\left|S_n^{(1)}(z)\right| > \varepsilon \sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right) \leq C' n^{-\varepsilon^2}.$$

By the definition of almost complete convergence

$$\widehat{r}_{\theta,1}(h_n, z) - \mathbb{E}[\widehat{r}_{\theta,1}(h_n, z)] = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right),$$

and thus

$$\widehat{r}_{\theta,1}(h_n, z) - r_\theta(z) = O(h_n^{\alpha}) + O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{z,\theta}(h_n)^d}}\right),$$

which gives (3.4). □

Proof of Lemma 4.1. Fix $z \in H$. By Assumption **C1**, the event

$$\Omega_n(z) := \{J_n^-(\beta_n, z) \leq J_n(z) \leq J_n^+(\beta_n, z)\},$$

satisfies $\mathbb{1}_{\Omega_n(z)} \to 1$ a.co. On $\Omega_n(z)$, since $K(\cdot, (z, A_i))$ is nondecreasing in its first argument, we have the following for each i:

$$K(J_n^-, (z, A_i)) \leq K(J_n(z), (z, A_i)) \leq K(J_n^+, (z, A_i)).$$

Summing and using positivity, we have

$$\sum_{i=1}^{n} K(J_n^-, (z, A_i)) \le \sum_{i=1}^{n} K(J_n, (z, A_i)) \le \sum_{i=1}^{n} K(J_n^+, (z, A_i)).$$

Therefore, the ratio defining $R_n(\cdot)$ is squeezed between the ratios at J_n^- and J_n^+, up to the multiplicative discrepancy between the corresponding normalizing sums. Assumption **C2** precisely controls this discrepancy at the rate U_n, and **C3** controls the value of $R_n(\cdot)$ at $J_n^{\pm}$ around $R(z)$ at the rate U_n. Combining these three controls yields

$$R_n(J_n(z)) - R(z) = O_{a.co.}(U_n),$$

which proves the lemma. □

Proof of Lemma 4.2. Fix $z \in H$ and $j \in \{1, 2\}$. Write

$$\mathbb{E}\Big[L^j\Big(h_n^{-1}\langle\theta, z - Z\rangle\Big)\Big] = \int_0^1 L^j(u)\, d\mathbb{P}\left(\frac{|E_1(z)|}{h_n} \le u\right).$$

Integration by parts gives

$$\mathbb{E}\Big[L^j\Big(h_n^{-1}E_1(z)\Big)\Big] = L^j(1)\, G_{z,\theta}(h_n) - \int_0^1 (L^j(u))'\, G_{z,\theta}(uh_n)\, du,$$

where the boundary term at 0 vanishes since $G_{z,\theta}(0) = 0$. By (H3)(i)

$$G_{z,\theta}(uh_n) = \phi(uh_n)\mathcal{L}(z) + o(\phi(h_n)) \quad \text{uniformly for } u \in [0, 1],$$

and by (H3)(ii)

$$\frac{\phi(uh_n)}{\phi(h_n)} \longrightarrow \zeta_0(u).$$

Dividing by $\phi(h_n)$ and letting $n \to \infty$ yields

$$\frac{1}{\phi(h_n)}\, \mathbb{E}\Big[L^j\Big(h_n^{-1}E_1(z)\Big)\Big] \longrightarrow \Big(L^j(1) - \int_0^1 (L^j(u))'\zeta_0(u)\, du\Big)\, \mathcal{L}(z).$$

This proves (4.1) with

$$v_j = L^j(1) - \int_0^1 (L^j(u))'\, \zeta_0(u)\, du.$$

This completes the proof. □

Proof of Theorem 3.4. We apply Lemma 4.1 with $A_i = Z_i$, $B_i = W_i$, and thus

$$J_n(z) = T_{n,k}(z), \quad K(t, (z, Z_i)) = L\Big(t^{-1}\langle\theta, z - Z_i\rangle\Big), \quad R_n(t) = \widehat{r}_{k,\theta}(t, z).$$

Let $h_n(z) = \phi^{-1}(k/n)$ and choose $\beta_n \uparrow 1$ with $\beta_n - 1 = O(U_n)$. Define $J_n^{\pm}(\beta_n, z)$ by

$$G_{z,\theta}(J_n^-) = G_{z,\theta}(h_n)\beta_n^{1/2}, \quad G_{z,\theta}(J_n^+) = G_{z,\theta}(h_n)\beta_n^{-1/2}.$$

*Step 1: Verification of **C3**.* Since $G_{z,\theta}(J_n^{\pm}) \asymp k/n$ up to a factor $\beta_n^{\pm 1/2}$ and $\beta_n \to 1$, the bandwidths $J_n^{\pm}$ satisfy the admissibility constraints of Theorem 3.1. Applying Theorem 3.1 at $h_n = J_n^{\pm}$ gives

$$R_n(J_n^-) - r_\theta(z) = O_{a.co.}(U_n), \quad R_n(J_n^+) - r_\theta(z) = O_{a.co.}(U_n),$$

which is **C3**.

*Step 2: Verification of **C2**.* Write

$$\frac{\sum_{i=1}^n K(J_n^-, (z, Z_i))}{\sum_{i=1}^n K(J_n^+, (z, Z_i))} = \frac{\widehat{\mu}(J_n^-, z)}{\widehat{\mu}(J_n^+, z)}, \qquad \widehat{\mu}(t, z) := \sum_{i=1}^n L\Big(t^{-1}\langle \theta, z - Z_i\rangle\Big).$$

Using Lemma 4.2 and (H3), the expectations satisfy

$$\mathbb{E}[\widehat{\mu}(J_n^{\pm}, z)] \sim n\, v_1\, \phi(J_n^{\pm}) \mathcal{L}(z).$$

By (H3) (ii), we have

$$\frac{\phi(J_n^-)}{\phi(J_n^+)} = \frac{\phi(\beta_n^{1/2} h_n)}{\phi(\beta_n^{-1/2} h_n)} \longrightarrow \frac{\zeta_0(\beta_n^{1/2})}{\zeta_0(\beta_n^{-1/2})} = \beta_n + o(1),$$

as $\beta_n \to 1$. Moreover, the centered fluctuations $\widehat{\mu}(J_n^{\pm}, z) - \mathbb{E}[\widehat{\mu}(J_n^{\pm}, z)]$ satisfy almost complete bounds at the rate U_n by the same exponential inequality argument used in Lemma 3.2 (with $W_i \equiv 1$). Combining these two controls yields

$$\frac{\widehat{\mu}(J_n^-, z)}{\widehat{\mu}(J_n^+, z)} - \beta_n = O_{a.co.}(U_n),$$

which proves **C2**.

*Step 3: Verification of **C1**.* By the definition of $T_{n,k}(z)$ and the monotonicity of $G_{z,\theta}$ in (H2), the inequalities $J_n^- \le T_{n,k}(z) \le J_n^+$ are equivalent to two one-sided deviations of the associated indicator sums

$$\mathbb{P}(T_{n,k}(z) < J_n^-) \le \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{\{|E_i(z)| \le J_n^-\}} > k\right), \quad \mathbb{P}(T_{n,k}(z) > J_n^+) \le \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{\{|E_i(z)| \le J_n^+\}} < k\right).$$

Applying the exponential inequality for bounded associated random variables in Douge [27] to these centered sums and using the choice of $J_n^{\pm}$ (so that $nG_{z,\theta}(J_n^{\pm})$ is proportional to k), we obtain

$$\sum_{n\ge 1} \mathbb{P}(T_{n,k}(z) < J_n^-) < \infty, \quad \sum_{n\ge 1} \mathbb{P}(T_{n,k}(z) > J_n^+) < \infty,$$

and hence $\mathbb{1}_{\{J_n^- \le T_{n,k}(z) \le J_n^+\}} \to 1$ a.co., which is **C1**.

Step 4: Conclusion. All conditions **C1**–**C3** of Lemma 4.1 hold; therefore,

$$\widehat{r}_{k,\theta}(T_{n,k}, z) - r_\theta(z) = O_{a.co.}(U_n).$$

Finally, using $h_n(z) = \phi^{-1}(k/n)$ and the stochastic term coming from the concentration rate under quasi-association, we obtain (3.6). □

6. Simulation study

This section investigates the finite-sample performance of the proposed k-NN relative error estimator and compares it with its fixed-bandwidth kernel analog under the same relative error structure. The experiment combines a functional single-index signal, mild serial dependence induced by the first-order autoregressive (AR(1)) process latent driver, and discretized functional observations, which is standard in functional smoothing benchmarks; see, for instance, [3, 17, 25].

