

Article

K-Nearest Neighbor Estimation of Functional Nonparametric Regression Model under NA Samples

Xueping Hu ¹, Jingya Wang ¹, Liuliu Wang ¹ and Keming Yu ^{1,2,*}

¹ College of Mathematics and Physics, Anqing Normal University, Anqing 246133, China; hxprob@163.com (X.H.); alyssa.wang@lcfuturecenter.com (J.W.); wliu970529@163.com (L.W.)

² Department of Mathematics, Brunel University, London UB8 3PH, UK

* Correspondence: keming.yu@brunel.ac.uk

Abstract: Functional data, which provides information about curves, surfaces or anything else varying over a continuum, has become a commonly encountered type of data. The k-nearest neighbor (kNN) method, as a nonparametric method, has become one of the most popular supervised machine learning algorithms used to solve both classification and regression problems. This paper is devoted to the k-nearest neighbor (kNN) estimators of the nonparametric functional regression model when the observed variables take values from negatively associated (NA) sequences. The consistent and complete convergence rate for the proposed kNN estimator is first provided. Then, numerical assessments, including simulation study and real data analysis, are conducted to evaluate the performance of the proposed method and compare it with the standard nonparametric kernel approach.

Keywords: convergence rate; NA samples; functional data; nonparametric regression model; k-nearest neighbor estimator

MSC: 62G08; 62G20



Citation: Hu, X.; Wang, J.; Wang L.; Yu, K. K-Nearest Neighbor Estimation of Functional Nonparametric Regression Model under NA Samples. *Axioms* **2022**, *11*, 102. <https://doi.org/10.3390/axioms11030102>

Academic Editor: Christophe Chesneau

Received: 10 January 2022

Accepted: 21 February 2022

Published: 25 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Functional data analysis (FDA) is a branch of statistics that analyzes data providing information about curves, surfaces or anything else varying over a continuum. In its most general form, under an FDA framework, each sample element of functional data is considered to be a random function.

Popularized by Ramsay and Silverman [1,2], statistics for functional data analysis have attracted considerable research interest because of its wide applications in many practical fields, such as medicine, economics and linguistics. For an introduction to the topics, we can refer to the monographs of Ramsay and Silverman [3] for parametric models, and Ferraty and Vieu [4] for nonparametric models.

In this paper, the following functional non-parametric regression model is considered.

$$Y = m(\chi) + \epsilon, \quad (1)$$

where Y is a scalar response variable, χ is a covariate taking value in a subset $S_{\mathfrak{F}}$ of an infinite-dimensional functional space \mathfrak{F} endowed with a semi-metric $d(\cdot, \cdot)$. $m(\cdot)$ is the unknown regression operator from $S_{\mathfrak{F}}$ to \mathfrak{R} , and the random error ϵ satisfies $E(\epsilon|\chi) = 0, a.s.$

For the estimation of model (1), Ferraty and Vieu [5] investigated the classical functional Nadaraya-Watson (N-W) kernel type estimator of $m(\cdot)$ and obtained the asymptotic properties with rates in the case of α -mixing functional data. Ling and Wu [6] studied the modified N-W kernel estimate and derived the asymptotic distribution for strong mixing functional time series data, Baíllo and Grane [7] proposed a functional local linear estimate based on the local linear idea. In this paper, we focus on the k-nearest neighbors (kNN)

method for regression model (1). The kNN method, as one of the most simple and traditional nonparametric techniques, is often used as a nonparametric classification method. The kNN method was first developed by Evelyn Fix and Joseph Hodges in 1951 [8] and then expanded by Thomas Cover [9]. In our kNN regression, the input consists of the k-closest training examples in a dataset, whereas the output is the property value for the object. This value is the average of the values of the k-nearest neighbors. Under independent samples, research in kNN regression mostly focuses on the estimation of the continuous regression function $m(\chi)$. For example, Burba et al. [10] investigated the kNN estimator based on the idea of the local adaptive bandwidth of functional explanatory variables. The papers [11–18], and others, obtained the asymptotic behavior of nonparametric regression estimators for functional data in independent and dependent cases. Further, Kudraszow and Vieu [19] obtained asymptotic results for a kNN generalized regression estimator when the observed variables take values in an abstract space. Kara-Zaitri et al. [20] provided an asymptotic theory for several different target operators and some simulated experiences, including regression, conditional density, conditional distribution and hazard operators. However, functional observations often behave with correlation, including satisfying some form of negative dependence or negative association.

