



**HAL**  
open science

## **Scalar-on-Function Mode Estimation Using Entropy and Ergodic Properties of Functional Time Series Data**

Mohammed Alamari, Fatimah Almulhim, Ibrahim Almanjahie, Salim Bouzebda,  
Ali Laksaci

► **To cite this version:**

Mohammed Alamari, Fatimah Almulhim, Ibrahim Almanjahie, Salim Bouzebda, Ali Laksaci. Scalar-on-Function Mode Estimation Using Entropy and Ergodic Properties of Functional Time Series Data. *Entropy*, 2025, 27 (6), pp.552. <10.3390/e27060552>. <hal-05084108>

**HAL Id: hal-05084108**

**<https://hal.science/hal-05084108v1>**

Submitted on 15 Jan 2026

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Article

# Scalar-on-Function Mode Estimation Using Entropy and Ergodic Properties of Functional Time Series Data

Mohammed B. Alamari <sup>1</sup>, Fatimah A. Almulhim <sup>2</sup>, Ibrahim M. Almanjahie <sup>1</sup>, Salim Bouzebda <sup>3,\*</sup>  
and Ali Laksaci <sup>1</sup>

<sup>1</sup> Department of Mathematics, College of Science, King Khalid University, Abha 62223, Saudi Arabia

<sup>2</sup> Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

<sup>3</sup> Université de Technologie de Compiègne, LMAC (Laboratory of Applied Mathematics of Compiègne), 60203 Compiègne, France

\* Correspondence: salim.bouzebda@utc.fr

**Abstract:** In this paper, we investigate the recursive  $L^1$  estimator of the conditional mode when the input variable takes values in a pseudo-metric space. The new proposed estimator is constructed under an ergodicity assumption, which provides a robust alternative to the standard mixing processes in various practical settings. The particular interest of this contribution arises from the difficulty in incorporating the mathematical properties of a functional mixing process. In contrast, ergodicity is characterized by the Kolmogorov–Sinai entropy, which measures the dynamics, the sparsity, and the microscopic fluctuations of the functional process. Using an observation sampled from ergodic functional time series (fts), we establish the asymptotic properties of this estimator. In particular, we derive its convergence rate and show Borel–Cantelli (BC) consistency. The general expression for the convergence rate is then specialized to several notable scenarios, including the independence case, the classical kernel method, and the vector-valued case. Finally, numerical experiments on both simulated and real-world datasets demonstrate the superiority of the  $L^1$ -recursive estimator compared to existing competitors.

**Keywords:**  $L^1$ -modal regression; functional data; ergodic data; recursive estimate; nonparametric prediction; complete consistency; conditional mode; quantile regression

**MSC:** 62G08; 62G10; 62G35; 62G07; 62G32; 62G30; 62H12



check for updates

Academic Editor: José María Amigó

Received: 18 April 2025

Revised: 14 May 2025

Accepted: 22 May 2025

Published: 24 May 2025

**Citation:** Alamari, M.B.; Almulhim, F.A.; Almanjahie, I.M.; Bouzebda, S.; Laksaci, A. Scalar-on-Function Mode Estimation Using Entropy and Ergodic Properties of Functional Time Series Data. *Entropy* **2025**, *27*, 552. <https://doi.org/10.3390/e27060552>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Investigating the joint behavior of two random variables in a functional setting is an active area of applied statistics, as it facilitates quantifying the influence of a functional covariate on a scalar response. Numerous functional approaches have been proposed to capture this relationship, including conditional expectation, relative regression, and median regression. However, modeling the relationship via the conditional distribution function is often regarded as more informative because it sheds light on both central and extreme parts of the response. Consequently, the principal goal of this work is to introduce a new estimator for modal regression using the cumulative distribution function.

Despite substantial literature on conditional mode prediction, the predominant estimator remains the Nadaraya–Watson (NW) method. The earliest investigation of conditional mode estimation can be traced back to [1], which demonstrated that the mode can yield superior predictive performance compared with the conditional mean. In [2], the authors

proposed a mode-based predictor using the derivative of the conditional density, applicable to vector-valued input variables. Subsequently, ref. [3] derived the asymptotic distribution of the modal regression estimator under independence, and this result was generalized to dependent data by [4]. For more recent works, we refer the readers to [5].

A functional version of the NW-estimator for the conditional mode (CM) was first introduced in [6], where the authors established almost complete consistency of the estimator by identifying it as the maximizer of the conditional density. This result was extended to dependent processes in [7]. The asymptotic distribution of the NW-based functional CM estimator was studied under the i.i.d. assumption in [8], whereas [9] considered strong mixing functional time series under fractal conditions. The monograph of [10] represents a key contribution to nonparametric functional prediction, and further theoretical results on functional mode estimation can be found in [11], which addressed the  $L^p$ -convergence of NW-based functional mode estimators. In the context of ergodic functional time series, ref. [12] focused on conditional mode estimation and derived BC consistency under a missing-at-random framework for the functional covariate. Several alternative estimators to the NW approach have also emerged in functional data analysis (FDA). For example, ref. [13] established the asymptotic normality of a local linear CM estimator in the functional setting, ref. [14] investigated a kNN-based functional CM approach, and [15] developed a local linear functional-kNN version of the estimator. More extensive studies on functional CM estimation can be found in [16–18] and related references. Additionally, in the ergodic functional time series case, ref. [19] obtained BC consistency for the conditional mode estimator.

A distinctive contribution of the present paper is its focus on a recursive estimation algorithm, a direction that remains underexplored in FDA. One of the earliest examinations of recursive methods in this domain is [20], which addressed the recursive estimation of conditional mean functions. Later, ref. [21] investigated recursive procedures for functional time series under mixing conditions. More recent developments in functional nonparametric smoothing by means of recursive algorithms, along with relevant references, are presented in [22]. For additional perspectives on FDA and its applications, including dedicated survey articles and specialized journal issues, see [23–26], among others and recent papers [27–29].

### 1.1. Contributions of This Paper

The primary objective of this work is to propose a novel modal regression estimator and establish its asymptotic properties under a general framework of ergodic functional time series. Specifically, our estimator combines an  $L^1$ -based approach with a recursive procedure. In contrast to estimators built upon the NW or local linear methods, the newly constructed estimator offers multiple advantages. First, incorporating an  $L^1$ -technique promotes robustness, which mitigates the impact of outliers through a percentile-based approach. Moreover, harnessing the conditional distribution function to identify the conditional mode leverages comprehensive information about the functional covariate–response relationship, potentially enhancing the estimator’s precision. A further strength lies in the recursive structure, which seamlessly updates the estimator upon the arrival of each new data point—a feature especially valuable for real-time forecasting in ergodic functional time series. This adaptability is highly relevant in fields such as artificial intelligence, where continuous data processing is critical. From a theoretical standpoint, we derive the Borel–Cantelli convergence rate of the proposed estimator under mild ergodic conditions frequently satisfied by common processes (e.g., moving average (MA), generalized autoregressive conditional heteroskedasticity (GARCH), Volterra). Finally, we illustrate the

practical value of our algorithm through empirical investigations on both synthetic and real-world datasets.

### 1.2. Paper Organization

We introduce the  $L^1$ -based conditional mode and its recursive estimator in Section 2. The main theoretical results, including consistency and convergence rates, are presented in Section 3. Section 4 is devoted to some discussion on the real impact of the principal axes of the studied topic. In Section 5, we investigate the finite-sample performance of the proposed estimator through simulation studies and applications to real data. The proofs of the auxiliary results are provided in Section 6.

## 2. The $L^1$ -Recursive Estimation of the Mode

Consider a strictly stationary sequence of dependent input–output random variables, denoted by  $(I_i, O_i)_{i=1, \dots, n}$ , which takes values in  $\mathcal{F} \times \mathbb{R}$ . Here,  $\mathcal{F}$  is a semi-metric space endowed with a semi-metric  $d(\cdot, \cdot)$ . Let  $N_\theta$  be a neighborhood of a fixed curve  $\theta \in \mathcal{F}$ . We assume that the conditional distribution function  $G(\cdot | \theta)$  is strictly increasing and admits a continuous density  $g(y | \theta)$  with respect to the Lebesgue measure on  $\mathbb{R}$ . Recall that for a given  $p \in (0, 1)$ , the conditional quantile of  $O$  given  $I = \theta$ , denoted by  $Qu_p(\theta)$ , is obtained by inverting the conditional distribution function, namely

$$Qu_p(\theta) = G^{-1}(p | \theta).$$

Meanwhile, the conditional mode of  $O$  given  $I = \theta$ , denoted by  $CM(\theta)$ , is defined as the maximizer of the conditional density on a given compact set  $K \subset \mathbb{R}$ :

$$CM(\theta) = \arg \max_{y \in K} g(y | \theta).$$

By combining these two notions, one can re-express modal regression as

$$CM(\theta) = Qu_{p_\theta}(\theta) \quad \text{with} \quad p_\theta = \arg \min_{p \in G^{-1}(K|\theta)} \frac{\partial}{\partial p} Qu_p(\theta). \tag{1}$$

In the rest of this paper, we assume that the compact subset  $K$  locates one conditional mode  $CM(\theta)$  and that (1) holds.

The  $L^1$ -estimator of the conditional mode naturally connects to the  $L^1$ -quantile regression, which is determined by

$$Qu_p(\theta) = \arg \min_{t \in \mathbb{R}} \Psi_p(\theta, t),$$

where

$$\Psi_p(\theta, t) = \mathbb{E}[L_p(O - t) | I = \theta] \quad \text{and} \quad L_p(s) = (2p - 1)s + |s|.$$

An  $L^1$ -recursive estimator of the function  $Qu_p(\cdot)$  can thus be defined by

$$\widehat{Qu}_p(\theta) = \arg \min_{t \in \mathbb{R}} \widehat{\Psi}_p(\theta, t), \tag{2}$$

where

$$\widehat{\Psi}_p(\theta, t) = \frac{\sum_{i=1}^n \Gamma(a_i^{-1} d(\theta, I_i)) [(2p - 1)(O_i - t) + |O_i - t|]}{\sum_{i=1}^n \Gamma(a_i^{-1} d(\theta, I_i))}, \quad \forall t \in \mathbb{R},$$

where  $\Gamma$  is a kernel function, and  $\{a_i\}$  is a sequence of positive real numbers satisfying  $\lim_{n \rightarrow \infty} a_n = 0$ . Note that in the recursive estimation setting, in contrast to the Nadaraya–Watson-type (NW) approach, the bandwidth  $a_i$  depends on the specific input observations  $\{I_i\}$ , thereby allowing the estimator to be updated whenever a new observation is obtained.

Before constructing a  $L^1$ -recursive estimator of the modal regression, it is necessary to define estimators for both  $p_\theta$  and  $Qu'_p(\theta)$ . Recall that

$$Qu'_p(\theta) = \frac{\partial}{\partial p} Qu_p(\theta) = \lim_{b \rightarrow 0} \frac{Qu_{p+b}(\theta) - Qu_p(\theta)}{b}.$$

A natural estimator for  $Qu'_p(\theta)$  is then given by

$$\widehat{Qu}'_p(\theta) = \frac{\widehat{Qu}_{p+h_n}(\theta) - \widehat{Qu}_{p-h_n}(\theta)}{2h_n},$$

where  $\{h_n\}$  is a sequence of positive real numbers converging to 0. The conditional mode  $CM(\theta)$  is accordingly estimated by

$$\widehat{CM}(\theta) = \widehat{Qu}_{\widehat{p}_\theta}(\theta), \tag{3}$$

where

$$\widehat{p}_\theta = \arg \min_{p \in G^{-1}(K|\theta)} \widehat{Qu}'_p(\theta).$$

Of course,  $\widehat{CM}(\theta)$  is not necessarily unique, and if that is the case,  $\widehat{CM}(\theta)$  will concern any values verifying (3).