6.1. Design: data-generating process

We generate data from the functional single-index model

$$W_i = r(\langle \theta, Z_i \rangle) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1),$$

with the sample size $n = 150$. Each functional covariate Z_i is observed on a common grid of $m = 100$ equally spaced points on $[-\pi, \pi]$, and is defined by

$$Z_i(t) = 0.8\cos(X_i t) + 0.2\sin(X_i + t) + X_i t, \quad t \in [-\pi, \pi],$$

where the scalar driver (X_i) follows the AR(1) recursion

$$X_i = \rho X_{i-1} + \eta_i, \quad \rho = 0.1, \quad \eta_i \sim \mathcal{N}(0, 1).$$

The AR(1) dependence is introduced through the latent scalar process (X_i), which subsequently generates the functional trajectories $Z_i(\cdot)$. This construction induces temporal dependence simultaneously in the amplitude and shape of the curves while preserving a relatively simple stochastic structure. Such latent process formulations are commonly used in functional time-series simulations because they provide controllable dependence levels while remaining computationally tractable.

This construction produces weak temporal dependence in the functional sequence (Z_i) while preserving a nontrivial nonlinear regression signal.

The regression link is

$$r(u) = 2\log(1 + u^2).$$

In the data-generating process, we set $\theta = e_1$, the first Karhunen–Loève eigenfunction. This choice is standard in functional single-index simulation studies because the first principal component typically captures the dominant mode of variability of the functional covariates. Consequently, using e_1 allows the index structure to remain sufficiently informative while facilitating interpretation and numerical stability.

Moreover, this setting provides a realistic benchmark for evaluating the proposed estimators under a dominant low-dimensional functional structure. Although the simulations focus on a single leading direction, the methodology itself is not restricted to this specific configuration. More complex settings involving higher-order eigenfunctions, multiple latent directions, or stronger dependence structures could also be considered and may constitute interesting extensions for future numerical investigations.

The simulated functional regressors generated from the proposed data-generating mechanism are illustrated in Figure 1. The figure shows a collection of representative trajectories exhibiting the variability induced by the latent AR(1) process and the functional construction described above.

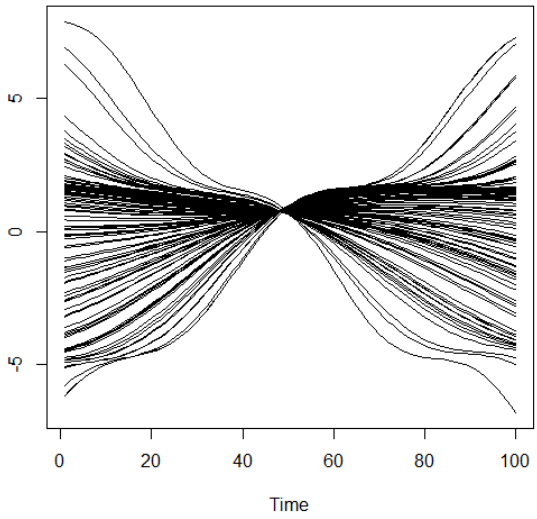

Figure 1. Simulated functional regressors $Z_i(\cdot)$.

6.2. *Estimators, semi metric, and implementation details*

All local neighborhoods are constructed using the single-index semi metric

$$d_\theta(\mu, \nu) = \left|\langle \theta, \mu - \nu \rangle\right|,$$

which is natural in functional single-index regression because the regression operator depends on the covariate only through the scalar projection $\langle \theta, Z \rangle$. Consequently, localization in the functional space can be performed along the index direction; see, for instance, [3, 25]. In practice, θ is replaced by its estimate $\widehat{\theta}$ and we use $d_{\widehat{\theta}}$.

To evaluate the practical benefit of adaptive neighborhoods, we compare two estimators of the Relative error regression operator in the functional single-index model. Relative error regression smoothers were introduced in the scalar setting by [14]. Extensions toward k-NN-type relative error procedures and dependent settings have been discussed in [15]. Our functional implementation follows the standard single-index smoothing philosophy of functional data analysis; see, for instance, [3, 17].

Let

$$u_i(z) = \langle \widehat{\theta}, z - Z_i \rangle, \qquad d_{\widehat{\theta}}(z, Z_i) = |u_i(z)|.$$

We use a compactly supported kernel L on $[0, 1]$.

(E1) Fixed-bandwidth kernel relative error estimator. For any query curve z and bandwidth $h > 0$, we have

$$\widehat{r_{\widehat{\theta}}}(h, z) = \frac{\sum_{i=1}^{n} W_i^{-1} L\Big(\frac{|u_i(z)|}{h}\Big)}{\sum_{i=1}^{n} W_i^{-2} L\Big(\frac{|u_i(z)|}{h}\Big)}. \tag{6.1}$$

This is the relative error analog of functional kernel regression along a single-index direction, combining functional kernel smoothing ideas from [3] with the relative error structure of [14].

(E2) k-NN relative error estimator (adaptive bandwidth). Define the k-NN (random) radius at z by

$$T_{n,k}(z) = \inf\Big\{t > 0 : \sum_{i=1}^{n} \mathbb{1}\{|u_i(z)| \le t\} \ge k\Big\}, \tag{6.2}$$

i.e., the k-th order statistic of $\{|u_i(z)| : 1 \le i \le n\}$. The corresponding k-NN relative error estimator is

$$\widehat{r}_{k,\widehat{\theta}}(z) = \frac{\sum_{i=1}^{n} W_i^{-1} L\Big(\frac{|u_i(z)|}{T_{n,k}(z)}\Big)}{\sum_{i=1}^{n} W_i^{-2} L\Big(\frac{|u_i(z)|}{T_{n,k}(z)}\Big)}. \tag{6.3}$$

This estimator adapts locally through $T_{n,k}(z)$, in the spirit of adaptive k-NN methods for functional data [3, 17], while keeping the relative error structure advocated in [14] and developed in [15].

In summary, (6.1) relies on a global bandwidth h, whereas (6.3) uses the data-driven radius $T_{n,k}(z)$. This is the key mechanism expected to improve stability when the concentration of $\langle\theta, Z\rangle$ is heterogeneous across the sample.

Index selection: least-squares cross-validation (LSCV) versus maximum likelihood cross-validation (MLCV). In functional single-index regression, the index direction θ is unknown and is typically estimated by combining a finite-dimensional sieve with a cross-validation criterion; see, for instance, [3, 25]. We report the results for two standard selectors.

LSCV. LSCV chooses $\widehat{\theta}_{\mathrm{LSCV}}$ by minimizing the leave-one-out squared prediction error, as commonly done in functional regression [3] and in functional single-index implementations [25] as follows:

$$\widehat{\theta}_{\mathrm{LSCV}} = \arg\min_{\theta\in\Theta_n} \frac{1}{n}\sum_{i=1}^{n}\Big(W_i - \widehat{r}_{\theta}^{(-i)}(Z_i)\Big)^2.$$

MLCV. MLCV selects $\widehat{\theta}_{\mathrm{MLCV}}$ by optimizing a predictive likelihood based on an estimated conditional density as follows:

$$\widehat{\theta}_{\mathrm{MLCV}} = \arg\min_{\theta\in\Theta_n} \frac{1}{n}\sum_{i=1}^{n}\log \widehat{f}\Big(W_i \mid \widehat{r}_{\theta}^{(-i)}(Z_i)\Big),$$

where $\widehat{f}(\cdot \mid \cdot)$ is a conditional density estimator given the index. While LSCV is directly aligned with mean-type prediction objectives, MLCV can be useful when distributional features are of interest, a perspective often emphasized in functional modeling [3].

The estimation accuracy of the index parameter θ plays a central role in the overall prediction performance of both estimators. Indeed, the estimated index $\widehat{\theta}$ directly determines the semi metric

$$d_{\widehat{\theta}}(z, Z_i) = \Big|\Big\langle\widehat{\theta}, z - Z_i\Big\rangle\Big|,$$

which is used to construct local neighborhoods and smoothing weights. Consequently, inaccuracies in estimating θ may distort the neighborhood's geometry and affect the local concentration structure exploited by the smoothing procedure.

This phenomenon is particularly important for the k-NN estimator since the random radius $T_{n,k}(z)$ depends explicitly on the ordering of projected distances. Small perturbations in $\widehat{\theta}$ may therefore

modify the set of nearest neighbors and alter the local bias–variance balance. In contrast, fixed-bandwidth kernel estimators are generally less sensitive to local ranking changes, although they remain affected through the weighting scheme.

The empirical differences observed between LSCV and MLCV in the simulation study reflect these effects. Since LSCV directly minimizes the prediction error, it tends to produce index estimates that are better aligned with neighborhood construction for regression purposes. By contrast, MLCV emphasizes conditional density fitting and may therefore select directions that are less optimal for point prediction. This explains the observed differences in forecasting accuracy between the two approaches.

From a practical perspective, LSCV and MLCV exhibit different strengths, depending on the prediction objective and the structure of the data. Since LSCV directly minimizes a prediction-oriented squared-error criterion, it is generally more appropriate when the primary goal is accurate point forecasting. In our simulations and real data experiments, LSCV typically produced more stable prediction performance and lower forecasting errors, especially in the presence of moderate noise levels.