Negatively associated (NA) sequences were introduced by Joag-Dev and Proschan in [21]. Random variables $\{Y_i\}_{1 \leq i \leq n}$ are said to be NA, if for every pair of disjoint subsets $A, B \subset \{1, 2, \dots, n\}$,

$$\text{Cov}(f(Y_i, i \in A)g(Y_j, j \in B)) \leq 0,$$

or equivalently,

$$E(f(Y_i, i \in A), g(Y_j, j \in B)) \leq E(f(Y_i, i \in A))E(g(Y_j, j \in B)),$$

where f and g are coordinatewise non-decreasing, such that this covariance exists. An infinite sequence $\{Y_n\}_{n \geq 1}$ is NA if every finite subcollection is NA.

For example, if $\{Y_i\}_{1 \leq i \leq n}$ follows permutation distributions, where $\{Y_1, Y_2, \dots, Y_n\} = \{y_1, y_2, \dots, y_n\}$ always and $y_1 \leq y_2 \leq \dots \leq y_n$ are n real numbers, then $\{Y_i\}_{1 \leq i \leq n}$ is NA.

Whereas kNN regression under NA sequences has not been explored in the literature, in this paper, we extend the kNN estimation of functional data from the case of independent samples to NA sequences.

Let a pair $\{(\chi_i, Y_i)\}_{i=1, \dots, n}$ be a sample of NA pairs in (χ, Y) , which is a random vector valued in the $\mathfrak{F} \times \mathfrak{R}$. (\mathfrak{F}, d) is a semi-metric space, \mathfrak{F} is not necessarily of the finite dimension and we do not suppose the existence of a density for the functional random variable χ . For a fixed $\chi \in \mathfrak{F}$, the closed ball with χ as the center and ϵ as the radius is denoted as:

$$d : B(\chi, \epsilon) = \{\chi' \in \mathfrak{F} | d(\chi', \chi) \leq \epsilon\}.$$

The kNN regression estimator [10] is defined as follows:

$$\hat{m}_{kNN}(\chi) = \frac{\sum_{i=1}^n Y_i K(H_{n,k}(\chi)^{-1} d(\chi_i, \chi))}{\sum_{i=1}^n K(H_{n,k}(\chi)^{-1} d(\chi_i, \chi))}, \chi \in \mathfrak{F}, \tag{2}$$

where $K(\cdot)$ is the kernel function supported on $[0, \infty)$. $H_{n,k}(\chi)$ is a positive random variable that depends on $(\chi_1, \chi_2, \dots, \chi_n)$ and is defined by:

$$H_{n,k}(\chi) = \min\{h \in R^+ : \sum_{i=1}^n I_{B(\chi, h)}(\chi_i) = k\},$$

obviously, the kNN estimator can be seen as an expansion to a random locally adaptive neighborhood of the traditional kernel method [5] defined as:

$$\hat{m}_n(\chi) = \frac{\sum_{i=1}^n Y_i K(h_n(\chi)^{-1}d(\chi_i, \chi))}{\sum_{i=1}^n K(h_n(\chi)^{-1}d(\chi_i, \chi))}, \chi \in \mathfrak{F}, \tag{3}$$

where $h_n(\chi)$ is a sequence of positive real numbers such as $h_n(\chi) \rightarrow 0$ a.s. $n \rightarrow \infty$.

This paper is organized as follows. The main results of our paper about the asymptotic behavior of the kNN estimators using a data-driven random number of neighbors are given in Section 2. Section 3 illustrates the numerical performance of the proposed method, including nonparametric functional regression analysis of the sea level surface temperature (SST) data for the El Niño area (0–100 S, 800–900 W). The technical proofs are postponed to Section 4. Finally, Section 5 is devoted to comments on the results and to related perspectives for the future.

2. Assumptions and Main Results

In this section, we focus on the asymptotic property of the kNN regression estimator and need to state the convergence rate of an estimator.

One says that the rate of almost complete convergence of a sequence $\{Y_n, n \geq 1\}$ to Y is of order u_n if only if for any $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P(|Y_n - Y| > \epsilon u_n) < \infty,$$

and we write $Y_n - Y = O_{a.co.}(u_n)$ (see for instance [5]). By the Borel-Cantelli lemma, this implies that $\frac{Y_n - Y}{u_n} \rightarrow 0$ almost surely, so almost complete convergence is a stronger result than almost sure convergence.