In the theoretical development, our principal goal is to establish a Borel–Cantelli convergence result for  $\widehat{CM}(\theta)$ . To achieve this, we adopt the classical ergodicity assumption, which is more general than ordinary mixing conditions. A process is typically considered ergodic if, over sufficient time, the entropy measured along a single evolving trajectory converges to the entropy of the system’s full ensemble of possible states. In the functional case, we employ the definition of ergodicity for functional statistics proposed by [7], as it provides a suitable framework for analyzing dependent functional data.

### 3. Main Results

We begin by letting  $C$  or  $C'$  denote generic strictly positive constants. We also assume that

$$G^{-1}(K | \theta) = [a_\theta, b_\theta].$$

Moreover, for each  $k = 1, \dots, n$ , define  $\mathfrak{G}_k$  as the  $\sigma$ -field generated by  $((I_1, O_1), \dots, (I_k, O_k), I_{k+1})$ , and let  $\mathfrak{F}_k$  be the  $\sigma$ -field generated by  $((I_1, O_1), \dots, (I_k, O_k))$ .

Our principal assumptions are as follows:

(Co1)

- (i) The function  $\zeta(\theta, a) := \mathbb{P}(I \in B(\theta, a))$  satisfies  $\zeta(\theta, a) > 0$  for every  $a > 0$ , where  $B(\theta, r) = \{\theta' \in \mathcal{F} : d(\theta', \theta) < r\}$ .
- (ii) For each  $i = 1, \dots, n$ , there exists a deterministic function  $\zeta_i(\theta, \cdot)$  such that almost surely,  $0 < \mathbb{P}(I_i \in B(\theta, a) \mid \mathfrak{F}_{i-1}) \leq \zeta_i(\theta, a)$ ,  $\forall a > 0$ , and  $\zeta_i(\theta, a) \rightarrow 0$  as  $a \rightarrow 0$ .
- (iii) For any positive sequence  $(a_i)_{i=1, \dots, n}$ , we have
 
$$\frac{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i) \mid \mathfrak{F}_{i-1})}{\sum_{i=1}^n \zeta(\theta, a_i)} \rightarrow 1 \quad \text{a.co.}$$

(Co2) The function  $Qu.(\theta)$  is three times continuously differentiable on  $[a_\theta, b_\theta]$ . In addition, suppose that  $G(\cdot \mid \cdot)$  satisfies the Lipschitz condition

$$\forall \theta_1, \theta_2 \in N_\theta, \quad \forall t_1, t_2 \in [a_\theta, b_\theta], \quad |G(t_1 \mid \theta_1) - G(t_2 \mid \theta_2)| \leq C d^b(\theta_1, \theta_2) + |t_1 - t_2|,$$

for some  $b > 0$ , where  $N_\theta$  is a neighborhood of  $\theta$ .

(Co3) The function  $\Gamma$  is supported on  $(0, 1)$  and fulfills

$$0 < C \mathbb{1}_{(0,1)} < \Gamma(t) < C' \mathbb{1}_{(0,1)} < \infty.$$

(Co4)

$$\lim_{n \rightarrow \infty} \frac{\zeta_n \ln n}{n^2 \zeta_n^2} = 0, \quad \text{where } \zeta_n = \sum_{i=1}^n \zeta_i(\theta, a_i) \quad \text{and} \quad \zeta_n = \frac{1}{n} \sum_{i=1}^n \zeta(\theta, a_i).$$

Clearly, conditions (Co1)–(Co4) are often encountered in nonparametric functional time series analysis. In particular, (Co1) describes the probabilistic concentration behavior of the input variable, including its conditional concentration with respect to the filtration, which underscores the impact of ergodicity on the asymptotic properties of the estimator. Assumption (Co2) is pivotal for the nonparametric structure of the model. Conditions (Co3) and (Co4) govern the behavior of the kernel  $\Gamma$  and the smoothing parameters  $a_n$  and  $h_n$ , ensuring the proper handling of the technical aspects of the estimator  $\widehat{CM}(\theta)$ . These requirements also allow us to express the convergence rate in a form analogous to the Nadaraya–Watson estimator, which can be viewed as arising from maximizing a double-kernel estimator of a conditional density. Thus, the assumptions under consideration effectively encompass the main components of the subject—namely, the model, the data structure, the correlation framework, and the convergence rate. These assumptions are not overly restrictive, especially given the complexity of the proposed functional time series model and the strength of the Borel–Cantelli (BC) type consistency. In fact, it is possible to establish a weaker form of consistency for the estimator under less stringent conditions. By employing techniques similar to those presented in [30], one can demonstrate weak consistency. Ultimately, the relationship between the assumptions and the resulting theoretical guarantees reflects a trade-off: stronger and more general results necessitate stronger assumptions. The next theorem establishes the almost-complete (a.co.) convergence of  $\widehat{CM}(\theta)$ .

**Theorem 1.** Suppose (Co1)–(Co4) hold. If

$$\inf_{p \in (0,1)} \frac{\partial^3}{\partial p^3} Qu_p(\theta) > 0,$$

then

$$\widehat{CM}(\theta) - CM(\theta) = O\left(\frac{1}{\zeta_n} \sum_{i=1}^n a_i^b \zeta_i(\theta, a_i)\right) + O(h_n) + O\left(\sqrt{\frac{\zeta_n \ln n}{n^2 \zeta_n^2}}\right), \quad a.co.$$

**Proof of the Main Result.** From the definitions of  $\widehat{CM}(\theta)$  and  $CM(\theta)$ , it follows that

$$\begin{aligned} |\widehat{CM}(\theta) - CM(\theta)| &= \left| \widehat{Qu}_{\widehat{p}_\theta}(\theta) - Qu_{p_\theta}(\theta) \right| \\ &= \left| \widehat{Qu}_{\widehat{p}_\theta}(\theta) - Qu_{\widehat{p}_\theta}(\theta) + Qu_{\widehat{p}_\theta}(\theta) - Qu_{p_\theta}(\theta) \right| \\ &\leq \sup_{p \in [a_\theta, b_\theta]} \left| \widehat{Qu}_p(\theta) - Qu_p(\theta) \right| \end{aligned} \tag{4}$$

$$+ \sup_{p \in [a_\theta, b_\theta]} \left| Qu_{\widehat{p}_\theta}(\theta) - Qu_{p_\theta}(\theta) \right|. \tag{5}$$

Using a Taylor expansion, we have

$$Qu_{\widehat{p}_\theta}(\theta) - Qu_{p_\theta}(\theta) = (\widehat{p}_\theta - p_\theta) \{Qu'\}_{p_\theta^*}(\theta), \quad \text{for some } p_\theta^* \in (\widehat{p}_\theta, p_\theta).$$

Since  $p_\theta$  is the minimizer of  $Qu'(\cdot | \theta)$ , we also obtain

$$Qu'_{\widehat{p}_\theta}(\theta) - Qu'_{p_\theta}(\theta) = (\widehat{p}_\theta - p_\theta)^2 \{Qu'''\}_{p_\theta^{**}}(\theta), \quad \text{for some } p_\theta^{**} \in (\widehat{p}_\theta, p_\theta).$$

Analogously to (4), it follows that

$$\left| Qu'_{\widehat{p}_\theta}(\theta) - Qu'_{p_\theta}(\theta) \right| \leq 2 \sup_{p \in [a_\theta, b_\theta]} \left| \widehat{Qu}'_p(\theta) - Qu'_p(\theta) \right|.$$

Because  $\inf_{p \in (0,1)} \frac{\partial^3}{\partial p^3} Qu_p(\theta) > 0$ , we obtain

$$\widehat{p}_\theta - p_\theta \leq C \sqrt{\sup_{p \in [a_\theta, b_\theta]} \left| \widehat{Qu}'_p(\theta) - Qu'_p(\theta) \right|}. \tag{6}$$

Combine (4) and (6) to obtain

$$\begin{aligned} &|\widehat{CM}(\theta) - CM(\theta)| \\ &\leq C \left( \sup_{p \in [a_\theta, b_\theta]} \left| \widehat{Qu}_p(\theta) - Qu_p(\theta) \right| + \sqrt{\sup_{p \in [a_\theta, b_\theta]} \left| \widehat{Qu}'_p(\theta) - Qu'_p(\theta) \right|} \right). \end{aligned} \tag{7}$$

Hence, determining the convergence rate reduces to studying

$$\sup_{p \in [a_\theta, b_\theta]} \left| \widehat{Qu}_p(\theta) - Qu_p(\theta) \right| \quad \text{and} \quad \sup_{p \in [a_\theta, b_\theta]} \left| \widehat{Qu}'_p(\theta) - Qu'_p(\theta) \right|.$$

Furthermore, asymptotically,

$$\begin{aligned} & \widehat{Qu}'_p(\theta) - Qu'_p(\theta) \\ &= \frac{\widehat{Qu}_{p+h_n}(\theta) - Qu_{p+h_n}(\theta) + Qu_{p-h_n}(\theta) - \widehat{Qu}_{p-h_n}(\theta)}{2h_n} \\ & \quad + \frac{Qu_{p+h_n}(\theta) - Qu_p(\theta) + Qu_p(\theta) - Qu_{p-h_n}(\theta) - 2h_n Qu'_p(\theta)}{2h_n} \\ &\leq C h_n^{-1} \sup_{q \in (a_\theta-h_n, b_\theta+h_n]} |\widehat{Qu}_q(\theta) - Qu_q(\theta)| + O(h_n). \end{aligned}$$

Finally, Theorem 1 follows from the following lemmas.  $\square$

**Lemma 1 ([16]).** Consider a family of real-valued random functions  $\{B_n\}$ , each of which is decreasing in  $\gamma$ . Let  $\{A_n\}$  be a real-valued random sequence. Suppose there exist positive constants  $\lambda, M > 0$  such that

$$A_n = o_{a.co.}(1) \quad \text{and} \quad \sup_{|\gamma| \leq M} |B_n(\gamma) + \lambda\gamma - A_n| = o_{a.co.}(1).$$

Then, for any real random sequence  $\{\gamma_n\}$  satisfying  $B_n(\gamma_n) = o_{a.co.}(1)$ , it follows that

$$\sum_{n=1}^{\infty} \mathbb{P}\{|\gamma_n| \geq M\} < \infty.$$

**Lemma 2.** Suppose (Co1) and (Co3)–(Co4) hold. Then, we have

$$\widehat{Q}_D(\theta) - \overline{Q}_D(\theta) = O\left(\sqrt{\frac{\zeta_n \ln n}{n^2 \zeta_n^2}}\right), \quad a.co.$$

Moreover, there exists a constant  $C > 0$  such that

$$\sum_n \mathbb{P}(\overline{Q}_D(\theta) < C) < \infty.$$

Here,

$$\widehat{Q}_D(\theta) := \frac{1}{n \zeta_n} \sum_{i=1}^n \Gamma(a_i^{-1} d(\theta, I_i))$$

and

$$\overline{Q}_D(\theta) := \frac{1}{n \zeta_n} \sum_{i=1}^n \mathbb{E}[\Gamma(a_i^{-1} d(\theta, I_i)) \mid \mathfrak{F}_{i-1}].$$

**Proposition 1.** Assume (Co1)–(Co4) hold. Then, there is a positive constant  $\lambda$  such that

$$\widehat{Qu}_p(\theta) - Qu_p(\theta) = \frac{1}{g(Qu_p \mid \theta)} A_n + O\left(\sup_{|\gamma| \leq M} |B_n(\gamma) + \lambda\gamma - A_n|\right),$$

where

$$B_n(\gamma) = \frac{1}{n \zeta_n} \sum_{i=1}^n [p - \mathbf{1}_{\{O_i \leq \gamma + Qu_p(\theta)\}}] \Gamma_i, \quad \text{and} \quad A_n = B_n(0).$$