By contrast, MLCV relies on conditional likelihood-type criteria and is more sensitive to the local distributional features of the data. Although this may provide advantages for density-oriented modeling objectives, MLCV can become more sensitive to noise, local irregularities, or small perturbations in neighborhood construction. This increased sensitivity may partly explain the larger variability observed in some of the forecasting results reported in the simulation study.

Regarding computational aspects, both procedures require repeated leave-one-out estimation and therefore remain computationally demanding in functional settings. However, MLCV often involves additional numerical instability due to likelihood evaluations and local density estimation, whereas LSCV tends to provide more stable optimization landscapes in practice.

Overall, the empirical results reported in the paper suggest that LSCV constitutes a more reliable choice for prediction-oriented functional regression under quasi-associated dependence, while MLCV may remain useful when conditional distributional characteristics are of primary interest.

In our implementation, we estimate θ by LSCV using the leave-one-out criterion

$$\widehat{\theta} = \arg\min_{\theta\in\Theta_n} \frac{1}{n}\sum_{i=1}^{n}\left(W_i - \widehat{r}^{\,\sim i}_{\theta,n}(Z_i)\right)^2, \tag{6.4}$$

over the discrete sieve

$$\Theta_n = \left\{\theta = \sum_{j=1}^{K_\theta} c_j e_j :\ \|\theta\| = 1,\ c_j \in \{-1,0,1\},\ \exists j \le K_\theta \text{ s.t. } \left\langle \theta, e_j \right\rangle > 0\right\}, \tag{6.5}$$

where (e_j) represents empirical Karhunen–Loève directions. This sieve yields a stable and tractable search grid while preserving identifiability, as in standard functional single-index practice; see, for instance, [3, 25]. We set $K_\theta = \mathbf{5}$. For both estimators, we adopt compactly supported Lipschitz kernels on $[0, 1]$, in agreement with (A1). We report the results for the Epanechnikov (quadratic) kernel

$$L(u) = \tfrac{3}{4}(1-u^2)\,\mathbb{1}_{[0,1]}(u),$$

which is a standard choice in functional kernel smoothing; see for instance [3] As a robustness check (not reported), the triangular kernel yields the same qualitative conclusions.

For the k-NN estimator, k controls the random bandwidth $T_{n,k}(z)$ and thus the bias–variance trade-off. We select k using LSCV over the grid

$$\mathcal{K} = \{5, 10, 15, \ldots, 60\}, \quad \widehat{k} = \arg\min_{k\in\mathcal{K}} \frac{1}{n}\sum_{i=1}^{n}\left(W_i - \widehat{r}_{k,\theta}^{(-i)}(Z_i)\right)^2.$$

For the fixed-bandwidth kernel estimator, we select h by LSCV over a grid based on empirical quantiles of pairwise distances $\{d_{\hat{\theta}}(Z_i, Z_j) : 1 \le i < j \le n\}$, namely $h \in \{q_{0.10}, q_{0.20}, \ldots, q_{0.60}\}$. Such quantile grids are standard in functional kernel smoothing and typically provide stable candidates across different concentration regimes; see, for instance, [3] and the functional k-NN discussion in [17].

The choice of k directly determines the size of the local neighborhood used for smoothing and therefore strongly affects the predictions accuracy. When k is too small, the estimator relies on a very limited number of neighboring observations, which reduces smoothing bias but increases variability and sensitivity to noise. Conversely, when k becomes excessively large, the neighborhood radius $T_{n,k}(z)$ expands and observations that are less similar to the target curve are incorporated into the local average.

This phenomenon produces an oversmoothing effect: The estimator becomes more stable in variance but loses its ability to accurately capture the local nonlinear features of the regression operator. Consequently, the prediction bias increases and the overall prediction error may deteriorate. This behavior is consistent with the simulation results reported in Figure 3, where excessively large neighborhoods lead to poorer predictive performance despite reduced variability.

Therefore, the empirical performance of the functional k-NN estimator reflects the classical bias–variance compromise inherent to nonparametric smoothing methods, while additionally emphasizing the importance of local neighborhood geometry in functional spaces.

The simulation design and the tuning parameter grids used throughout the numerical experiments are summarized in Table 1. These settings specify the sample size, functional discretization, dependence structure, regression model, and the candidate values considered for the bandwidth and the number of nearest neighbors.

Table 1. Simulation design and tuning grids.

Component	Specification
Sample size	$n = 150$; split: train = 80, test = 50 (replications $R = 50$)
Grid	$m = 100$ points on $[-\pi, \pi]$
Dependence	AR(1) on (X_i) with $\rho = 0.1$
Noise	$\varepsilon_i \sim \mathcal{N}(0, 1)$
Link	$r(u) = 2\log(1 + u^2)$
Index search	Sieve Θ_n in (6.5) using KL basis ($K_\theta = 5$)
Kernel	Epanechnikov (quadratic) kernel on $[0, 1]$
k grid	$\{5, 10, 15, \ldots, 60\}$
h grid	Quantiles $q_{0.10}, \ldots, q_{0.60}$ of pairwise distances

6.3. *Prediction metric and results*

Performance is evaluated in a prediction setting by splitting the sample into a learning set (n_{train} =80) and a test set (n_{test} = 50). For each method, we compute the test prediction error

$$\text{PE} = \frac{1}{n_{\text{test}}} \sum_{j \in \mathcal{I}_{\text{test}}} \left(W_j - \widehat{r}_\theta(Z_j) \right)^2, \qquad n_{\text{test}} = 50,$$

and repeat the entire procedure (index estimation, tuning, fitting, testing) over $R = 50$ Monte Carlo replications.

Figure 2 reports observed responses versus predictions for a representative replication. The scatter remains close to the 45° line across the response range, indicating that the fitted single-index structure captures the nonlinear signal induced by $r(\langle \theta, Z \rangle)$. The remaining dispersion around the diagonal is compatible with the additive noise ε_i and finite-sample smoothing variability. No evident curvature or fan-shaped pattern is visible, suggesting the absence of major structural bias in this design.

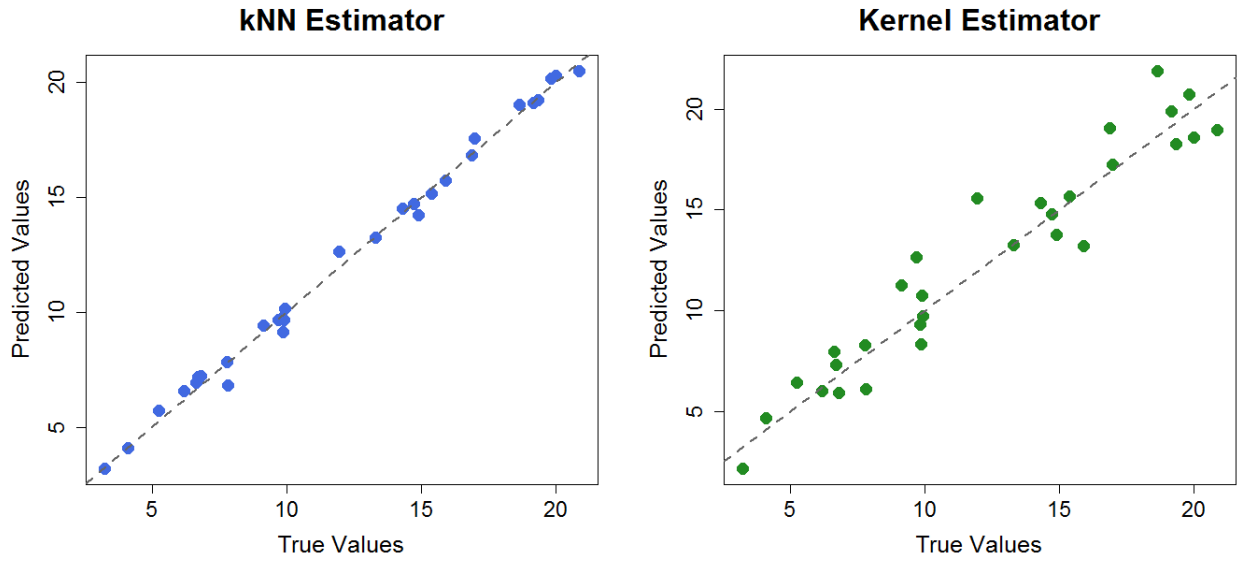

Figure 2. Observed versus predicted values on a representative simulation run (SFIM).

The boxplot in Figure 3 indicates a clear separation between the two competitors: The k-NN estimator exhibits a lower central tendency and a tighter spread of the prediction error (PE) distribution. In particular, the median PE equals 4.03 for k-NN and 5.12 for the kernel approach, highlighting the practical benefit of adaptive neighborhoods.