Our results are stated under some mild assumptions we gather below for easy references. Throughout the paper, we will denote by C, C_1, C' some positive generic constants, which may be different in various places.

Assumption 1. $\forall \epsilon > 0, P(\chi \in B(\chi, \epsilon)) = \varphi_\chi(\epsilon) > 0$ and $\varphi_\chi(\cdot)$ is a continuous function, and strictly monotonically increasing at the origin with $\varphi_\chi(0) = 0$.

Assumption 2. There exist a function $\phi(\cdot) \geq 0$ and a bounded function $f(\cdot) > 0$ such that:

- (i) $F\phi(0) = 0$, and $\lim_{\epsilon \rightarrow \infty} \phi(\epsilon) = 0$.
- (ii) $\lim_{\epsilon \rightarrow \infty} \frac{\phi(u\epsilon)}{\phi(\epsilon)} = 0$, for any $u \in [0, 1]$.
- (iii) $\exists \tau > 0$ such that $\sup_{\chi \in S_{\mathfrak{F}}} \left| \frac{\varphi_\chi(\epsilon)}{\phi(\epsilon)} - f(\chi) \right| = O(\epsilon^\tau), \epsilon \rightarrow 0$.

Assumption 3. $K(t)$ is a nonnegative bounded kernel function with support $[0, 1]$, and if $K(1) > 0$, the derivative $K'(t)$ exists on $[0, 1]$ satisfying:

$$-\infty < C < K'(t) < C' < \infty, \text{ for } \forall t \in [0, 1].$$

Assumption 4. $m(\cdot)$ is a bounded Lipschitz operator with order β on $S_{\mathfrak{F}}$, and there exists $\beta > 0$ such that:

$$\forall \chi_1, \chi_2 \in S_{\mathfrak{F}}, |m(\chi_1) - m(\chi_2)| \leq Cd(\chi_1, \chi_2)^\beta.$$

Assumption 5. $\forall m \geq 2, E(|Y|^m | X = \chi) = \delta_m(\chi) < C$ with $\delta_m(\cdot)$ continuous on $S_{\mathfrak{F}}$.

Assumption 6. Kolmogorov’s ϵ -entropy of $S_{\mathfrak{F}}$ satisfies:

$$\sum_{n=1}^{\infty} \exp \left\{ (1 - \omega) \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right) \right\} < \infty, \text{ for some } \omega > 1.$$

For $\forall \epsilon > 0$, the Kolmogorov’s ϵ -entropy of some set $S_{\mathfrak{F}} \subset \mathfrak{F}$ is defined by $\Psi_{S_{\mathfrak{F}}} = \log(N_{\epsilon}(S_{\mathfrak{F}}))$, where $N_{\epsilon}(S_{\mathfrak{F}})$ is the minimal number of open balls, which can cover $S_{\mathfrak{F}}$ with $\chi_1, \chi_2, \dots, \chi_{N_{\epsilon}(S_{\mathfrak{F}})}$ as the center and ϵ as the radius in \mathfrak{F} .

Remark 1. Assumption 1, Assumption 2((i)–(iii)) and Assumption 4 are the standard assumptions for small ball probability and regression operators in nonparametric FDA, see Kudraszow and Vieu [19]. Assumption 2(ii) will play a key role in the methodology particularly when we compute the asymptotic variance and permit it to be explicit in Ling and Wang [6]. Assumption 2(iii) shows that the small ball probability can be written as the product of the two independent functions $\phi(\cdot)$ and $f(\cdot)$, which has been used many times in Masry [11], Laib and Louani [12] and other literatures. Assumption 5 is standard in the nonparametric setting and concerns the existence of the conditional moments in Masry [11] and Burba [10], which aims to obtain the rate of uniform almost complete convergence. Assumption 6 assumes the Kolmogorov’s ϵ -entropy condition, which we will use in the following proof of the rate of uniform almost complete convergence.

Theorem 1. Under Assumptions 1–6, suppose that sequence $\{k_n, n \geq 1\}$ satisfies $\frac{k_n}{n} \rightarrow 0, n \rightarrow \infty, \frac{\log^2 n}{k_n} < \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right) < \frac{k_n}{\log n}$ and $0 < C_1 < \frac{k_n}{\log^2 n} < C_2 < \infty$, for n large enough, then we have:

$$\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{kNN}(\chi) - m(\chi)| = O_{a.co.} \left(\phi^{-1} \left(\frac{k_n}{n} \right)^{\beta} + \sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right)}{n^2}} \right).$$

Remark 2. The Theorem extends the kNN estimation result of Theorem 2 in Kudraszow and Vieu [19] from the independent case to the NA mixed dependent case, and obtains the same convergence rate under the assumptions. Second, the almost complete convergence rate of the prediction operator is divided into two parts, one part affected by strong mixing and Kolmogorov’s ϵ -entropy, and the other part depends on the smoothness of the regression operator and smoothness parameter k .