**Proposition 2.** Under the same assumptions (Co1)–(Co4), we also have

$$\sup_{p \in (0,1)} |\widehat{Qu}_p(\theta) - Qu_p(\theta)| = O\left(\frac{1}{\zeta_n} \sum_{i=1}^n a_i^b \zeta_i(\theta, a_i)\right) + O\left(\sqrt{\frac{\zeta_n \ln n}{n^2 \zeta_n^2}}\right), \quad a.co.$$

## 4. Discussion and Comments

### 4.1. On the Ergodic Functional Time Series

Similarly to multivariate statistics, ergodicity plays a crucial role in functional statistics. In particular, ergodicity ensures that temporal averages converge to their corresponding stochastic means. This property is especially important, as it justifies the use of sample mean and covariance functions as consistent estimators of the true mean function and the true covariance operator. These estimators, in turn, allow for efficient estimation of eigenfunctions in functional principal component analysis (FPCA) and for accurate curve smoothing using a chosen basis, such as splines or Fourier functions. All these methodologies fundamentally rely on the sample mean and the empirical covariance operator. The ergodic behavior and functional characteristics of the time series under consideration are governed by assumption **(Co1)**, which quantifies the concentration properties of the functional variables. This assumption is thoroughly discussed in [6], where it is shown that **Co1(i)** holds for a wide class of continuous processes whose probability measures are absolutely continuous with respect to the Wiener measure. Examples include the Poisson process, the Ornstein–Uhlenbeck process, fractional Brownian motion, and general diffusion processes. In this work, we also focus on the conditional version of this assumption, namely **(Co1)(ii–iii)**. This extension enables us to account for the dependence structure of the process by analyzing its long-memory behavior, a standard approach in dynamic systems modeling and time series analysis. In such contexts, conditional distributions with respect to the past filtration  $\mathcal{F}_{i-1}$  are frequently employed to control process evolution, verify the martingale property, and assess predictability. Using arguments similar to those used in the unconditional case **(Co1)(i)**, one can show that a trivial example of a functional ergodic process satisfying **(Co1)(ii–iii)** is when its conditional distribution, given the past, is absolutely continuous with respect to the Wiener measure. Additionally, the Karhunen–Loève decomposition can be used to represent such processes explicitly (see [6] for examples of functional processes admitting such a decomposition). It is worth emphasizing that while both mixing and ergodicity describe forms of dependence among observations, they are fundamentally different. Specifically, the mixing property implies that any two subsets of the state space become asymptotically independent over time, whereas ergodicity implies that the system’s trajectory visits all regions of the space in proportion to their probability measure. Importantly, ergodicity is generally easier to verify than mixing. It is well known that ergodicity does not imply mixing, and there exist numerous ergodic time series that fail to satisfy any form of mixing assumption. Prominent examples include the following:

- First-order autoregressive processes with Bernoulli innovations (see [31]);
- Gaussian processes with Hurst exponent  $H > 0.5$  (see [32]);
- Gaussian processes with non-decaying covariance structures (see [33]).

Additional examples are discussed in [34], and these models can be naturally extended to the functional setting.

### 4.2. The Conditional Mode Versus the Conditional Mean

In predictive modeling, particularly when the conditional distribution is asymmetric or multimodal, the conditional mode often yields more accurate and meaningful predictions than the conditional mean. While the conditional mean represents the average outcome given certain inputs, it can be significantly affected by outliers or skewness in the distribution. In contrast, the conditional mode reflects the most probable outcome, making it more robust, reliable, and informative. A similar conclusion applies to the conditional median, which is also less sensitive to extreme values than the mean. As a result, combining the conditional mode and median can significantly outperform the conditional mean in predictive

tasks. This advantage becomes even more crucial in the context of ergodic functional time series, where providing robust predictors is essential. Ergodicity ensures that time averages converge to ensemble averages, offering a sound statistical basis for long-term forecasting. Moreover, the environmental data under study often exhibits seasonality, which can distort conditional mean predictions. In particular, repeating seasonal patterns can oversmooth the conditional expectation, reducing forecasting accuracy. The conditional mode, however, better captures the most likely outcomes within each seasonal segment, making it especially suitable for forecasting applications.

#### 4.3. The Recursive Estimation in Action

As with all smoothing approaches, the choice of the bandwidth parameter  $a_i$  is critical to the quality of the estimation. Typically, the mean squared error serves as a fundamental criterion for selecting this parameter. In the recursive framework considered here, we adopt the selection algorithm proposed by [20]. The smoothing parameter  $a_i$  is fixed to  $a_i = Ci^{-v}$ , where

$$C = \max d_i(\theta, I_i),$$

and  $v$  is selected by the cross-validation rule as follows

$$v_{opt} = \arg \min_{v \in (0,1)} \sum_{j=1}^n (O_j - \widehat{CM}^{-j}(I_j))^2, \quad (8)$$

where  $\widehat{CM}^{-j}$  is the leave-out-one estimator of the estimator. The rule (8) is similar to the cross-validation criterion considered by [10]. In our empirical analysis, the rule (8) is optimized over  $m$  equidistant points in the interval  $(0, 1)$ . Finally, it is worth noting that, although this selection approach has demonstrated good empirical performance, establishing its asymptotic optimality remains an important direction for future research.

#### 4.4. The Computational Cost

Recall that, unlike traditional kernel estimators, which compute the estimate independently at each point, the recursive version updates the estimate sequentially with each new observation, potentially reducing execution time. Consequently, computational efficiency is a significant advantage of the recursive estimator. Quantifying this efficiency is particularly important in the context of large datasets or real-time applications. Specifically, if each update involves a constant number of operations, i.e., of order  $O(1)$ , then the total computational cost becomes  $O(n)$ , which is considerably more efficient than the  $O(n^2)$  complexity of standard kernel estimators. However, in practical scenarios where the bandwidth is selected via adaptive tuning, the computational cost may increase. Despite its advantages, the recursive approach has a notable drawback: it requires storing past data, which negatively impacts memory usage. This limitation becomes especially critical when dealing with large sample sizes or high-dimensional data.

## 5. Simulation Study

In this simulation study, our objective is to investigate the feasibility and effectiveness of the proposed method. In particular, we seek to assess how dependence impacts the convergence rate comparing the algorithm's performance under different scenarios such as varying dependence levels, signal-to-noise ratios, or outlier contamination. To achieve this, we first generate an artificial dataset following a nonparametric form:

$$\text{heteroscedastic (Het.) Model : } O_i = 4 \int_0^1 \sin(3 + I_i^3(t)) dt + \cos(3 + I_i^3(t)) \epsilon_i,$$

$$\text{homoscedastic (Hom.) Model : } O_i = 5 \int_0^1 \log\left(\frac{2 + I_i^2(t)}{3 + I_i^3(t)}\right) dt + \epsilon_i,$$

and heteroscedastic model with signal-to-noise ratio (SNRHet.) Model:

$$O_i = 4 \int_0^1 \sin(3 + I_i^3(t)) dt + \sigma_{SNR} \epsilon_i,$$

where  $\epsilon_i$  and  $I_i$  are independent, and  $\sigma_{SNR}$  is controlled by considering various values of signal-to-noise ratio  $SNR_k = 5\%$  and  $40\%$ , where

$$SNR_k = \frac{\sigma_{SNR}^2}{\frac{1}{n} \sum_{i=1}^n (R_i - \bar{R})^2}, \quad R_i = 4 \int_0^1 \sin(3 + I_i^3(t)).$$

The functional input variable is generated from dependent functional processes using the R-package *FTSgof* (<https://www.r-project.org/>). We have generated  $n = 150$  observations. The resulting functional variables are presented in Figures 1–3.

Clearly, this sampling process encompasses three types of FTS-dependence, specifically functional autoregressive processes of order 2 (FAR(2)) involving two distinct kernels

$$\text{Gaussian kernel } \psi(t, s) = \exp\left\{\frac{t^2 + s^2}{2}\right\} \quad t, s \in [0, 1],$$

$$\text{Wiener kernel } \psi(t, s) = t(1 - t)s(1 - s) \quad t, s \in [0, 1].$$

The third illustration of a functional covariate setting is the functional ARCH(1) model, whose conditional volatility depends on the following kernel:

$$\text{Default kernel } \alpha(t, s) = 12 t(1 - t) s(1 - s), \quad t, s \in [0, 1]. \tag{9}$$

This kernel is the default choice in the *FTSgof* R-package. In our experimental design, the correlation level is adjusted via the function *fACF*. It is evident that this aspect emphasizes the effect of dependency levels on the accuracy of estimates. Meanwhile, the impact of outliers is managed by adjusting the observation responses  $(O_i)_i$  with a multiplicative factor *MF*, which can be either  $MF = 1$  or  $MF = 10$ . Meanwhile, the true values of the conditional mode, denoted by  $CM(\theta)$ , are obtained from the distribution of the underlying white noise  $\epsilon_i$ . This step is crucial because it allows us to evaluate how the nonparametric component influences the prediction task.

To investigate robustness, we consider three distinct distributions: Weibull, Laplace, and Log normal. These distributions are selected for their invariance under translation and varying degrees of heavy-tailed behavior, which in turn help to gauge the estimator’s resilience to outliers. We then compare the  $L^1$ -robustness of our estimator against multiple existing predictors.

To examine how recursion affects estimation, we contrast our proposed recursive estimator with a non-recursive version, in which the bandwidth parameter  $a_i$  remains

fixed at  $a_n$ . We further assess robustness by comparing  $\widehat{CM}$  against the double-kernel (DK) estimators given by

$$\text{The NW-estimator: } \widehat{CM}(\theta) = \arg \max_y \frac{\sum_{i=1}^n \Gamma(a_n^{-1}d(\theta, I_i)) \Gamma(b_n^{-1}(y - O_i))}{\sum_{i=1}^n b_n \Gamma(a_n^{-1}d(\theta, I_i))}, \quad (10)$$

and

$$\text{The DK-recursive estimator: } \overline{CM}(\theta) = \arg \max_y \frac{\sum_{i=1}^n \Gamma(a_i^{-1}d(\theta, I_i)) \Gamma(b_n^{-1}(y - O_i))}{\sum_{i=1}^n b_n \Gamma(a_i^{-1}d(\theta, I_i))}. \quad (11)$$

The performance of these three estimators,  $\widehat{CM}$ ,  $\overline{CM}$ , and  $\overline{CM}$  depends on the choice of parameters  $(a_n, b_n)$ . Selecting semi-metric  $d$  and kernel  $\Gamma$  also influences efficiency. In particular,  $\Gamma$  is chosen to satisfy (Co3), while  $d$  controls the smoothing level of the functional predictors  $I_i$ . To examine the feasibility of the selector algorithm discussed in Section 4 we simulate with  $a_i = C i^{-\nu}$ , where  $\nu$  is chosen from a 10 equidistant grid in  $(0, 1)$  and  $C = \max d_i(\theta, I_i)$ . Regarding the kernel  $\Gamma$ , we employ a quadratic kernel on  $(0, 1)$ , which is consistent with (Co3) and frequently used in nonparametric functional statistics. Moreover, the PCA metric proves especially suitable for cases in which the explanatory curves  $I_i$  exhibit discontinuities. In this empirical analysis, we proceed with the PCA associated with the third eigenfunction. Finally we point out that we took  $b_n = a_n = C n^{-\nu}$  for the estimators  $\widehat{CM}$ , and  $\overline{CM}$ ,  $\nu$  being selected by the same manner as in the estimator  $\widehat{CM}$ .