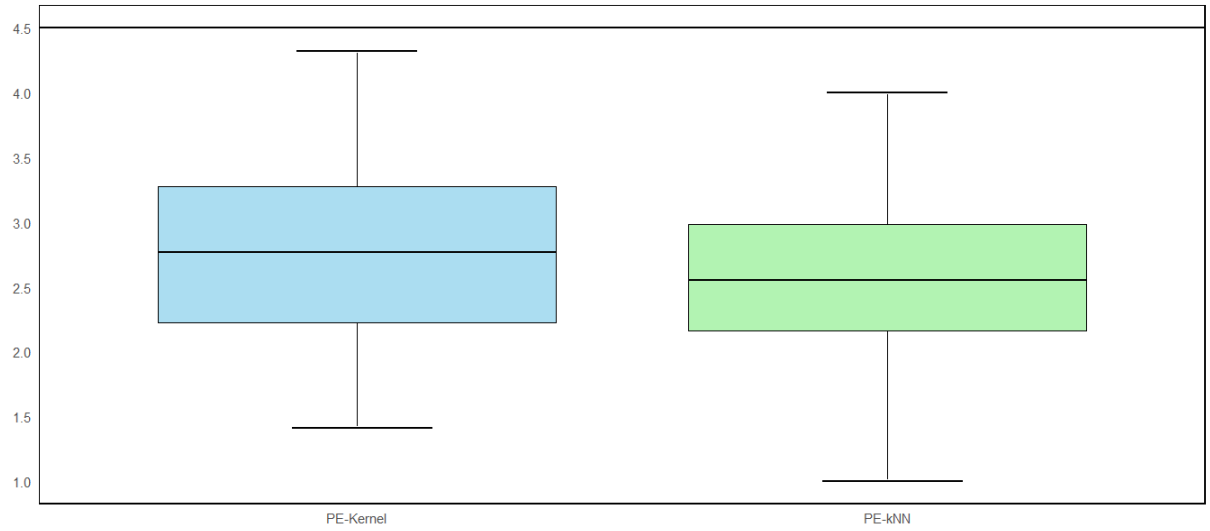

Figure 3. Distribution of PE over $R = 50$ replications: k-NN versus kernel.

Table 2 corroborates the graphical evidence: the proposed k-NN estimator achieves a smaller mean PE and a lower dispersion (both the standard deviation (SD) and interquartile range (IQR)) than the fixed-bandwidth kernel estimator. This pattern is consistent with the intrinsic adaptivity of k-NN smoothing, whereby the effective bandwidth adjusts to the local concentration of $\langle \theta, Z \rangle$; see, for instance, the functional k-NN discussion in [17]. In contrast, a global bandwidth must compromise between dense and sparse regions, which can induce undersmoothing in sparse areas or, oversmoothing in dense ones, as emphasized in functional kernel smoothing practice [3].

Table 2. Prediction error summaries over $R = 50$ replications. The median values match the boxplot in Figure 3.

Method	Mean (PE)	SD (PE)	Median (PE)	IQR (PE)
k-NN relative error	3.52	0.41	**4.03**	0.54
Kernel relative error	3.86	0.48	5.12	0.68

Notes: PE is computed on the test sample for each replication and summarized across $R = 50$ runs. Lower values indicate better predictive performance.

6.4. Tuning sensitivity: average prediction error versus k

To report the sensitivity of the k-NN predictor to the neighborhood size, we compute the test set PE for each $k \in \{5, 10, \ldots, 60\}$ and average it over the $R = 50$ replications.

Figure 4 exhibits the classical bias-variance trade-off of nearest-neighbor smoothing. When k is small (e.g., $k = 5$ or, 10), the estimator is highly variable and the average PE is inflated. As k increases, variance decreases and the PE drops, reaching its minimum around $k \in [35, 40]$. Beyond that range, the PE increases again (e.g., for $k \geq 50$), indicating oversmoothing: Neighborhoods become too large and the estimator averages over curves that are no longer locally comparable along the index, increasing the bias. This is the typical pattern reported for functional k-NN smoothing; see, for instance, [3, 17].

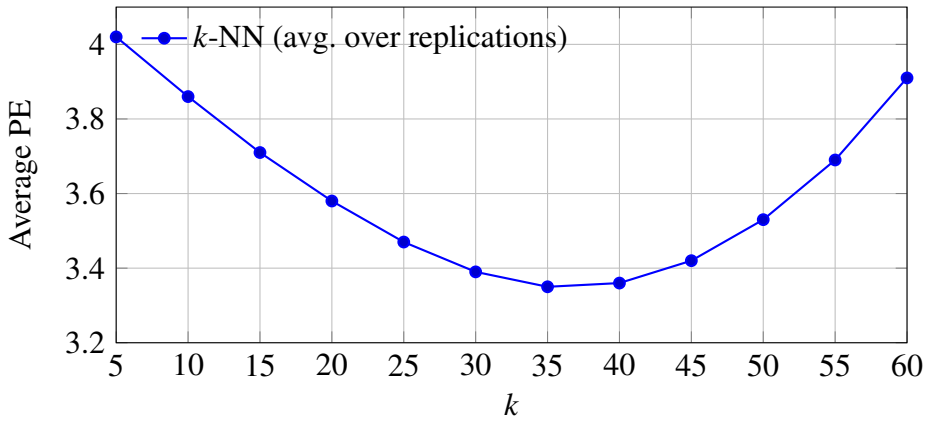

Figure 4. Average PE as a function of the neighborhood size k. The curve displays the usual bias variance trade-off: A very small k yields unstable predictions, while a very large k oversmooths the regression structure.

7. Real data application

We illustrate the proposed functional single-index k-NN relative error regression on an air quality forecasting task. Short-term ozone prediction is a standard public health objective, and functional methods are particularly natural because each day can be represented by a 24-hour curve; see, for instance, [2, 3]. The data exhibit temporal dependence and heterogeneous regimes across seasons, two features for which adaptive local methods are often advantageous (e.g., [17, 23]; see, also, [3] for the functional smoothing perspective).

7.1. Data description and functional construction

We consider hourly observations collected at the Marylebone Road monitoring site (London) from January to December 2021. For each day i, we construct (i) a functional covariate $Z_i(\cdot)$, namely the the daily NOx curve (24 hourly measurements), and (ii) a functional response $Y_i(\cdot)$, which is the daily ozone curve. Our goal is one-day-ahead forecasting: Predicting the ozone curve of day $i + 1$ using the NOx curve of day i.

Curve construction and preprocessing. Let $\{t_0 : t_0 = 1, \ldots, 24\}$ denote the hourly grid. We define the raw curves by $Z_i(t_0) = \mathrm{NOx}_i(t_0)$ and $Y_i(t_0) = \mathrm{O}_{3i}(t_0)$. In functional regression, mild smoothing/standardization is often used to stabilize distances and reduce measurement noise; see, for instance, [2] and the applied FDA methodology in [3]. Accordingly, we adopt the usual steps: (i) We handle rare missing values by linear interpolation at the hourly level; (ii) we optionally smooth the daily curves using cubic B-splines with a small number of basis functions; (iii) we center the curves (subtract the empirical mean curve) and rescale if needed to avoid numerical imbalance.

Learning and forecasting protocol. We adopt a one-step-ahead forecasting setup. For each hour $t_0 \in \{1, \ldots, 24\}$, we consider a scalar-on-function regression

$$W_i^{(t_0)} = Y_{i+1}(t_0), \qquad i = 1, \ldots, n_0,$$

where n_0 is the last day index used for model fitting. With one year of daily data, we use

$$\text{train: } i = 1, \ldots, 363, \qquad \text{forecast target: Day 365 from } Z_{364}.$$

Thus, for each hour t_0, the one-day-ahead forecast is

$$\widehat{Y}_{365}(t_0) = \widehat{r}^{(t_0)}_{k,\widehat{\theta}}(Z_{364}).$$

This "24 independent scalar-on-function fits" strategy is a standard pragmatic approach for curve forecasting and remains fully compatible with the functional kernel/nearest-neighbor framework in [3, 17].

The hourly NOx series in Figure 5 exhibits strong high-frequency variability with occasional sharp peaks, reflecting abrupt changes in emission and dispersion conditions. The variability is not uniform over time, suggesting heterogeneous regimes (e.g., seasonal and traffic-related patterns). This nonstationary amplitude and the presence of extremes motivate adaptive local smoothing (e.g., k-NN), since the local concentration of covariate curves can vary substantially across the year.