Corollary 1. Under the condition of the Theorem, we have:

$$\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{kNN}(\chi) - m(\chi)| = O_{a.s.} \left(\phi^{-1} \left(\frac{k_n}{n} \right)^{\beta} + \sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right)}{n^2}} \right).$$

Corollary 2. Under the condition of the Theorem, we have:

$$\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{kNN}(\chi) - m(\chi)| = O_P \left(\phi^{-1} \left(\frac{k_n}{n} \right)^{\beta} + \sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right)}{n^2}} \right).$$

3. Simulation

3.1. A simulation Study

In this section, we aim at illustrating the performance of the nonparametric functional regression model and we will make a comparison with traditional kernel density estimation methods. We consider the nonparametric functional regression model:

$$Y_i = m(\chi_i) + \epsilon_i,$$

where $m(\chi_i) = \left(\int_0^{\frac{\pi}{5}} \chi_i'(t) dt \right)^2$, ε_i is distributed according to $N(0, 0.05)$, the functional curve $\chi_i(t)$ is generated in the following way:

$$\chi_i(t) = a_i t^3 + \arctan\left(b_i \left(t - \frac{\pi}{5}\right)\right), t \in \left[0, \frac{\pi}{5}\right], i = 1, 2, \dots, n.$$

where $\{a_i\} \sim N(0, \frac{\pi}{10}), i = 1, 2, \dots, n, \{b_1, b_2, \dots, b_n\} \sim N_n(\mathbf{0}, \Sigma), \mathbf{0}$ represents zero vector and the covariance matrix is defined as:

$$\Sigma = \begin{pmatrix} 1 + \theta^2 & -\theta & 0 & \dots & 0 & 0 & 0 \\ -\theta & 1 + \theta^2 & -\theta & \dots & 0 & 0 & 0 \\ 0 & -\theta & 1 + \theta^2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 + \theta^2 & -\theta & 0 \\ 0 & 0 & 0 & \dots & -\theta & 1 + \theta^2 & -\theta \\ 0 & 0 & 0 & \dots & 0 & -\theta & 1 + \theta^2 \end{pmatrix}_{n \times n}, 0 < \theta < 1.$$

By the definition of NA, it can be seen that (b_1, b_2, \dots, b_n) is an NA vector for each $n \geq 3$ with a finite moment of any order (see Wu and Wang [22]).

We choose casually that $\theta = 0.4$, the sample sizes n as $n = 330$, t takes 1000 equispaced values in $[0, \frac{\pi}{5}]$. We carry out the simulation of the curve $\chi(t)$ for the 330 samples (see Figure 1).

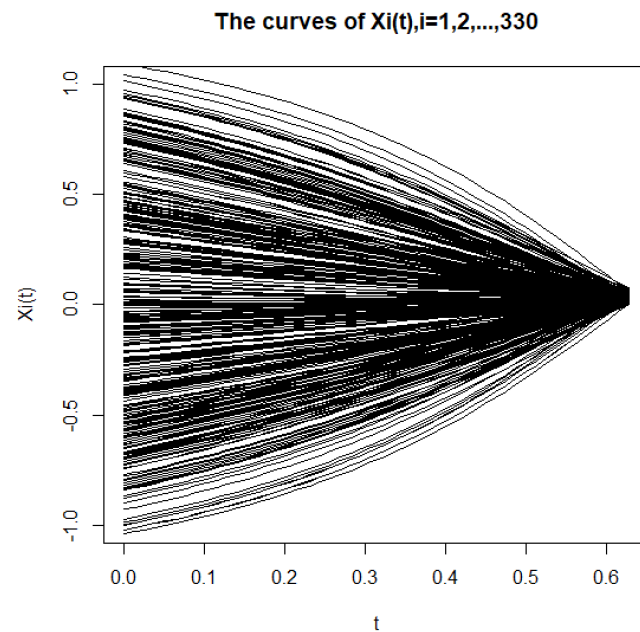


Figure 1. Curve-sample with sample size of $n = 330$.