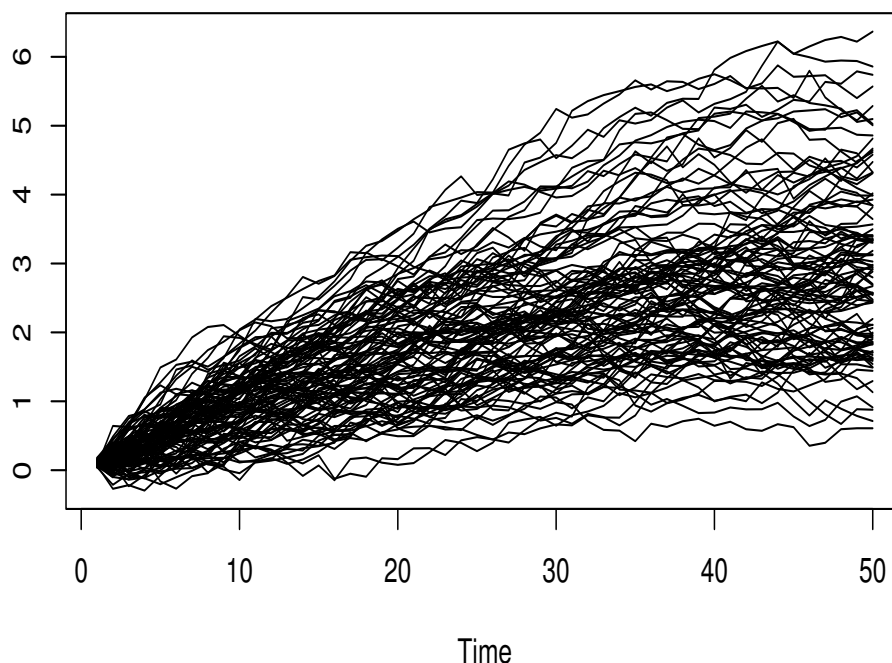


Figure 1. Functional autoregressive order 2: Wiener kernel.

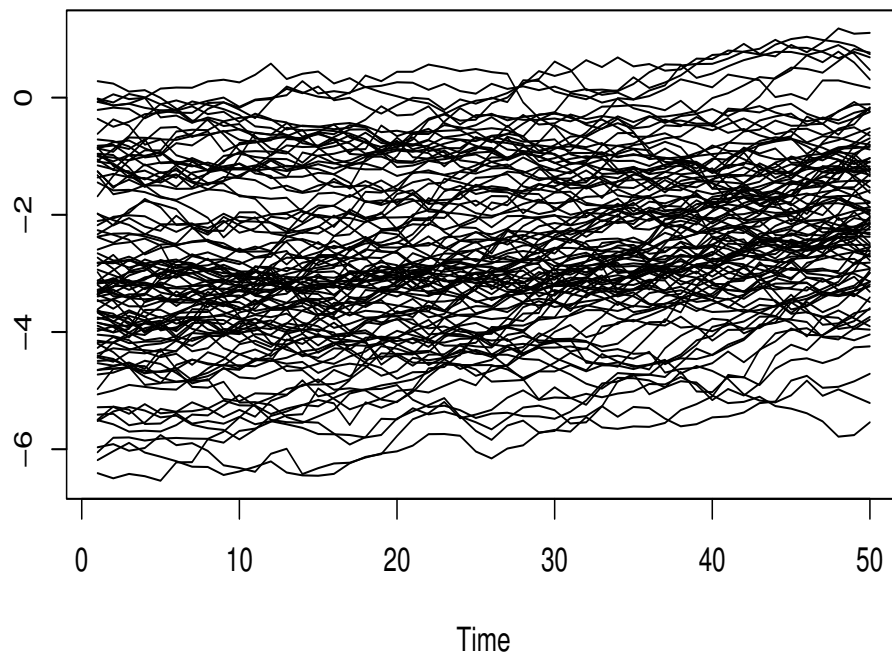


Figure 2. Functional autoregressive order 2: Gaussian kernel.

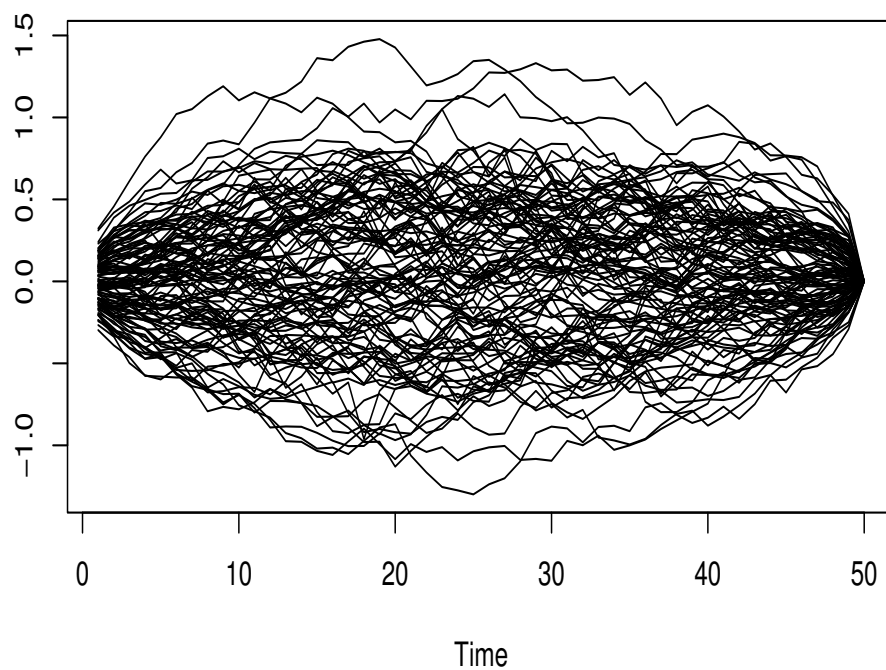


Figure 3. Functional GRCH order 1.

To compare the effectiveness of the estimators, we compute their mean square error (MSE) across all simulated scenarios,

$$\text{MSE}(\hat{C}\tilde{M}) = \frac{1}{n} \sum_{i=1}^n (O_i - \hat{C}\tilde{M}(I_i))^2, \tag{12}$$

where  $\hat{C}\tilde{M}$  can represent either  $\widehat{C}\tilde{M}$ ,  $\overline{C}\tilde{M}$ , or  $\widetilde{C}\tilde{M}$ . This metric enables us to contrast their performance under varying distributional assumptions dependence levels (GFAR(2) (Gaussian kernel based FAR(2)), WFAR(2) (Wiener kernel based FAR(2)), and FARCH(1)), signal-to-noise ratios, or outliers contamination. The results are reported in Tables 1–3.

**Table 1.** MSE errors of the estimator  $\widehat{CM}$ .

MF	FTS	Dist.	Het.	Hom.	SNRHet (5%)	SNRHet (40%)
MF = 1	GFAR(2)	Laplace	0.176	0.154	1.174	2.197
		Log normal	0.183	0.166	1.187	2.198
		Weibull	0.196	0.172	1.189	2.208
	WFAR(2)	Laplace	0.165	0.143	0.158	2.171
		Log normal	0.173	0.141	1.153	2.185
		Weibull	0.170	0.153	1.170	2.192
	FARCH(1)	Laplace	0.201	0.182	1.204	2.209
		Log normal	0.223	0.209	1.311	2.534
		Weibull	0.240	0.223	1.412	2.626
MF = 10	GFAR(2)	Laplace	0.256	0.261	1.353	2.413
		Log normal	0.317	0.312	1.487	2.507
		Weibull	0.296	0.542	1.618	2.698
	WFAR(2)	Laplace	0.356	0.334	1.385	0.497
		Log normal	0.432	0.513	1.635	2.758
		Weibull	0.408	0.443	1.489	2.595
	FARCH(1)	Laplace	0.513	0.523	1.641	2.691
		Log normal	0.434	0.419	1.453	2.554
		Weibull	0.504	0.514	1.562	1.516

**Table 2.** MSE errors of the estimator  $\widetilde{CM}$ .

MF	FTS	Dist.	HeT.	Hom.	SNRHet (5%)	SNRHet (40%)
MF = 1	GFAR(2)	Laplace	1.161	1.109	2.008	4.378
		Log normal	1.304	0.963	2.789	4.678
		Weibull	3.239	2.107	3.896	4.894
	WFAR(2)	Laplace	0.876	0.403	1.097	1.856
		Log normal	0.606	0.236	1.765	2.785
		Weibull	1.690	1.327	2.045	2.976
	FARCH(1)	Laplace	2.332	1.763	2.435	4.554
		Log normal	2.204	1.398	3.971	5.861
		Weibull	5.109	3.712	5.023	6.432
MF = 10	GFAR(2)	Laplace	4.201	4.216	7.312	8.417
		Log normal	4.230	4.736	6.789	7.678
		Weibull	5.117	6.107	7.186	8.243
	WFAR(2)	Laplace	3.654	3.212	4.178	5.164
		Log normal	3.902	4.561	5.605	6.194
		Weibull	4.310	4.127	6.205	6.817
	FARCH(1)	Laplace	6.231	7.862	8.333	9.352
		Log normal	4.315	5.493	6.771	8.662
		Weibull	7.101	7.513	8.224	9.533

The effectiveness of these estimators is substantially influenced by both the structure of the functional time series and the level of correlation. In addition, the predictor's accuracy depends on the choice of nonparametric modeling. Nevertheless, empirical findings indicate that the recursive approach generally surpasses the NW method in terms of precision and that the  $L^1$  method exhibits more stable mean squared error (MSE) variability compared to double-kernel techniques. Consequently, the estimator  $\widehat{CM}$  emerges as notably precise and robust since it combines the advantages of recursive algorithms

with those of  $L^1$ -based techniques. Finally, we observe that all considered functional estimators are straightforward to implement and maintain acceptable accuracy across a variety of scenarios.

**Table 3.** MSE errors of the estimators  $\overline{CM}$ .

MF	FTS	Dist.	Het.	Hom.	SNRHet (5%)	SNRHet (40%)
MF = 1	GFAR(2)	Laplace	1.535	1.331	2.103	4.414
		Log normal	1.202	1.106	2.452	4.786
		Weibull	2.119	1.811	3.02	4.949
	WFAR(2)	Laplace	0.167	0.156	1.861	3.843
		Log normal	0.134	0.136	1.451	3.073
		Weibull	0.109	0.117	0.698	1.785
	FARCH(1)	Laplace	3.101	2.512	3.972	4.952
		Log normal	2.603	2.363	3.861	5.045
		Weibull	4.009	3.227	4.961	5.895
MF = 10	GFAR(2)	Laplace	2.552	2.312	4.132	6.447
		Log normal	2.221	3.166	5.421	8.761
		Weibull	4.191	5.823	6.211	8.991
	WFAR(2)	Laplace	2.171	2.164	4.812	6.832
		Log normal	2.142	0.162	1.412	3.264
		Weibull	2.192	2.173	3.682	5.751
	FARCH(1)	Laplace	8.112	8.521	9.128	9.921
		Log normal	4.611	4.169	6.161	10.012
		Weibull	9.018	10.271	11.031	12.185

### 6. A Real Data Analysis

The purpose of this section is to evaluate how the  $L^1$ -recursive estimator of the conditional mode predictor performs in comparison with other recursive approaches. Specifically, we juxtapose it with recursive estimators of both the conditional mean and the conditional median. To conduct this forecasting problem, we rely on an environmental functional time series dataset. In particular, we focus on predicting air quality at a predetermined lead time by exploiting historical observations. The recursive predictors considered in this study are

$$(\text{cond.mean}) \quad \widehat{ME}(\theta) = \frac{\sum_{i=1}^n \Gamma(a_i^{-1}d(\theta, I_i)) O_i}{\sum_{i=1}^n \Gamma(a_i^{-1}d(\theta, I_i))},$$

and

$$(\text{cond.median}) \quad \widehat{CM}(\theta) = \arg \min_y \frac{\sum_{i=1}^n \Gamma(a_i^{-1}d(\theta, I_i)) |O_i - y|}{\sum_{j=1}^n \Gamma(a_j^{-1}d(\theta, I_j))}.$$

The dataset utilized for this comparative analysis is available online at <https://gaftp.epa.gov/castnet> (accessed on 8 January 2025). It contains information pertaining to the city of Bondville in Champaign, IL, USA. The monitoring station in question has the following geographical attributes, as shown in the following table.