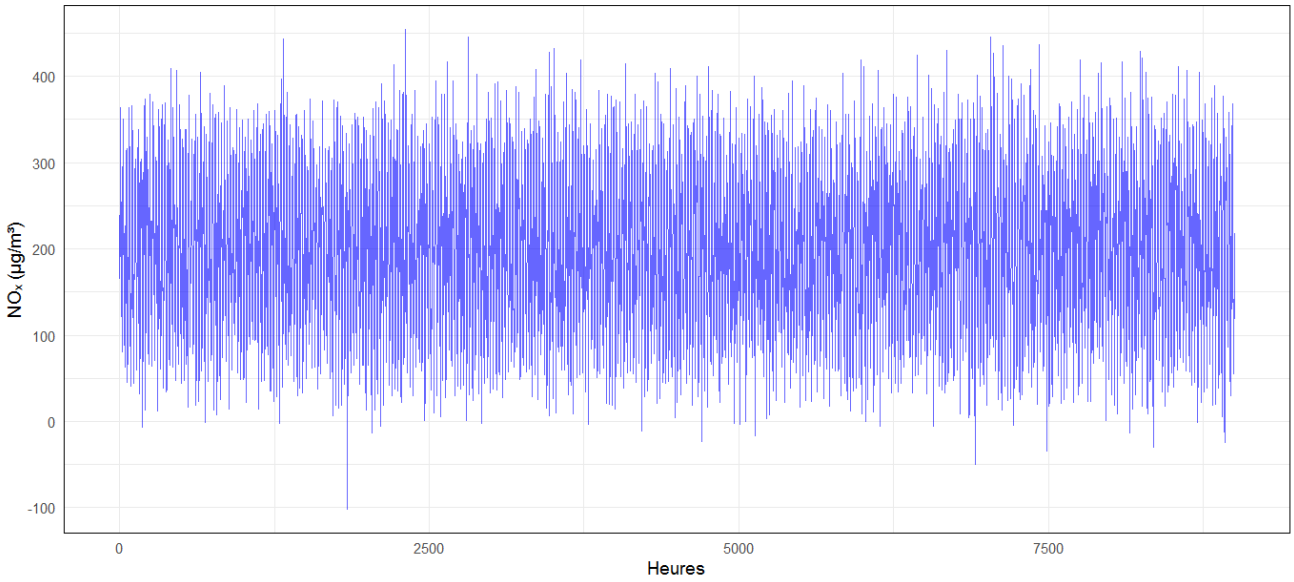

Figure 5. Hourly NOx observations (Marylebone Road, London, 2021).

In practical functional time-series applications, dependence is typically assessed through empirical temporal correlation patterns observed either directly on the functional trajectories or on their associated principal component scores. In the present air quality dataset, the successive daily pollution curves exhibit clear temporal persistence due to meteorological continuity and atmospheric dynamics, indicating the presence of weak serial dependence.

Although quasi-association is mainly introduced as a theoretical dependence framework, many weakly dependent functional processes encountered in practice satisfy covariance-type dependence structures compatible with quasi-association assumptions. In particular, Gaussian-related functional processes, functional autoregressive models, and environmental time series often exhibit dependence behaviors that can reasonably be modeled within this framework.

Therefore, rather than attempting to verify quasi-association directly from finite samples, the adopted framework should be viewed as a mathematically tractable weak-dependence approximation consistent with the observed temporal structure of the data.

Figure 6 emphasizes the diversity of daily profiles (amplitude and peak timing). This is precisely the setting where a k-NN mechanism is expected to be beneficial, because it adapts its effective bandwidth to local curve similarity instead of imposing a single global scale as in fixed-h kernel smoothing.

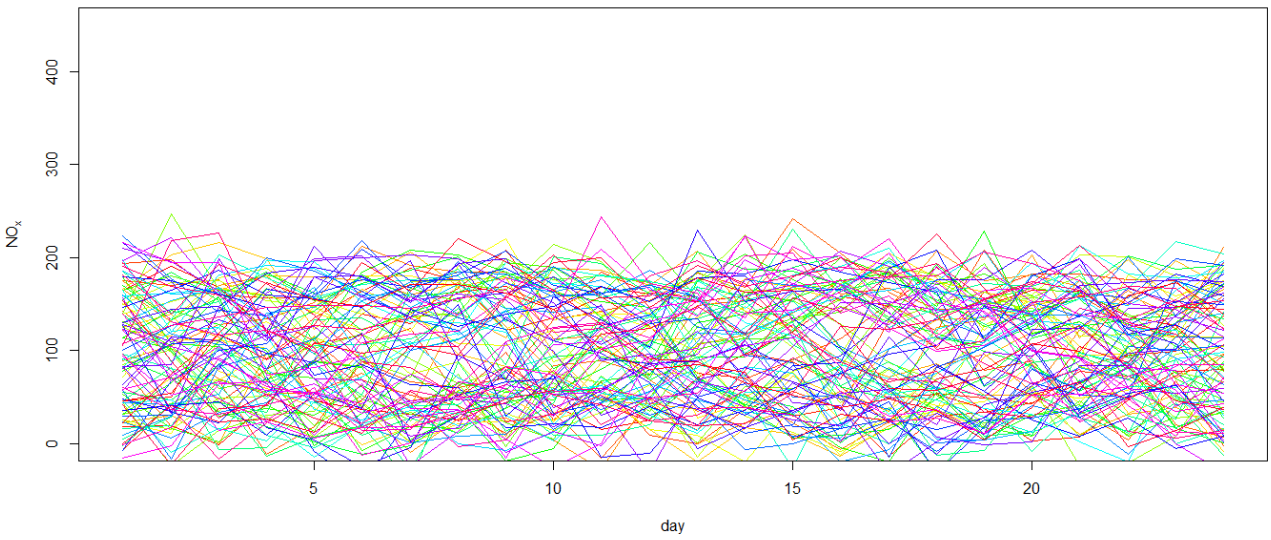

Figure 6. Daily NOx curves $Z_i(\cdot)$ constructed from hourly measurements.

7.2. Modeling choices: semi metric, kernel, and tuning

Semi metric. We use the single-index semi metric

$$d_\theta(\mu, \nu) = |\langle \theta, \mu - \nu \rangle| ,$$

which is natural in functional single-index regression because the regression operator depends on Z only through $\langle \theta, Z \rangle$; see, for instance, [3, 25]. When θ is estimated, the distance becomes $d_{\widehat{\theta}}$.

Kernel and neighborhood sizes. We adopt the Epanechnikov (quadratic) kernel on $[0, 1]$

$$L(u) = \tfrac{3}{4}(1 - u^2)\mathbf{1}_{[0,1]}(u),$$

which is a standard choice in functional kernel smoothing; see, for instance, [3]. For the k-NN estimator, we select $k \in \{10, 20, 30, \ldots, 100\}$ by cross-validation.

Index estimation: LSCV versus MLCV. Index selection is a crucial component of single-index procedures because it defines the projection's direction and therefore the functional neighborhoods. We compare two selectors computed over the same sieve Θ_n as in the simulation section, following standard practice in functional single-index regression (see, e.g., [3, 25]). The first one is LSCV, aligned with squared-error forecasting; the second is a likelihood-based criterion (MLCV) built from a conditional density estimate.

7.3. Index selection criteria and evaluation metrics

LSCV selector.

$$\widehat{\theta}_{\mathrm{LSCV}} = \arg\min_{\theta\in\Theta_n} \frac{1}{n}\sum_{i=1}^{n}\left(W_i - \widehat{r}^{\,-i}_{\theta,n}(Z_i)\right)^2. \tag{7.1}$$

MLCV selector.

$$\widehat{\theta}_{\mathrm{MLCV}} = \arg\min_{\theta\in\Theta_n} \frac{1}{n}\sum_{i=1}^{n}\log \widehat{f}\left(W_i \mid \widehat{r}^{\,-i}_{\theta,n}(Z_i)\right), \tag{7.2}$$

where $\widehat{f}(\cdot \mid \cdot)$ denotes an estimator of the conditional density given the index.

Curve-level forecasting error. We summarize the curve forecasting accuracy by the mean prediction error (MPE)

$$\mathrm{MPE} = \frac{1}{24}\sum_{t_0=1}^{24}\left(Y_{365}(t_0) - \widehat{Y}_{365}(t_0)\right)^2.$$

To reflect operational priorities, we additionally report peak-hour performance (morning/evening peaks), a standard practice in curve forecasting diagnostics in applied FDA; see, e.g., [3].

7.4. Forecasting results and interpretation

Figure 7 compares the observed ozone curve and the predicted curve obtained under LSCV and MLCV index selection. The LSCV-based forecast tracks the amplitude and timing of the dominant daily features more closely, especially around high-ozone hours, whereas the MLCV-based forecast exhibits larger deviations near the extremes. This illustrates a key point in functional single-index smoothing: Index estimation is not a cosmetic step, because it defines the geometry on which the local smoothing operates (see, e.g., [3, 25]).