We consider the Epanechnikov kernel given by $K(u) = \frac{3}{4}(1 - u^2)I_{[0,1]}(u)$, and the semi-metrics $d(\cdot, \cdot)$ based on derivatives of order q .

$$d(\chi_i, \chi_j) = \sqrt{\int_0^{\frac{\pi}{5}} (\chi_i^{(q)}(t) - \chi_j^{(q)}(t))^2 dt}, \forall \chi_i, \chi_j \in \mathfrak{F}, q = \{0, 1, 2, \dots\}.$$

Our purpose is to compare the mean square error (MSE) of the kNN method with the NW kernel approach on finite simulated datasets. In the finite sample simulation, the following steps are followed.

Step 1: We take 300 curves to construct the training samples $\{\chi_i, Y_i\}_{i=1}^{300}$, and the other 30 constitute the test samples $\{\chi_i, Y_i\}_{i=301}^{330}$.

Step 2: In the training sample, the parameters k and h in the kNN method and NW kernel method are automatically selected based on the cross-validation method, respectively.

Step 3: Based on the MSE standard (see [4] for details), we obtain that the respective semi-metric parameters q in both the kNN method and the NW method takes $q = 1$.

Step 4: The response values $\{\hat{Y}_i\}_{i=301}^{330}$ and $\{\tilde{Y}_i\}_{i=301}^{330}$ of the test sample $\{Y_i\}_{i=301}^{330}$ are calculated by using the kNN method and the NW method, respectively, and their MSE and scatter plots against the true value $\{Y_i\}_{i=301}^{330}$ are represented by Figure 2.

As we can see in Figure 2, the MSE of the kNN method is much smaller than that of the NW method, and the scattered points in Figure 2 are more densely distributed around the linear function $y = x$, which shows that the kNN method has a better fit and higher prediction accuracy for the NA dependent functional samples.

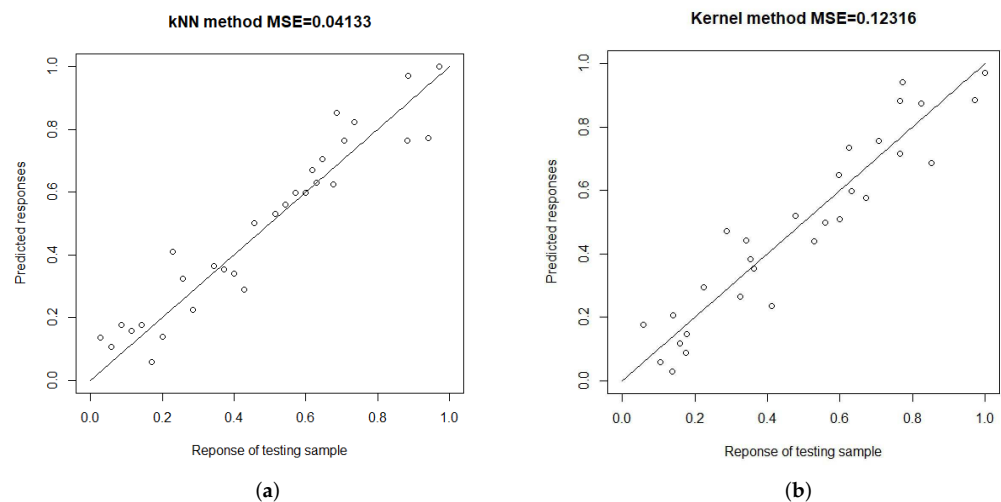


Figure 2. Prediction effects of the two estimation methods. (a) kNN estimation method. (b) NW estimation method.

The kNN method and NW method were used to conduct 100 independent replicated experiments at sample sizes of $n = 200, 300, 500, 800$, respectively. AMSE was calculated for both methods at different sample sizes using the following equation.

$$AMSE = \frac{1}{100} \sum_{j=1}^{100} \frac{1}{30} \sum_{i=n-30}^n (\bar{Y}_i - Y_i)^2, \bar{Y}_i = \hat{Y}_i, \tilde{Y}_i, n = 200, 300, 500, 800$$

As can be seen from Table 1, the AMSE of the kNN method is much smaller than that of the NW kernel method when the sample size is fixed at $n = 200, 300, 500, 800$, respectively; when the estimation method is fixed, the AMSE of the two estimation methods have the same trend—they both decrease as the sample size increases. However, the decreasing speed of the kNN method is significantly faster than that of the NW kernel method.