Country	Sate	County	Code of Station	Geographical Coordinates
USA	Illinois	Champaign	BVL130	40.051981–88.372495

The dataset was recorded at hourly intervals from January through December 2024. The original set of observations is illustrated in Figure 4.

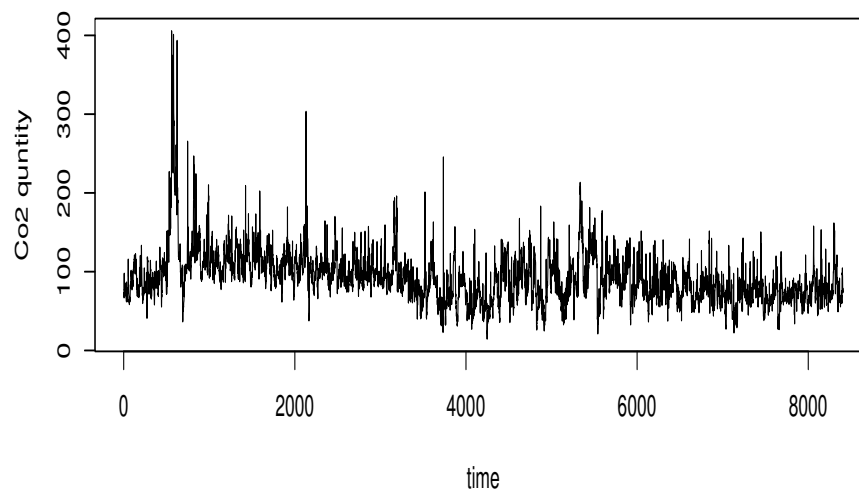


Figure 4. The CO<sub>2</sub> emission during 2024.

Recall that predicting ozone levels based on CO<sub>2</sub> concentrations is crucial for environmental sustainability. In Champaign, a city with a mix of urban traffic and surrounding agricultural activity, ozone levels can spike during hot, stagnant summer days, worsening air quality. Therefore, air quality in this region is significantly influenced by seasonal variations. Naturally, this effect can be managed by applying suitable seasonal data preprocessing techniques. Initially, we replaced missing values with the average of the nearest four values and employed correlation and causality analysis to pinpoint the right covariate variables. Following this preprocessing, we concluded that predicting CO<sub>2</sub> emissions three hours ahead using the past 24 h of historical data is more advantageous. To do that, we segment the 8736-h dataset into  $N + 1 = 364$  intervals, each denoted by  $I_i$ , and each interval  $I_i$  spans 24 h (i.e., one full day). Following this procedure, we define the output variable as

$$O_i = I_{i+1}(3). \tag{13}$$

The curve data are given in the following graph (Figure 5).

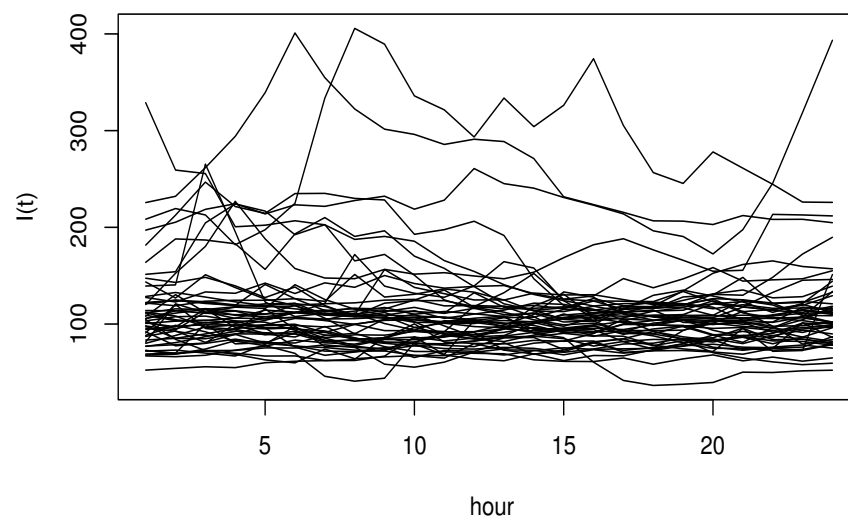


Figure 5. A sample of 50 curves of daily CO<sub>2</sub> emission.

To construct the estimators  $\widehat{ME}$ ,  $\widehat{CM}$ , and  $\widehat{MO}$ , we maintain the same smoothing approach, the same kernel function, and the same distance metric (PCA metric associated to the third eigenfunction). We subsequently evaluate and compare these estimators using the following procedure:

- *Step 1.* Randomly partition the dataset into two parts:
  - A training set,  $\{(I_j, O_j)\}_{j \in J}$ , consisting of 300 observations;
  - A test set,  $\{(I_i, O_i)\}_{i \in I}$ , consisting of 64 observations.
- *Step 2.* For each  $I_i$  in the training set, predict the corresponding response  $O_i$  by applying:
  - **Method 1 (Conditional mean):**

$$\widehat{O}_i^{ME} = \widehat{ME}(I_i);$$

- **Method 2 (Conditional mode):**

$$\widehat{O}_i^{MO} = \widehat{MO}(I_i);$$

- **Method 3 (Conditional median):**

$$\widehat{O}_i^{CM} = \widehat{CM}(I_i).$$

- *Step 3.* For each  $I_{\text{new}}$  in the test set, identify

$$i^* = \arg \min_{I_i \in \text{training set } J} d(I_{\text{new}}, I_i),$$

where  $d(\cdot, \cdot)$  denotes the chosen distance function.

- *Step 4.* Use the identified index  $i^*$  to predict  $O_{\text{new}}$ :

- **Method 1 (Conditional mean):**

$$\widehat{O}_{\text{new}}^{ME} = \widehat{ME}(I_{i^*});$$

- **Method 2 (Conditional mode):**

$$\widehat{O}_{\text{new}}^{MO} = \widehat{MO}(I_{i^*});$$

- **Method 3 (Conditional median):**

$$\widehat{O}_{\text{new}}^{CM} = \widehat{CM}(I_{i^*}).$$

- *Step 5.* To assess the prediction accuracy among the methods, compute the square root of the mean squared error (SMSE):

$$SMSE = \sqrt{\frac{1}{64} \sum_{i \in \text{Test set}} (O_i - \widehat{T}(I_i))^2},$$

where  $\widehat{T}$  can be  $\widehat{ME}$ ,  $\widehat{CM}$ , or  $\widehat{MO}$ .

- *Step 6.* Plot the actual response values versus the predicted values for each method.

Consistent with our expectations, the  $L^1$ -based recursive predictor  $\widehat{CM}$  demonstrates superior performance compared to the alternative models,  $\widehat{ME}$  and  $\widehat{MO}$ , see Figure 6. This

improvement in predictive accuracy is notably significant. To support this assertion, we computed the standardized mean squared error (SMSE). The SMSE for  $\widehat{CM}$  was 3.26, while  $\widehat{ME}$  and the second instance of  $\widehat{CM}$  yielded SMSE values of 5.42 and 4.87, respectively. These predictive error measures are broadly consistent with the findings reported in [35], although one must consider the considerable differences in the climatic conditions of the regions studied. Moreover, to evaluate the sensitivity of the proposed predictors to various parameter settings, we re-ran the algorithm using the  $L^2$  metric associated with the B-spline basis function, as well as the  $\beta$ -kernel with shape parameters (2,3). Subsequently, we computed the SMSE for each scenario.

Once again,  $\widehat{CM}$  appears to outperform the other models,  $\widehat{ME}$  and  $\widehat{CM}$ . Table 4 outlines the SMSE error for the different situations. It is evident that prediction is significantly influenced by the chosen parameters of the estimators. However, the choice of the metric has a greater impact compared to the choice of the kernel. It is clear that the PCA metric is more suitable for these data. This conclusion confirms the connection between the metric and the smoothing level of the curves. In fact, using the spline metric for discontinuous curves can over-smooth the functional covariate, leading to less accurate outcomes.

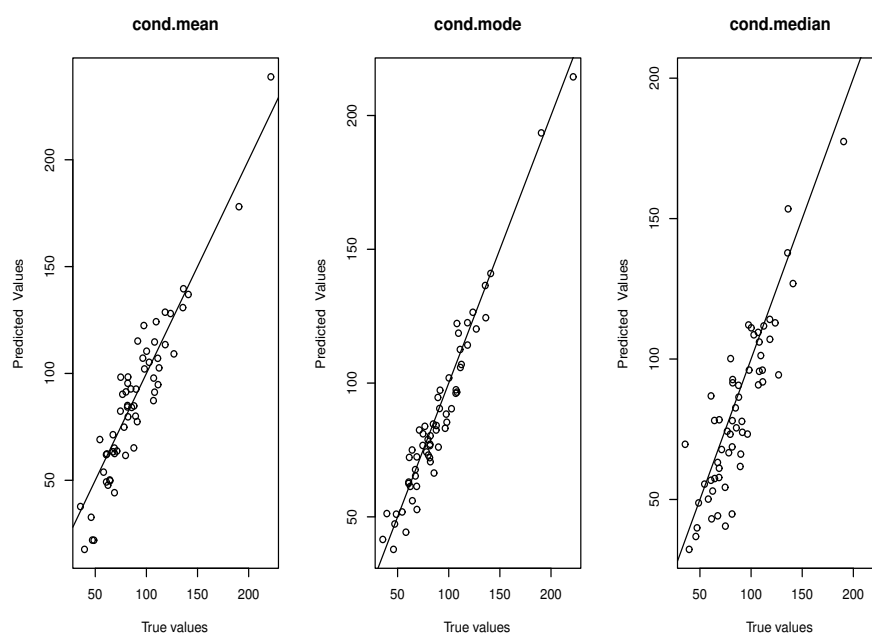


Figure 6. Comparison between the three predictors.

Table 4. SMSE errors of the three predictors with respect to different metrics and different kernels.

Model	Metric	Kernel	SMSE
$\widehat{CM}$	PCA (3th eigenfunction)	Quadratic kernel	3.26
	PCA (3th eigenfunction)	$\beta$ -kernel	3.37
	8th eigenfunction	Quadratic kernel	4.03
	8th eigenfunction	$\beta$ -kernel	4.11
	$L_2$ Spline metric	Quadratic kernel	4.39
	$L_2$ Spline metric	$\beta$ -kernel	4.52
$\widehat{ME}$	PCA (3th eigenfunction)	Quadratic kernel	5.42
	PCA (3th eigenfunction)	$\beta$ -kernel	5.61
	8th eigenfunction	Quadratic kernel	7.56
	8th eigenfunction	$\beta$ -kernel	8.22
	$L_2$ Spline metric	Quadratic kernel	6.45
	$L_2$ Spline metric	$\beta$ -kernel	6.82

**Table 4.** Cont.

Model	Metric	Kernel	SMSE
$\widehat{CM}$	PCA (3th eigenfunction)	Quadratic kernel	4.87
	PCA (3th eigenfunction)	$\beta$ -kernel	5.11
	8th eigenfunction	Quadratic kernel	8.62
	8th eigenfunction	$\beta$ -kernel	8.34
	$L_2$ Spline metric	Quadratic kernel	6.12
	$L_2$ Spline metric	$\beta$ -kernel	6.73

### 7. Conclusions

This work introduces a new predictor based on the estimation of the  $L^1$ -modal regression, constructed by means of a recursive procedure. The theoretical discussion provides the essential mathematical underpinnings that enable the straightforward, practical application of the proposed estimator. More specifically, we establish its asymptotic behavior under the fts-ergodic assumption, an alternative condition to the conventional correlation-based criteria.