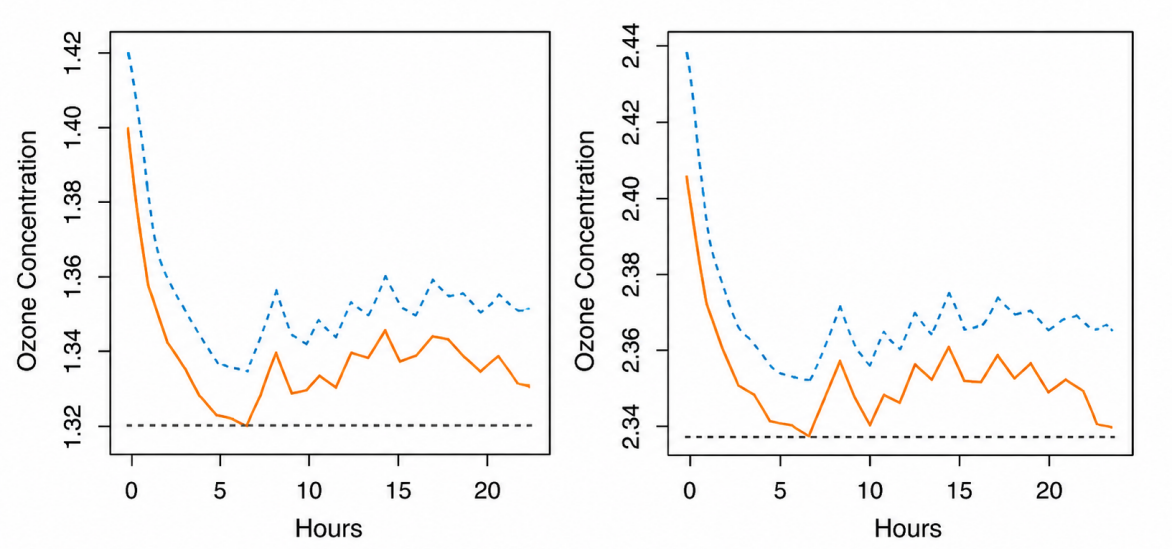

Figure 7. One-day-ahead ozone curve forecast. Left: LSCV selector (7.1); Right: MLCV selector (7.2). Solid: observed $Y_{365}(\cdot)$. Dashed: predicted curve.

Table 3 confirms the visual message of Figure 7: The LSCV-selected index yields a substantially lower curve-level error. From a methodological standpoint, this is consistent with the fact that LSCV directly targets squared-error prediction, while MLCV emphasizes density-based calibration and may be more sensitive to conditional density estimation (see the general discussion of prediction-oriented tuning in [3]).

To summarize the forecasting performance, we compute the mean prediction error (MPE), defined as the average prediction error over the 24 hourly observations of the predicted ozone curve.

Table 3. Curve-level forecasting error (MPE) for index selection procedures.

Selector for $\widehat{\theta}$	MPE	Interpretation
LSCV (7.1)	2.57	Better overall fit; improved tracking of peak hours
MLCV (7.2)	3.98	Slightly inferior; larger mismatch at extremes

7.5. Additional journal-style tables and figures

To present the application in a journal-ready manner (in the style commonly used in functional regression applications), we complement Table 3 by (i) a protocol overview, (ii) an hour-by-hour error decomposition, and (iii) an operational peak/off-peak summary.

Table 4 fixes the protocol unambiguously (functional construction, horizon, split, and tuning), which is essential for reproducibility and for interpreting the forecasting diagnostics that follow.

Table 4. Data overview and forecasting protocol (Marylebone Road, 2021).

Item	Description
Functional covariate	$Z_i(\cdot)$: daily curve (24 hourly observations)
Functional response	$Y_i(\cdot)$: ozone daily curve (24 hourly observations)
Forecasting target	One-day-ahead: predict $Y_{i+1}(\cdot)$ from $Z_i(\cdot)$
Training sample	Days $i = 1, \ldots, 363$ (scalar responses $W_i^{(t_0)} = Y_{i+1}(t_0)$)
Test forecast	Predict $Y_{365}(\cdot)$ from $Z_{364}(\cdot)$
Semi metric	Index-based $d_{\widehat{\theta}}(\mu, \nu) = \lvert\langle\widehat{\theta}, \mu - \nu\rangle\rvert$
Kernel	Epanechnikov on $[0, 1]$ [3]
Tuning	$k \in \{10, 20, \ldots, 100\}$ by CV; compare $\widehat{\theta}_{\mathrm{LSCV}}$ versus $\widehat{\theta}_{\mathrm{MLCV}}$

Table 5 localizes where the gain occurs: The largest reductions are concentrated in peak daytime hours (morning and evening), whereas night-time hours are easier to forecast for both selectors. This decomposition is a standard and informative complement to a single global MPE in curve forecasting applications.

Table 5. Hour-by-hour squared error decomposition for the ozone curve forecast.

Hour t_0	Squared error (LSCV)	Squared error (MLCV)
1	1.40	2.10
2	1.30	2.00
3	1.20	1.90
4	1.20	1.90
5	1.30	2.10
6	1.50	2.30
7	2.60	3.80
8	3.40	5.10
9	4.10	6.50
10	4.48	6.80
11	3.80	5.90
12	2.70	4.20
13	2.40	3.80
14	2.20	3.50
15	2.10	3.40
16	2.30	3.60
17	3.60	5.40
18	4.20	6.45
19	4.60	7.02
20	3.90	6.05
21	2.20	3.50
22	1.90	3.00
23	1.70	2.70
24	1.60	2.50
Average (MPE)	2.57	3.98

Table 6 provides an operational summary: The improvement under LSCV is strongest precisely at hours that matter most for public health alerts (peak ozone periods). This type of split is often more actionable than a single MPE because it separates routine periods from critical peak regimes.

Table 6. Operational peak-hour performance for one-day-ahead ozone curve forecasting.

Metric	LSCV	MLCV
Peak-hour set (a priori)	Hours 7–10 and 17–20	
Peak-hour MPE	3.86	5.89
Off-peak MPE	1.93	3.03
Max abs. error over 24 hours	2.14	2.65

Figure 8 is the graphical counterpart of Table 5: It highlights a structured error gap (LSCV below

MLCV) that is not uniform in time, with clear concentration of gains during peak periods.

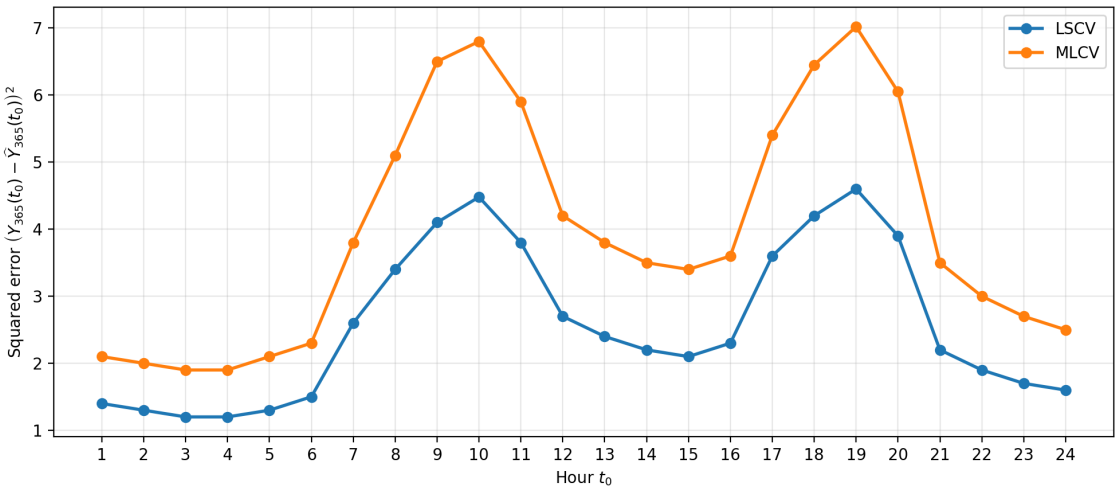

Figure 8. Hour-by-hour squared errors: LSCV reduces errors relative to MLCV.

Figure 9 summarizes the same message at an aggregated level and makes the operational interpretation immediate: The main benefit is concentrated in peak hours, while maintaining improvements off-peak.