Table 1. The AMSE of the predicted response variables of the two methods under different sample sizes.

n	200	300	500	800
kNN- AMSE	0.0623	0.0470	0.0327	0.0291
NW- AMSE	0.2764	0.2593	0.2329	0.2117

3.2. A Real Study

This section applies the proposed kNN regression analysis of the data, which consist of the sea level surface temperature (SST) for the El Niño area (0–100 S, 800–900 W) for a total of 31 years from 1 January 1990 to 31 December 2020. The data are available online at

the website: <https://www.cpc.ncep.noaa.gov/data/indices/> (accessed on 1 January 2022). More relevant discussions of these data can be found in Ezzahrioui et al. [13,14], Delsol et al. [23], and Ferraty et al. [24]. The 1618 weekly SST data from the original data were preprocessed and averaged by month to obtain 372 monthly average SST discrete data. Figure 3 displays the decomposition of the multiplicative time series of the monthly SST.

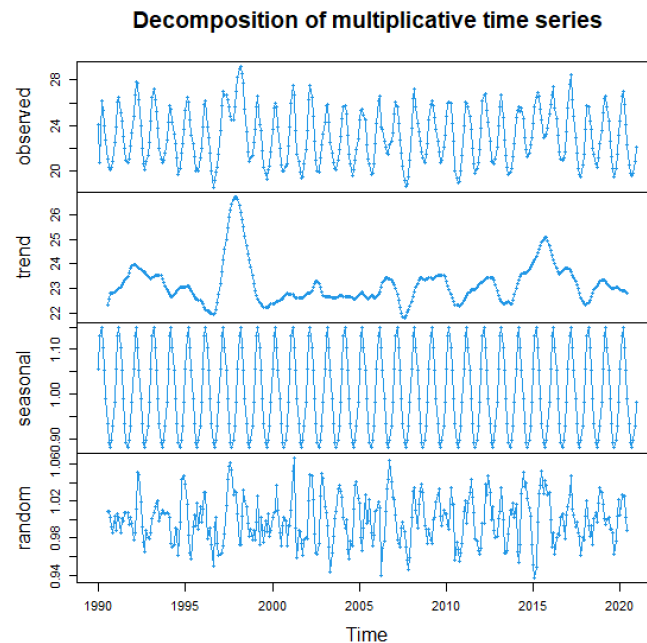


Figure 3. Monthly mean SST factor decomposition fitting comprehensive output diagram.

Figure 4 shows that the monthly average SST in El Niño regions from 1990 to 2020 had a clear seasonal variation, and the monthly trend of SST can also clearly be observed from the seasonal index plot of the monthly mean SST.

The main factors affecting the temperature variation can be generally summarized as seasonal factors and random fluctuations. If the seasonal factor is removed, the SST should be left with only random fluctuations, i.e., the values fluctuate up and down at some mean value. At the same time, if the effect of random fluctuations is not considered, the SST is left with only the seasonal factor, i.e., the SST will have similar values in the same month in different years.

The following steps implement the kNN regression estimation method for the analysis of the SST data and display the comparison with the NW sum estimation method in Figure 5.

Step 1: Transform 372 months (31 years) of SST data $\{Z_i, i = 1, \dots, 372\}$ into functional data.

Step 2: Divide the 31 samples of data $(\chi_j, Y_j(s))_{j=1, \dots, 31}$ into two parts: 30 training samples of data $(\chi_j, Y_j(s))_{j=1, \dots, 30}$ for model fitting and 1 test sample of data $(\chi_{31}, Y_{31}(s))$ for prediction assessment.

Step 3: Here, the functional principal component analysis (FPCA) is applicable to semi-measures for rough curves such as SST data (see Chapter 3 of Ferraty et al. [25] for the methodology). A quadratic kernel function used in Section 3.1 is used in kNN regression.

Step 4: The SST values $(\hat{Y}_{31}(s), s = 1, \dots, 12)$ for 12 months in 2020 are predicted by the kNN method and the NW method, respectively, along with obtaining their MSEs for both methods.

Then, in step 1, we split the discrete monthly average temperature data of 372 months into 31 years of temperature profiles and express them as $\chi_i = \{Z_i(t), 12(j-1) < t < 12j\}, i = 1, \dots, 31$. Therefore, the response variable can be expressed as $Y_j(s) = \{Z_{12j+s}, s =$

$1, \dots, 12\}, j = 1, \dots, 30$. Thus, $(\chi_j, Y_j(s))_{j=1, \dots, 30}$ is the sample set of dependent function type with a sample size of 30, where χ_j is the function type data, and $Y_j(s)$ is a real value.