Empirical evidence from artificial and real datasets confirms that the outlined implementation accommodates the theoretical assumptions. In particular, the accuracy of the estimator depends on the degree of correlation in the data, the smoothness of the underlying nonparametric model, and the careful selection of tuning parameters such as kernel and bandwidth. Notably, combining the  $L^1$  framework with a recursive approach offers improvements in terms of robustness and predictive precision.

In addition to these findings, the present study highlights several potential avenues for future investigation. One promising direction involves identifying the asymptotic distribution of the normalized estimator under various forms of fts, such as association or Markovian sequences. Another important extension concerns spatial modeling, which considers the geographical arrangement of the observations and supports more intricate prediction tasks. Although these extensions primarily focus on dependencies in the data, further generalizations to other smoothing methods, including the kNN approach, local linear estimators, and semi-partial linear techniques, remain equally compelling.

### 8. Proof of Propositions

**Proof of Lemma 2.** First, we define

$$\Gamma_i = \Gamma(a_i^{-1}d(\theta, I_i)) \quad \text{and} \quad \mathcal{L}_i = \Gamma_i - \mathbb{E}[\Gamma_i | \mathfrak{F}_{i-1}].$$

Thus,

$$\widehat{Q}_D(\theta) - \overline{Q}_D(\theta) = \frac{1}{n\bar{\zeta}_n} \sum_{i=1}^n \mathcal{L}_i.$$

As  $\mathcal{L}$  is a martingale difference for  $q \geq 2$ ,

$$\begin{aligned} \mathbb{E} \left[ \mathcal{L}_i^q | \mathfrak{F}_{i-1} \right] &\leq \text{CE} \left[ \mathcal{L}_i^2 | \mathfrak{F}_{i-1} \right] \\ &\leq \text{CE} \left[ \mathcal{L}_i^2 | \mathfrak{F}_{i-1} \right] \\ &< \text{CP} (I_i \in B(\theta, a_i) | \mathfrak{F}_{i-1}) \end{aligned}$$

$$\leq C\tilde{\zeta}_i(\theta, a_i).$$

Now, apply the exponential to obtain

$$\begin{aligned} \mathbb{P}\left\{\left|\hat{Q}_D(\theta) - \bar{Q}_D(\theta)\right| > \varepsilon\right\} &= \mathbb{P}\left\{\left|\frac{1}{n\tilde{\zeta}_n} \sum_{i=1}^n \mathcal{L}_i\right| > \varepsilon\right\} \\ &\leq 2 \exp\left\{-\frac{\varepsilon^2 n^2 \tilde{\zeta}_n^2(\theta)}{2(\zeta_n + C\varepsilon n\tilde{\zeta}_n)}\right\} \\ &= 2 \exp\left\{-\frac{\varepsilon^2 n^2 \tilde{\zeta}_n^2(\theta)}{2\zeta_n} \left(\frac{1}{1 + \frac{C\varepsilon n\tilde{\zeta}_n}{\zeta_n}}\right)\right\}. \end{aligned}$$

Putting  $\varepsilon = \varepsilon_0 \frac{\sqrt{\zeta_n \ln n}}{n\tilde{\zeta}_n}$ , then,

$$\mathbb{P}\left\{\left|\hat{Q}_D(\theta) - \bar{Q}_D(\theta)\right| > \varepsilon_0 \frac{\sqrt{\zeta_n \ln n}}{n\tilde{\zeta}_n}\right\} \leq 2 \exp\left\{-\frac{\varepsilon_0^2 \ln n}{2} \left(\frac{1}{1 + \frac{C\varepsilon_0 \sqrt{\ln n}}{\sqrt{\zeta_n}}}\right)\right\}.$$

Since

$$\frac{\ln n}{\zeta_n} \leq \frac{\ln n}{Cn\tilde{\zeta}_n} \leq C' \frac{\ln n}{n\tilde{\zeta}_n} \frac{\zeta_n}{n\tilde{\zeta}_n}.$$

Therefore,

$$\lim_{n \rightarrow \infty} \frac{\ln n}{\zeta_n} = 0.$$

This implies that

$$\mathbb{P}\left\{\left|\hat{Q}_D(\theta) - \bar{Q}_D(\theta)\right| > \varepsilon_0 \frac{\sqrt{\zeta_n \ln n}}{n\tilde{\zeta}_n}\right\} \leq 2 \exp\left\{-\frac{\varepsilon_0^2 \ln n}{2C}\right\} \leq 2n^{-C'\varepsilon_0^2}.$$

Hence,

$$\hat{Q}_D(\theta) - \bar{Q}_D(\theta) = O\left(\sqrt{\frac{\zeta_n \ln n}{n^2 \tilde{\zeta}_n^2}}\right), \quad a.co.$$

For the second result, we have the following for  $C > 0$

$$0 < C \frac{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i) | \mathfrak{F}_{i-1})}{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i))} \leq \bar{Q}_D(\theta) \leq \left|\bar{Q}_D(\theta) - \hat{Q}_D(\theta)\right| + \hat{Q}_D(\theta).$$

Thus,

$$C \frac{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i) | \mathfrak{F}_{i-1})}{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i))} - \left|\hat{Q}_D(\theta) - \bar{Q}_D(\theta)\right| < \hat{Q}_D(\theta).$$

So,

$$\begin{aligned} &\mathbb{P}\left(\hat{Q}_D(\theta) \leq \frac{C}{2}\right) \\ &\leq \mathbb{P}\left(C \frac{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i) | \mathfrak{F}_{i-1})}{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i))} - \left|\hat{Q}_D(\theta) - \bar{Q}_D(\theta)\right| < \frac{C}{2}\right) \end{aligned}$$

$$\leq \mathbb{P}\left(\left|C \frac{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i) | \mathfrak{F}_{i-1})}{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i))} - |\widehat{Q}_D(\theta) - \overline{Q}_D(\theta)| - C\right| > \frac{C}{2}\right).$$

It follows that

$$\begin{aligned} & \sum_n \mathbb{P}\left(\widehat{Q}_D(\theta) \leq \frac{C}{2}\right) \\ & \leq \sum_n \mathbb{P}\left(\left|C \frac{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i) | \mathfrak{F}_{i-1})}{\sum_{i=1}^n \mathbb{P}(I_i \in B(\theta, a_i))} - |\widehat{Q}_D(\theta) - \overline{Q}_D(\theta)| - C\right| > \frac{C}{2}\right) < \infty. \end{aligned}$$

Finally, from the first result, we obtain

$$\sum_n \mathbb{P}\left(\overline{Q}_D(\theta) \leq \frac{C}{2}\right) < \infty.$$

□

**Proof of Propositions 1.** Let

$$\gamma_n = \widehat{Q}u_p(\theta) - Qu_p(\theta).$$

Clearly,  $B_n(\gamma_n) = 0$ . Now, we check

$$A_n = o_{a.co.}(1), \tag{14}$$

and there exist  $M, \lambda > 0$  such that

$$\sup_{|\gamma| \leq M} |B_n(\gamma) + \lambda\gamma - A_n| = o_{a.co.}(1). \tag{15}$$

For (14), we evaluate

$$A_n - \overline{A}_n = O_{a.co.}\left(\sqrt{\frac{\zeta_n \ln n}{n^2 \zeta_n^2}}\right) \quad \text{and} \quad \overline{A}_n = O\left(\frac{1}{\zeta_n} \sum_{i=1}^n a_i^b \zeta_i(\theta, a_i)\right),$$

where

$$\overline{A}_n = \frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E}\left[\left(p - \mathbf{1}_{[O_i \leq Qu_p(\theta)]}\right) \Gamma_i | \mathfrak{F}_{i-1}\right].$$

Firstly, we have

$$\begin{aligned} \forall \varepsilon > 0 \quad \mathbb{P}\{|A_n - \overline{A}_n| > \varepsilon\} &= \mathbb{P}\left\{\left|\frac{1}{n\zeta_n} \sum_{i=1}^n \Psi_i\right| > \varepsilon\right\} \\ &\leq \mathbb{P}\left\{\left|\sum_{i=1}^n \Psi_i\right| > \varepsilon n\zeta_n\right\}, \end{aligned}$$

with

$$\Psi_i = \frac{1}{n\zeta_n} \left(p - \mathbf{1}_{[O_i \leq Qu_p(\theta)]}\right) \Gamma_i - \mathbb{E}\left[\left(p - \mathbf{1}_{[O_i \leq Qu_p(\theta)]}\right) \Gamma_i | \mathfrak{F}_{i-1}\right].$$

Clearly,  $\Psi_i$  is a martingale difference with respect to the  $\sigma$ -algebra, and  $(\mathfrak{F}_{i-1})_i$  satisfies the following for  $q \geq 2$

$$\mathbb{E}\left[\Psi_i^q | \mathfrak{F}_{i-1}\right] \leq C \mathbb{E}\left[\Psi_i^2 | \mathfrak{F}_{i-1}\right]$$

$$\begin{aligned} &\leq C\mathbb{E}\left[\Gamma_i^2|\mathfrak{F}_{i-1}\right] \\ &< C\mathbb{P}(I_i \in B(\theta, a_i)|\mathfrak{F}_{i-1}) \\ &\leq C\zeta_i(\theta, a_i). \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{P}\{|A_n - \overline{A}_n| > \varepsilon\} &= \mathbb{P}\left\{\left|\frac{1}{n\zeta_n} \sum_{i=1}^n \Psi_i\right| > \varepsilon\right\} \\ &\leq 2 \exp\left\{-\frac{\varepsilon^2 n^2 \zeta_n^2}{2(\zeta_n + C\varepsilon n\zeta_n)}\right\} \\ &= 2 \exp\left\{\frac{-\varepsilon^2 n^2 \zeta_n^2}{2\zeta_n} \left(\frac{1}{1 + \frac{C\varepsilon n\zeta_n}{\zeta_n}}\right)\right\}. \end{aligned}$$

So, for  $\varepsilon = \epsilon_0 \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n}$ , we have,

$$\mathbb{P}\left\{|A_n - \overline{A}_n| > \epsilon_0 \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n}\right\} \leq 2 \exp\left\{\frac{-\epsilon_0^2 \ln n}{2} \left(\frac{1}{1 + \frac{C\epsilon_0 \sqrt{\ln n}}{\sqrt{\zeta_n}}}\right)\right\}.$$

Since

$$\frac{\ln n}{\zeta_n} \leq \frac{\ln n}{Cn\zeta_n} \leq C' \frac{\ln n}{n\zeta_n} \frac{\zeta_n}{n\zeta_n}.$$

Then,

$$\lim_{n \rightarrow \infty} \frac{\ln n}{\zeta_n} = 0.$$

Therefore,

$$\mathbb{P}\left\{|A_n - \overline{A}_n| > \epsilon_0 \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n}\right\} \leq 2 \exp\left\{\frac{-\epsilon_0^2 \ln n}{2C}\right\} \leq 2n^{-C'\epsilon_0^2}.$$