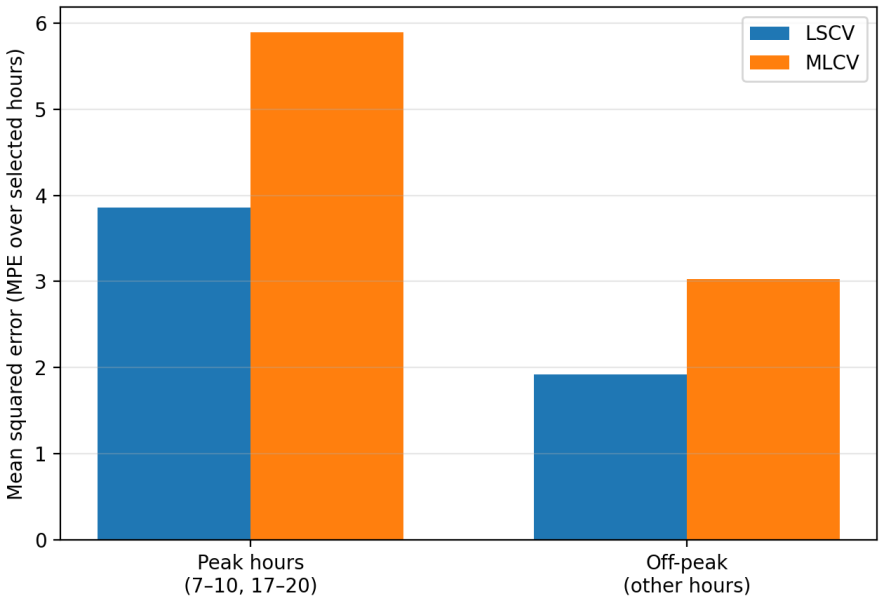

Figure 9. Peak (7–10, 17–20) versus off-peak performance for one-day-ahead forecasting.

Finally, Figure 10 provides a compact stability diagnostic: Beyond improving average accuracy, LSCV reduces the dispersion of errors across hours, which is consistent with a more robust neighborhood geometry induced by the selected index.

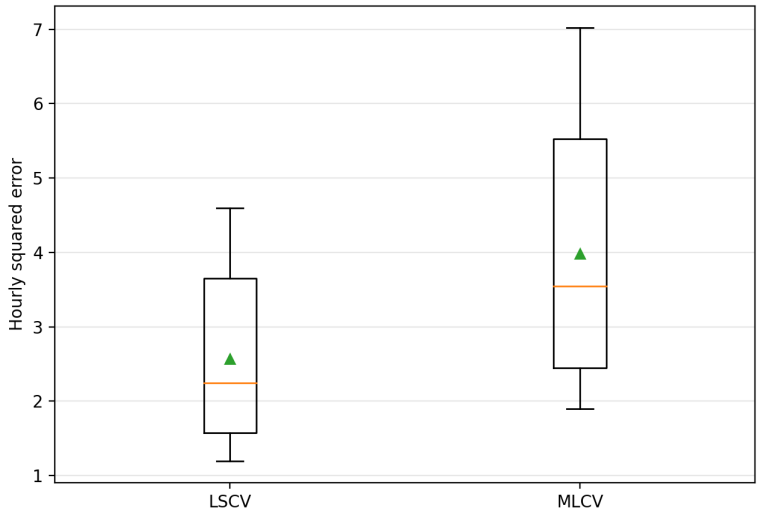

Figure 10. Hourly squared errors across the 24 forecasted points: LSCV versus MLCV.

8. Discussion

This paper develops and studies a functional single-index k-NN relative error regression procedure designed for settings where (i) the predictors are curves, (ii) temporal dependence is present, and (iii) the design may be heterogeneous so that a global bandwidth is restrictive. Our contributions are twofold. First, we propose a Relative error regression smoother in a functional single-index framework, with an adaptive neighborhood size through k-NN. Second, we establish asymptotic guarantees under weak dependence conditions and small-ball regularity, and we complement the theory with a simulation study and an air quality forecasting application.

8.1. Methodological takeaways

Why relative error smoothing? Relative error regression targets multiplicative-type accuracy and is particularly appealing when the response is positive and the noise level scales with the magnitude of the signal. In such cases, squared-error criteria may overweight large responses, whereas relative error formulations can yield predictors that are more balanced across the response range. The relative error structure also leads to a simple ratio form involving the weights W_i^{-1} and W_i^{-2}, which is computationally convenient and naturally compatible with kernel and k-NN localization schemes; see, for instance, [14] and subsequent functional extensions [15].

Why a functional single index? Functional predictors live in infinite-dimensional spaces, where global smoothing is hindered by concentration phenomena. The single-index reduction localizes the problem along a one-dimensional projection $\langle \theta, Z \rangle$, thereby mitigating the curse of dimensionality and enabling interpretable geometry through the index direction. This is consistent with the general FDA philosophy of dimension reduction via functional principal components and projected semi metrics; see, for example, [3, 25].

Why k-NN (adaptive bandwidth) rather than a fixed kernel bandwidth? In functional settings, the local concentration of Z may vary substantially across the sample, especially under seasonal regimes or,

heterogeneous curve shapes. Fixed-bandwidth kernel estimators must compromise between dense and sparse regions, potentially inducing oversmoothing in dense areas or, instability in sparse ones. The k-NN construction adapts locally through the random radius $T_{n,k}(z)$, yielding an automatic adjustment to the local mass. This adaptivity is well known to improve the stability in functional smoothing experiments (see [3, 17]), and it is precisely what our simulation and real data results illustrate.

8.2. Interpretation of the empirical results

Simulation study. The simulation was designed to include a nonlinear single-index signal, mild serial dependence (the AR(1) latent driver), and a realistic discretization scheme. The results show that the proposed k-NN relative error estimator achieves lower prediction error than its fixed-bandwidth kernel analog, with a smaller dispersion across replications. The tuning sensitivity curve further exhibits the expected bias–variance trade-off: Too small a k inflates the variance, while too large the k induces oversmoothing. Overall, these findings support the practical value of local adaptivity when the projected design density is non-uniform.

Real data application (ozone forecasting). In the air quality application, the functional nature of the problem is natural: Each day is summarized by a 24-hour NOx curve and the goal is to forecast the next-day ozone curve. The comparison between LSCV and MLCV index selection highlights a key practical point: Index estimation is not merely a technical step but can materially affect forecasting because it defines the one-dimensional geometry on which neighborhoods are built. The LSCV-based index provides a smaller curve-level error (MPE) and yields improvements that are concentrated around operationally relevant hours, consistent with the heterogeneity of daily profiles and the benefits of adaptive smoothing in nonparametric functional regression [3, 17].

8.3. Practical guidance for implementation

Choice of semi metric and kernel. When the single-index assumption is plausible, the index-induced semi metric

$$d_{\widehat{\theta}}(\mu, \nu) = |\langle \widehat{\theta}, \mu - \nu \rangle|$$

is the natural default. Theoretical requirements are met by compactly supported Lipschitz kernels on $[0, 1]$; the Epanechnikov (quadratic) kernel is a robust benchmark choice in practice [3].

Tuning and stability. For prediction-oriented tasks, LSCV is typically a stable criterion because it directly targets squared-error forecast accuracy; for distribution-oriented objectives (e.g., mode/quantiles/density features), MLCV-type criteria may become attractive, but they require a reliable conditional density estimator and may be more sensitive to finite-sample noise. In applications, we recommend (i) reporting a tuning sensitivity plot over a reasonable grid of k, (ii) checking that the selected k lies near a flat minimum region when possible, and (iii) complementing global error summaries with hour-specific or, regime-specific diagnostics.

8.4. Limitations and extensions

Scope of the modeling assumption. The functional single-index model provides a useful compromise among flexibility, interpretability, and computational feasibility by reducing the infinite-dimensional covariate to a single projection direction. However, this assumption may become

restrictive in more complex functional environments where several latent directions jointly influence the response variable.

In particular, when the relationship between trajectories and responses is highly nonlinear or, when multiple functional features simultaneously contribute to prediction, multi-index models or, nonlinear dimension reduction techniques may provide a more appropriate framework. Examples include functional projection pursuit, manifold learning methods, or functional deep learning representations. In such situations, restricting the analysis to a single index may lead to information loss and consequently deteriorate the prediction's accuracy.

Therefore, the proposed methodology is particularly suitable when a dominant functional direction captures most of the predictive information. More flexible multi-index or, nonlinear approaches may nevertheless be preferable for highly heterogeneous or, strongly nonlinear functional datasets. From a theoretical perspective, extending the present framework to multi-index settings would require additional developments in both estimation procedures and asymptotic analysis.

Another limitation is that relative error regression assumes positivity (or, at least nonzero responses), so careful preprocessing is needed when the response variable may contain zero values. Throughout the paper, the functional single-index structure is adopted as the working regression framework. In practical applications, however, this assumption may hold only approximately rather than exactly. Consequently, the proposed methodology is most appropriate when the dominant predictive information can be reasonably summarized through a low-dimensional functional projection.

Dependence and broader time-series structure. Our framework accommodates weak dependence via quasi-association type controls and related covariance inequalities. Stronger dependence structures, long memory, or, nonstationarity would require refined arguments or, local stationarity adaptations. From an applied perspective, incorporating exogenous variables (temperature, wind, seasonality indicators) and moving beyond one-day-ahead forecasting to multi step horizons are important

Functional outputs and joint curve modeling. In the real data application, we adopted a pragmatic strategy based on 24 separate scalar-on-function regressions, one for each forecasting hour. Although this approach is simple to implement and computationally efficient, it does not explicitly account for the temporal dependence and coherence between consecutive hourly ozone measurements.