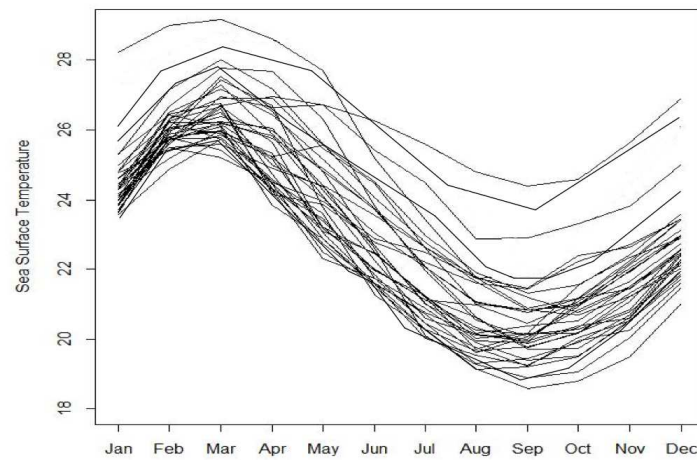


Figure 4. Time series curve of SST in El Niño during 31 years.

In Step 3, the choice of parameters q for the kNN method and NW method is performed via computation of cross-validation in R, which gives $q = 3$ and $q = 1$ for the kNN regression method and NW method, respectively. The selection of parameters k and h is similar to Section 3.1.

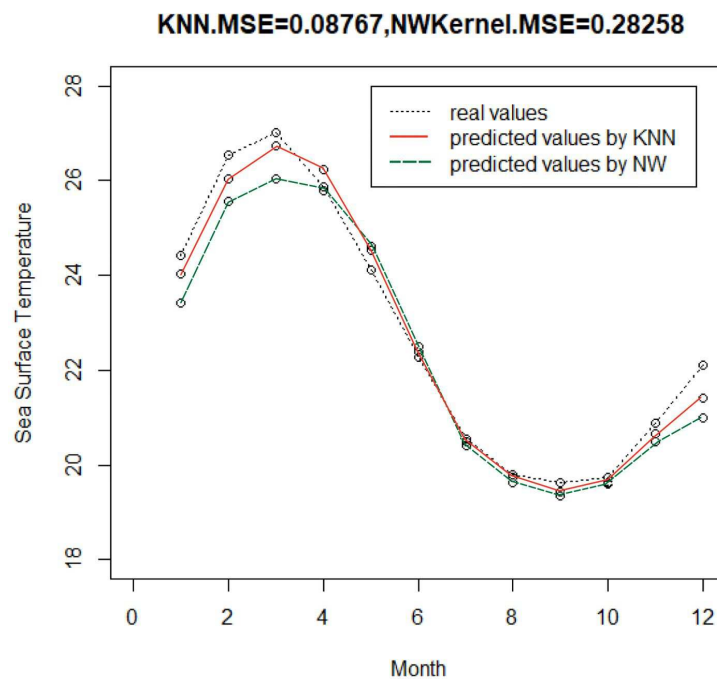


Figure 5. Forecast value of SST in 2020 by KNN method and NW method.

From Figure 5, which compares the MSE values calculated by the two methods, it can be seen that the MSE of the kNN method is much smaller than that of the NW method. Further, noting that the degree of fit between the curves fitted by the two methods to the true curve (dotted line), the predicted curves by two methods are generally closer to the true curve, indicating that the prediction effect of both methods is very good. However, a closer look reveals that the predicted values of the kNN method obviously have better fitting at the inflection points of the curves, such as January, February, March, November

and December, which fully reflect the fact that the kNN method pays more attention to the local variation than the NW method when processing the data like this, including the abnormal or extreme distribution of the response variable.