For the second one, we have

$$\mathbf{1}_{B(\theta)}(I_1)|G(t, I_i) - G(t, \theta)| \leq Ca_i^b.$$

Then,

$$\begin{aligned} \overline{A}_n &= \frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E}\left[\left(p - \mathbf{1}_{[O_i \leq Qu_p(\theta)]}\right)\Gamma_i|\mathfrak{F}_{i-1}\right] \\ &= \frac{1}{n\zeta_n} \sum_{i=1}^n \left(\mathbb{E}\left[\Gamma_i\left(G(Qu_p(\theta)|\theta) - \mathbb{E}\left[\mathbf{1}_{[O_i \leq Qu_p(\theta)]}|I_i\right]\right)\right]|\mathfrak{F}_{i-1}\right) \\ &\leq \frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E}\left[\Gamma_i \mathbf{1}_{B(\theta, a_i)}(I_i)|G(Qu_p(\theta)|\theta) - G(Qu_p(\theta)|I_i)|\mathfrak{F}_{i-1}\right] \\ &\leq \frac{1}{n\zeta_n} \sum_{i=1}^n a_i^b \mathbb{E}[\Gamma_i|\mathfrak{F}_{i-1}]. \end{aligned} \tag{16}$$

Consequently,

$$|A_n - \overline{A}_n| = O_{a.co.} \left( \sqrt{\frac{\zeta_n \ln n}{n^2 \zeta_n^2}} \right)$$

and

$$\overline{A}_n = O \left( \frac{1}{\zeta_n} \sum_{i=1}^n a_i^b \zeta_i(\theta, a_i) \right).$$

For (15), similar to (14), we split the required result into parts

$$\sup_{|\gamma| \leq M} \left| B_n(\gamma) - A_n - \frac{1}{n \zeta_n} \sum_{i=1}^n \mathbb{E}[(B_n(\gamma) - A_n) | \mathfrak{F}_{i-1}] \right| = O_{a.co.} \left( \sqrt{\frac{\zeta_n \ln n}{n^2 \zeta_n^2}} \right), \tag{17}$$

and the bias term

$$\sup_{|\gamma| \leq M} \left| \frac{1}{n \zeta_n} \sum_{i=1}^n \mathbb{E}[B_n(\gamma) - A_n | \mathfrak{F}_{i-1}] + g(Qu_p(\theta))\gamma \right| = O \left( \frac{1}{\zeta_n} \sum_{i=1}^n a_i^b \zeta_i(\theta, a_i) \right). \tag{18}$$

Let us start with the dispersion term in (17). We employ the compactness of the interval  $[-M, M]$  and write

$$[-M, M] \subset \bigcup_{j=1}^{d_n} [\gamma_j - l_n, \gamma_j + l_n], \text{ for } \gamma_j \in [-M, M] \text{ and } l_n = d_n^{-1} = 1/\sqrt{n}.$$

So, for all  $\gamma \in [-M, M]$ , we put  $j(\gamma) = \arg \min_j |\gamma - \gamma_j|$  and use the monotony of  $B_n(\cdot)$  and  $\mathbb{E}[B_n(\gamma) | \mathfrak{F}_{i-1}]$ , which leads to the following for all  $1 \leq j \leq d_n$ ,

$$B_n(\gamma_j + l_n) \leq \sup_{y \in (\gamma_j - l_n, \gamma_j + l_n)} B_n(\gamma) \leq B_n(\gamma_j - l_n)$$

and

$$\mathbb{E}[B_n(\gamma_j + l_n) | \mathfrak{F}_{i-1}] \leq \sup_{\gamma \in (\gamma_j - l_n, \gamma_j + l_n)} \mathbb{E}[B_n(\gamma) | \mathfrak{F}_{i-1}] \leq \mathbb{E}[B_n(\gamma_j - l_n) | \mathfrak{F}_{i-1}].$$

We deduce, for any  $\gamma_1, \gamma_2 \in [-M, M]$ ,

$$\left| \frac{1}{n \zeta_n} \sum_{i=1}^n \mathbb{E}[B_n(\gamma_1) | \mathfrak{F}_{i-1}] - \frac{1}{n \zeta_n} \sum_{i=1}^n \mathbb{E}[B_n(\gamma_2) | \mathfrak{F}_{i-1}] \right| \leq C |\gamma_1 - \gamma_2|^b \overline{Q}_D(x).$$

It follows that

$$\begin{aligned} & \sup_{|\gamma| \leq M} \left| B_n(\gamma) - A_n - \frac{1}{n \zeta_n} \sum_{i=1}^n \mathbb{E}[(B_n(\gamma) - A_n) | \mathfrak{F}_{i-1}] \right| \\ & \leq \max_{1 \leq j \leq d_n} \max_{z \in \{\gamma_j - l_n, \gamma_j + l_n\}} \left| B_n(z) - A_n - \frac{1}{n \zeta_n} \sum_{i=1}^n \mathbb{E}[(B_n(z) - A_n) | \mathfrak{F}_{i-1}] \right| \\ & \quad + 2^b C l_n^b \overline{Q}_D(x). \end{aligned}$$

Concerning  $l_n^b$ , we write

$$\frac{l_n^b}{\sqrt{\frac{\zeta_n \log n}{n^2 \zeta_n^2}}} = \frac{l_n^b n \zeta_n}{\sqrt{\zeta_n \log n}}$$

$$\begin{aligned}
 &= \frac{n\tilde{\zeta}_n}{\sqrt{n\tilde{\zeta}_n \log n}} \\
 &= \sqrt{\frac{\sum_{i=1}^n \tilde{\zeta}(\theta, a_i)}{n}} \sqrt{\frac{\sum_{i=1}^n \tilde{\zeta}(\theta, a_i)}{\sum_{i=1}^n \tilde{\zeta}_i(\theta, a_i)}} \sqrt{\frac{1}{\log n}}.
 \end{aligned}$$

Furthermore, as  $\tilde{\zeta}(\theta, a_i) \leq 1$  we have, for all  $n$ ,

$$\frac{\sum_{i=1}^n \tilde{\zeta}(\theta, a_i)}{n} \leq 1,$$

and by **((C1)(iii))**

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \tilde{\zeta}(\theta, a_i)}{\sum_{i=1}^n \tilde{\zeta}_i(\theta, a_i)} \leq \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \tilde{\zeta}(\theta, a_i)}{\sum_{i=1}^n \mathbb{P}(I_i \in B(x, r_i) | \mathfrak{F}_{i-1})} = 1.$$

Finally, we can write

$$l_n^b = O\left(\sqrt{\frac{\zeta_n \log n}{n^2 \tilde{\zeta}_n^2}}\right).$$

Dealing with

$$\sup_{|\gamma| \leq M} \left| B_n(\gamma_j) - A_n - \frac{1}{n\tilde{\zeta}_n} \sum_{i=1}^n \mathbb{E}[B_n(\gamma_j) - A_n | \mathfrak{F}_{i-1}] \right|,$$

we set

$$B_n(\gamma_j) - A_n - \frac{1}{n\tilde{\zeta}_n} \sum_{i=1}^n \mathbb{E}[B_n(\gamma_j) - A_n | \mathfrak{F}_{i-1}] = \frac{1}{n\tilde{\zeta}_n} \sum_{i=1}^n Y_i,$$

with

$$\begin{aligned}
 Y_i &= \left( \mathbf{1}_{O_i \leq Qu_p(\theta)} - \mathbf{1}_{O_i \leq \gamma_j + Qu_p(\theta)} \right) \Gamma_i \\
 &\quad - \mathbb{E} \left[ \left( \mathbf{1}_{O_i \leq Qu_p(\theta)} - \mathbf{1}_{O_i \leq \gamma_j + Qu_p(\theta)} \right) \Gamma_i | \mathfrak{F}_{i-1} \right].
 \end{aligned}$$

As in  $A_n$ ,

$$\begin{aligned}
 \mathbb{E}[Y_i^q | \mathfrak{F}_{i-1}] &\leq C \mathbb{E}[Y_i^2 | \mathfrak{F}_{i-1}] \\
 &\leq C \mathbb{E}[\Gamma_i^2 | \mathfrak{F}_{i-1}] \\
 &< C \mathbb{P}(I_i \in B(\theta, a_i) | \mathfrak{F}_{i-1}) \\
 &\leq C \tilde{\zeta}_i(\theta, a_i).
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \mathbb{P} \left\{ \left| B_n(\gamma_j) - A_n - \frac{1}{n\tilde{\zeta}_n} \sum_{i=1}^n \mathbb{E}[B_n(\gamma_j) - A_n | \mathfrak{F}_{i-1}] \right| > \epsilon_0 \frac{\sqrt{\zeta_n \ln n}}{n\tilde{\zeta}_n} \right\} \\
 \leq 2 \exp \left\{ \frac{-\epsilon_0^2 \ln n}{2C} \right\} \leq 2n^{-C' \epsilon_0^2}.
 \end{aligned}$$

Consequently,

$$\begin{aligned} & \sum_n \mathbb{P} \left( \sup_{|\gamma| \leq M} \left| B_n(\gamma_{j(\gamma)}) - A_n - \frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E} [B_n(\gamma_{j(\gamma)}) - A_n] \right| \geq \epsilon_0 \left( \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n} \right) \right) \\ & \leq \sum_n d_n \max_j \mathbb{P} \left( \left| B_n(\gamma_j) - A_n - \frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E} [B_n(\gamma_j) - A_n] \right| \geq \epsilon_0 \left( \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n} \right) \right) < \infty, \end{aligned}$$

implying (17). Concerning (18), we write

$$\begin{aligned} & \frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E} [B_n(\gamma) - A_n | \mathfrak{F}_{i-1}] \\ & = -\frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E} \left[ \left( \mathbf{1}_{[O_1 \leq \gamma + Qu_p(\theta)]} - \mathbf{1}_{[O_1 \leq Qu_p(\theta)]} \right) \Gamma_i | \mathfrak{F}_{i-1} \right] \\ & = -\frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E} \left[ (G(\gamma + Qu_p(\theta) | I_1) - G(Qu_p(\theta) | I_1)) \Gamma_i | \mathfrak{F}_{i-1} \right] \\ & = -\frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E} \left[ (G(\gamma + Qu_p(\theta) | \theta) - g(Qu_p(\theta) | \theta)) \Gamma_i | \mathfrak{F}_{i-1} \right] + O(a_i^b) \\ & = -\frac{\gamma g(Qu_p(\theta) | \theta)}{\zeta_n} \frac{1}{n} \sum_{i=1}^n \mathbb{E} [\Gamma_i | \mathfrak{F}_{i-1}] + O \left( \frac{1}{\zeta_n} \sum_{i=1}^n a_i^b \zeta_i(\theta, a_i) \right) + o(\gamma). \end{aligned}$$

It follows that

$$\begin{aligned} & \mathbb{E} [B_n(\gamma) - A_n | \mathfrak{F}_{i-1}] \\ & = -g(Qu_p | \theta) \bar{Q}_D(x) \gamma + O \left( \frac{1}{\zeta_n} \sum_{i=1}^n a_i^b \zeta_i(\theta, a_i) \right) + o(\gamma). \end{aligned}$$

Therefore, the Bahadur representation of  $\widehat{Qu}_p(\theta) - Qu_p(\theta)$  is

$$\widehat{Qu}_p(\theta) - Qu_p(\theta) = \frac{1}{g(Qu_p | \theta)} A_n + O \left( \sup_{|\gamma| \leq M} |B_n(\gamma) + \lambda \gamma - A_n| \right).$$