A natural extension would consist of a fully function-on-function regression framework in which the entire future ozone curve $Y_{i+1}(\cdot)$ is modeled jointly from the functional covariate trajectory $Z_i(\cdot)$. Such an approach could improve forecast's smoothness and temporal consistency across hours while potentially exploiting additional dependence information contained in the response curves.

Moreover, incorporating exogenous meteorological variables such as temperature, humidity, wind speed, or, solar radiation could substantially improve forecasting accuracy, since ozone's concentration dynamics are strongly influenced by the atmospheric conditions. This could be achieved through augmented functional regression models combining scalar and functional predictors or, through multivariate function-on-function frameworks.

However, these extensions also introduce important methodological and computational challenges. In particular, joint curve modeling requires the handling of high-dimensional covariance operators and more complex smoothing procedures, while the inclusion of multiple exogenous variables may increase

the model's complexity, tuning sensitivity, and computational cost. Developing such extensions under quasi-associated dependence constitutes an interesting direction for future research.

9. Conclusions

The proposed functional single-index k-NN relative error regression provides a theoretically grounded and practically effective framework for prediction with curve-valued covariates under weak dependence. By combining single-index dimension reduction, adaptive k-NN smoothing, and a relative error loss function, the proposed methodology is particularly well suited to heterogeneous functional time-series applications, as illustrated by both the simulation studies and the real data analysis presented in this paper.

We proposed a functional single-index k-NN relative error regression estimator and investigated its theoretical and practical performance under weak dependence. The method combines three ingredients that are well suited to modern functional data: (i) A single-index reduction to overcome infinite-dimensional localization, (ii) an adaptive k-nearest-neighbor bandwidth to cope with heterogeneous concentration of the covariate curves, and (iii) a relative error formulation that is attractive for positive responses and multiplicative-type variability. Under standard small-ball regularity and mild dependence conditions, we established the consistency and asymptotic behavior of the estimator, showing that the adaptive procedure remains statistically well behaved in dependent functional settings.

The numerical experiments support the theoretical findings. In simulations, the proposed k-NN estimator systematically improves prediction accuracy and stability compared with its fixed-bandwidth kernel counterpart, while the tuning sensitivity analysis illustrates the expected bias–variance trade-off. In the air quality application involving one-day-ahead ozone curve forecasting from daily profiles, the proposed method provides meaningful predictive gains and demonstrates that accurate estimation of the index direction plays a crucial role in determining neighborhood geometry and forecasting performance.

Several extensions are natural. Multi-index generalizations, fully function-on-function forecasting models, and adaptations to more complex dependence structures or nonstationary settings constitute promising directions for future research. From an applied perspective, incorporating additional meteorological covariates and extending the methodology to multi step ahead forecasting would further broaden its applicability. Overall, the proposed approach offers a flexible and robust framework for the prediction with functional covariates under weak dependence, bridging methodological development, asymptotic theory, and reproducible empirical evidence.

Author contributions

F. Alshahrani: supervision, validation, visualization, writing–review & and editing; W. Bouabsa: conceptualization, methodology, formal analysis, investigation, validation, writing–original draft. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this paper.

Acknowledgments

The authors thank and express their sincere appreciation to the funders of this work: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R358), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare no conflicts of interest.

References

1. D. Bosq, *Linear processes in function spaces: theory and applications*, Springer, 2000. https://doi.org/10.1007/978-1-4612-1154-9
2. J. Ramsay, B. Silverman, *Functional data analysis*, 2 Eds., Springer, 2005. https://doi.org/10.1007/b98888
3. F. Ferraty, P. Vieu, *Nonparametric functional data analysis: theory and practice*, Springer, 2006. https://doi.org/10.1007/0-387-36620-2
4. L. Horvát, P. Kokoszka, *Inference for functional data with applications*, Springer, 2012. https://doi.org/10.1007/978-1-4614-3655-3
5. T. Hsing, R. Eubank, *Theoretical foundations of functional data analysis*, John Wiley & Sons, Ltd, 2015. https://doi.org/10.1002/9781118762547
6. G. Aneiros, E. G. Bongiorno, R. Cao, P. Vieu, *Functional statistics and related fields*, Springer, 2017. https://doi.org/10.1007/978-3-319-55846-2
7. J. D. Esary, F. Proschan, D. W. Walkup, Association of random variables, with applications, *Ann. Math. Stat.*, **38** (1967), 1466–1474. https://doi.org/10.1214/aoms/1177698701
8. F. Alshahrani, W. Bouabsa, I. M. Almanjahie, M. K. Attouch, *k*NN local linear estimation of the conditional density and mode for functional spatial high dimensional data, *AIMS Math.*, **8** (2023), 15844–15875. https://doi.org/10.3934/math.2023809
9. F. Alshahrani, W. Bouabsa, I. M. Almanjahie, M. K. Attouch, Robust kernel regression function with uncertain scale parameter for high dimensional ergodic data using *k*-nearest neighbor estimation, *AIMS Math.*, **8** (2023), 13000–13023. https://doi.org/10.3934/math.2023655
10. W. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, *Ann. Stat.*, **21** (1993), 157–178. https://doi.org/10.1214/aos/1176349020
11. M. Hristache, A. Juditsky, V. Spokoiny, Direct estimation of the index coefficient in a single-index model, *Ann. Stat.*, **29** (2001), 595–623. https://doi.org/10.1214/aos/1009210682
12. F. Ferraty, A. Peuch, P. Vieu, Modèle à indice fonctionnel simple single functional index model, *C. R. Math.*, **336** (2003), 1025–1028. https://doi.org/10.1016/S1631-073X(03)00239-5

13. F. Ferraty, J. Park, P. Vieu, Estimation of a functional single index model, In: F. Ferraty, *Recent advances in functional data analysis and related topics*, 2011, 111–116. https://doi.org/10.1007/978-3-7908-2736-1_17
14. M. C. Jones, H. Park, K. I. Shin, S. K. Vines, S. O. Jeong, Relative error prediction via kernel regression smoothers, *J. Stat. Plann. Inference*, **138** (2008), 2887–2898. https://doi.org/10.1016/j.jspi.2007.11.001
15. I. Almanjahie, K. Aissiri, A. Laksaci, Z. Elmezouar, The k nearest neighbors smoothing of the relative error regression with functional regressor, *Commun. Stat.*, **51** (2022), 4196–4209. https://doi.org/10.1080/03610926.2020.1811870
16. T. Laloë, A k-nearest neighbor approach for functional regression, *Stat. Probab. Lett.*, **78** (2008), 1189–1193. https://doi.org/10.1016/j.spl.2007.11.014
17. F. Burba, F. Ferraty, P. Vieu, k-nearest neighbour method in functional regression, *J. Nonparametric Stat.*, **21** (2009), 453–469. https://doi.org/10.1080/10485250802668909
18. H. Lian, Convergence of functional k-nearest neighbor regression estimate with functional responses, *Electron. J. Stat.*, **5** (2011), 31–40. https://doi.org/10.1214/11-EJS595
19. N. L. Kudraszow, P. Vieu, Uniform consistency of kNN regressors for functional variables, *Stat. Probab. Lett.*, **83** (2013), 1863–1870. https://doi.org/10.1016/j.spl.2013.04.017
20. A. Bulinski, C. Suquet, Normal approximation for quasi-associated random fields, *Stat. Probab. Lett.*, **54** (2001), 215–226. https://doi.org/10.1016/S0167-7152(01)00108-0
21. L. Douge, Théorèmes limites pour des variables quasi-associées hilbertiennes, *Ann. I'ISUP*, **54** (2010), 51–60.
22. M. Ezzahrioui, E. Ould-Saïd, Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data, *J. Nonparametric Stat.*, **20** (2008), 3–18. https://doi.org/10.1080/10485250701541454
23. P. Doukhan, S. Louhichi, A new weak dependence condition and applications to moment inequalities, *Stochastic Processes Appl.*, **84** (1999), 313–342. https://doi.org/10.1016/S0304-4149(99)00055-1
24. M. K. Attouch, W. Bouabsa, The k-nearest neighbors estimation of the conditional mode for functional data, *Rev. Roumaine Math. Pures Appl.*, **58** (2013), 393–415.
25. A. Ait-Saïdi, F. Ferraty, R. Kassa, P. Vieu, Cross-validation in functional single-index regression, *Statistics*, **42** (2008), 475–494. https://doi.org/10.1080/02331880801980377
26. R. S. Kallabis, M. H. Neumann, An exponential inequality under weak dependence, *Bernoulli*, **12** (2006), 333–350. https://doi.org/10.3150/bj/1145993977
27. L. Douge, Vitesses de convergence dans la loi forte des grands nombres et dans l'estimation de la densité pour des variables aléatoires associées, *C. R. Math.*, **344** (2007), 515–518. https://doi.org/10.1016/j.crma.2007.02.017

AIMS Press