□

**Proof of Propositions 2.** The uniform consistency of  $\widehat{Qu}_p(\theta) - Qu_p(\theta)$  is based on

$$\sup_{p \in [0,1]} |A_n - \mathbb{E}[A_n]| = O_{a.co.} \left( \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n} \right), \tag{19}$$

and

$$\sup_{p \in [0,1]} \sup_{|\gamma| \leq M} |B_n(\gamma) + g(Qu_p | \theta) \gamma - A_n| = O_{a.co.} \left( \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n} \right). \tag{20}$$

Since the inequalities in the bias terms are uniform on  $p \in [0, 1]$ . We only focus on the dispersion terms of

$$\sup_{p \in [0,1]} |A_n(p)| \quad \text{and} \quad \sup_{p \in [0,1]} \sup_{|\gamma| \leq M} |F_n(\gamma, p)|,$$

where

$$A_n(p) = \left| A_n - \frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E} [A_n | \mathfrak{F}_{i-1}] \right|,$$

and

$$F_n(\gamma, p) = \left| B_n(\gamma) - A_n - \frac{1}{n\zeta_n} \sum_{i=1}^n \mathbb{E}[(B_n(\gamma) - A_n) | \mathfrak{F}_{i-1}] \right|.$$

We focus on the first-term, the second one is similar. Indeed,

$$[0, 1] \subset \bigcup_{k=1}^{d_n} [p_k - l_n, p_k + l_n], \text{ for } p_k \in [0, 1].$$

Next, for all  $p \in [0, 1]$ , we put  $\eta_p = \arg \min_k |p - p_k|$ . Then,

$$\sup_{p \in [0,1]} |A_n(p)| \leq \max_{1 \leq j \leq d_n} \max_{z_p \in \{p_j - l_n, p_j + l_n\}} |A_n(z_p)| + 2^b C_l^b \bar{Q}_D(x).$$

It is shown in Lemma that for all  $p \in (0, 1)$  that

$$\mathbb{P} \left( |A_n(z_p)| \geq \epsilon_0 \left( \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n} \right) \right) \leq 2n^{-C'} \epsilon_0^2.$$

Therefore,

$$\begin{aligned} & \sum_n \mathbb{P} \left( \sup_{p \in [0,1]} |A_n(p)| \geq \epsilon_0 \left( \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n} \right) \right) \\ & \leq \sum_n d_n \max_j \mathbb{P} \left( |A_n(p)| \geq \epsilon_0 \left( \frac{\sqrt{\zeta_n \ln n}}{n\zeta_n} \right) \right) < \infty. \end{aligned}$$

The uniform consistency of  $(\sup_{p \in [0,1]} |\mathbb{E}[A_n]|)$  is obtained by taking the uniform version of (16), which allows us to conclude that

$$\sup_{p \in [0,1]} |\widehat{Q}u_p(\theta) - Qu_p(\theta)| = O \left( \frac{1}{\zeta_n} \sum_{i=1}^n a_i^b \zeta_i(\theta, a_i) \right) + O \left( \sqrt{\frac{\zeta_n \ln n}{n^2 \zeta_n^2}} \right).$$

□

**Author Contributions:** Conceptualization, M.B.A., F.A.A., I.M.A., S.B. and A.L.; Methodology, M.B.A., F.A.A., I.M.A., S.B. and A.L.; Software, M.B.A., F.A.A., I.M.A., S.B. and A.L.; Validation, F.A.A., I.M.A., S.B. and A.L.; Formal analysis, M.B.A., F.A.A., I.M.A., S.B. and A.L.; Investigation, M.B.A., I.M.A., S.B. and A.L. All authors have read and agreed to the final version of the manuscript.

**Funding:** This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R515), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia; and the Deanship of Scientific Research and Graduate Studies at King Khalid University through the Small Research Groups Program under grant number RGP1/41/46.

**Data Availability Statement:** The data used in this study are available through the link <https://kithub.cmu.edu> (accessed on 2 February 2025).

**Acknowledgments:** The authors thank and extend their appreciation to the funders of this work: This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R515), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, and the Deanship of Scientific Research and Graduate Studies at King Khalid University through the Small Research Groups Program under grant number RGP1/41/46. The authors extend their sincere gratitude to the Editor-in-Chief, the Associate Editor, and the reviewers for their invaluable feedback and for pointing out a number of oversights in the version initially submitted. Their insightful comments have greatly refined and focused the original work, resulting in markedly improved presentation.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Collomb, G.; Härdle, W.; Hassani, S. A note on prediction via estimation of the conditional mode function. *J. Stat. Plan. Inference* **1986**, *15*, 227–236. [[CrossRef](#)]
2. Quintela-Del-Rio, A.; Vieu, P. A nonparametric conditional mode estimate. *J. Nonparametr. Stat.* **1997**, *8*, 253–266. [[CrossRef](#)]
3. Ioannides, D.; Matzner-Løber, E. A note on asymptotic normality of convergent estimates of the conditional mode with errors-in-variables. *J. Nonparametr. Stat.* **2004**, *16*, 515–524. [[CrossRef](#)]
4. Louani, D.; Ould-Saïd, E. Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametr. Stat.* **1999**, *11*, 413–442. [[CrossRef](#)]
5. Allaoui, S.; Bouzebda, S.; Chesneau, C.; Liu, J. Uniform almost sure convergence and asymptotic distribution of the wavelet-based estimators of partial derivatives of multivariate density function under weak dependence. *J. Nonparametr. Stat.* **2021**, *33*, 170–196 [[CrossRef](#)]
6. Ferraty, F.; Laksaci, A.; Vieu, P. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.* **2006**, *9*, 47–76. [[CrossRef](#)]
7. Dabo-Niang, S.; Kaid, Z.; Laksaci, A. On spatial conditional mode estimation for a functional regressor. *Stat. Probab. Lett.* **2012**, *82*, 1413–1421. [[CrossRef](#)]
8. Ezzahrioui, M.H.; Ould-Saïd, E. Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.* **2008**, *20*, 3–18. [[CrossRef](#)]
9. Ezzahrioui, M.H.; Saïd, E.O. Some asymptotic results of a non-parametric conditional mode estimator for functional time-series data. *Stat. Neerl.* **2010**, *64*, 171–201. [[CrossRef](#)]
10. Ferraty, F.; Vieu, P. Functional Nonparametric Prediction Methodologies. In *Nonparametric Functional Data Analysis: Theory and Practice*; Springer: New York, NY, USA, 2006; pp. 49–59.
11. Dabo-Niang, S.; Kaid, Z.; Laksaci, A. Asymptotic properties of the kernel estimate of spatial conditional mode when the regressor is functional. *ASTA Adv. Stat. Anal.* **2015**, *99*, 131–160. [[CrossRef](#)]
12. Bouanani, O.; Laksaci, A.; Rachdi, M.; Rahmani, S. Asymptotic normality of some conditional nonparametric functional parameters in high-dimensional statistics. *Behaviormetrika* **2019**, *46*, 199–233. [[CrossRef](#)]
13. Ling, N.; Liu, Y.; Vieu, P. Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics* **2016**, *50*, 991–1013. [[CrossRef](#)]
14. Almanjahie, I.M.; Kaid, Z.; Laksaci, A.; Rachdi, M. Estimating the conditional density in scalar-on-function regression structure: K-NN local linear approach. *Mathematics* **2022**, *10*, 902. [[CrossRef](#)]
15. Attouch, M.; Bouabsa, W. The k-nearest neighbors estimation of the conditional mode for functional data. *Rev. Roum. Math. Pures Appl.* **2013**, *58*, 393–415.
16. Azzi, A.; Belguerna, A.; Laksaci, A.; Rachdi, M. The scalar-on-function modal regression for functional time series data. *J. Nonparametr. Stat.* **2024**, *36*, 503–526. [[CrossRef](#)]
17. Wang, T. Non-parametric Estimator for Conditional Mode with Parametric Features. *Oxf. Bull. Econ. Stat.* **2024**, *86*, 44–73. [[CrossRef](#)]
18. Schouten, B.; Klausch, T.; Buelens, B.; Van Den Brakel, J. A Cost–Benefit Analysis of Reinterview Designs for Estimating and Adjusting Mode Measurement Effects: A Case Study for the Dutch Health Survey and Labour Force Survey. *J. Surv. Stat. Methodol.* **2024**, *12*, 790–813. [[CrossRef](#)]
19. Guenani, S.; Bouabsa, W.; Omar, F.; Kadi Attouch, M.; Khardani, S. Some asymptotic results of a kNN conditional mode estimator for functional stationary ergodic data. *Commun. Stat.-Theory Methods* **2024**, *54*, 3094–3113. [[CrossRef](#)]
20. Thiam, A.; Thiam, B.; Crambes, C. Recursive estimation of nonparametric regression with functional covariate. *Qual. Control Appl. Stat.* **2014**, *59*, 527–528.
21. Slaoui, Y. Recursive nonparametric regression estimation for dependent strong mixing functional data. *Stat. Inference Stoch. Process.* **2020**, *23*, 665–697. [[CrossRef](#)]
22. Alamari, M.B.; Almulhim, F.A.; Litimein, O.; Mechab, B. Strong Consistency of Incomplete Functional Percentile Regression. *Axioms* **2024**, *13*, 444. [[CrossRef](#)]
23. Shang, H.L.; Yang, Y. Nonstationary functional time series forecasting. *J. Forecast.* **2024**, *early view*. [[CrossRef](#)]
24. Aneiros, G.; Horová, I.; Hušková, M.; Vieu, P. Special Issue on Functional Data Analysis and Related Fields. *J. Multivar. Anal.* **2022**, *189*, 104908. [[CrossRef](#)]
25. Moindjié, I.A.; Preda, C.; Dabo-Niang, S. Fusion regression methods with repeated functional data. *Comput. Stat. Data Anal.* **2025**, *203*, 108069. [[CrossRef](#)]
26. Gertheiss, J.; Rügamer, D.; Liew, B.X.; Greven, S. Functional data analysis: An introduction and recent developments. *Biom. J.* **2024**, *66*, e202300363. [[CrossRef](#)]

27. Agua, B.M.; Bouzebda, S. Single index regression for locally stationary functional time series. *AIMS Math.* **2024**, *9*, 36202–36258. [[CrossRef](#)]
28. Bouanani, O.; Bouzebda, S. Limit theorems for local polynomial estimation of regression for functional dependent data. *AIMS Math.* **2024**, *9*, 23651–23691. [[CrossRef](#)]
29. Bouzebda, S. Weak convergence of the conditional single index  $U$ -statistics for locally stationary functional time series. *AIMS Math.* **2024**, *9*, 14807–14898. [[CrossRef](#)]
30. Xu, R.; Wang, J.  $L_1$ -estimation for spatial nonparametric regression. *J. Nonparametr. Stat.* **2008**, *20*, 523–537. [[CrossRef](#)]
31. Andrews, D. *First Order Autoregressive Processes and Strong Mixing*; Cowles Foundation Discussion Papers 664. Cowles Foundation for Research in Economics; Yale University: New Haven, CT, USA, 1983.
32. Beran, J. *Statistics for Long-Memory Processes*; Routledge: London, UK, 2017.
33. Ibragimov, I.A.; Rozanov, Y.A.E. *Gaussian Random Processes*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 9.
34. Magdziarz, M.; Weron, A. Ergodic properties of anomalous diffusion processes. *Ann. Phys.* **2011**, *326*, 2431–2443. [[CrossRef](#)]
35. Aneiros-Pérez, G.; Cardot, H.; Estévez-Pérez, G.; Vieu, P. Maximum ozone concentration forecasting by functional non-parametric approaches. *Environmetrics* **2004**, *15*, 675–685. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.