4. Proof of Theorem

In order to prove the main results, we give some lemmas. Let $(A_i, B_i)_{i=1,2,\dots,n}$ be n random pairs valued in $(\Omega \times \mathfrak{R}, \mathfrak{A} \times \mathfrak{B}(\mathfrak{R}))$, where (Ω, \mathfrak{A}) is a general measurable space. Let S_Ω be a fixed subset of Ω , $G(\cdot, (\chi, \cdot)) : \mathfrak{R} \times (S_\Omega \times \Omega) \rightarrow \mathfrak{R}^+$ be a measurable function, for $\forall t, t' \in \mathfrak{R}$,

$$(L_0) : t \leq t' \Rightarrow G(t, z) \leq G(t', z), \forall z \in S_\Omega \times \Omega,$$

$D_n(\chi)_{n \in \mathbb{N}}$ is a sequence of random real variables (r.r.v.), and $c(\cdot) : S_\Omega \rightarrow \mathfrak{R}$ is a nonrandom function such that $\sup_{\chi \in S_\Omega} |c(\chi)| < \infty$. For $\forall \chi \in S_\Omega, n \in \mathbb{N} / \{0\}$, we define:

$$c_{n,\chi}(t) = \frac{\sum_{i=1}^n B_i G(t, (\chi, A_i))}{\sum_{i=1}^n G(t, (\chi, A_i))}.$$

Lemma 1 ([10]). Let $u_n(\chi)_{n \in \mathbb{N}}$ be a decreasing positive real sequence satisfying $\lim_{n \rightarrow \infty} u_n = 0$. For any increasing sequences $\beta_n \in (0, 1)$ and $\beta_n - 1 = O(u_n)$, there exist two real random sequences $\{D_n^-(\beta_n, \chi)\}_{n \in \mathbb{N}}$ and $\{D_n^+(\beta_n, \chi)\}_{n \in \mathbb{N}}$ such that:

$$(L_1) D_n^-(\beta_n, \chi) \leq D_n^+(\beta_n, \chi), \forall n \in \mathbb{N}, \forall \chi \in S_\Omega,$$

$$(L_2) I_{\{D_n^-(\beta_n, \chi) \leq D_n(\beta_n, \chi) \leq D_n^+(\beta_n, \chi), \forall \chi \in S_\Omega\}} \rightarrow 1, a.co. n \rightarrow \infty,$$

$$(L_3) \sup_{\chi \in S_\Omega} \left| \frac{\sum_{i=1}^n G(D_n^-(\beta_n, \chi))}{\sum_{i=1}^n G(D_n^+(\beta_n, \chi))} - \beta_n \right| = O_{a.co.}(u_n),$$

$$(L_4) \sup_{\chi \in S_\Omega} |c_{n,\chi}(D_n^-(\beta_n, \chi)) - c(\chi)| = O_{a.co.}(u_n),$$

$$(L_5) \sup_{\chi \in S_\Omega} |c_{n,\chi}(D_n^+(\beta_n, \chi)) - c(\chi)| = O_{a.co.}(u_n),$$

then, we have:

$$\sup_{\chi \in S_\Omega} |c_{n,\chi}(D_n(\beta_n, \chi)) - c(\chi)| = O_{a.co.}(u_n).$$

The proof of Lemma 1 is not presented here because it follows, step by step, the same argument in Burba et al. [10], Kudraszow and Vieu [19].

Lemma 2 ([26]). Let $\{X_n, n \in \mathbb{N}\}$ be an NA random sequence with zero mean, and there exists a positive constant $c_k, k = 1, 2, \dots, n$ such that $|X_k| \leq c_k$, let $S_n = X_1 + X_2 + \dots + X_n$. For any $\epsilon > 0$, we get:

$$P(S_n \geq n\epsilon) \leq \exp \left\{ \frac{-n^2 \epsilon^2}{2 \sum_{i=1}^n c_i^2} \right\}, \tag{4}$$

and

$$P(|S_n| \geq n\epsilon) \leq 2 \exp \left\{ \frac{-n^2 \epsilon^2}{2 \sum_{i=1}^n c_i^2} \right\}.$$

Lemma 3. Suppose that Assumptions 1–6 hold, and $h_n(\chi) \rightarrow 0$ a.s. $n \rightarrow \infty$ in model (3) satisfying:

$$\lim(\varphi_\chi(H_{n,k}(\chi)) - \varphi_\chi(h_n(\chi))) = 0, \tag{5}$$

$$0 < C_1 h_n \leq \inf_{\chi \in S_{\mathfrak{F}}} h_n(\chi) \leq \sup_{\chi \in S_{\mathfrak{F}}} h_n(\chi) \leq C_2 h_n < \infty, \tag{6}$$

and for n large enough,

$$\frac{\log^2 n}{n\phi(h_n)} < \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right) < \frac{n\phi(h_n)}{\log n}, \tag{7}$$

$$0 < C_1 < \frac{n\phi(h_n)}{\log^2 n} < C_2 < \infty, \tag{8}$$

then we have:

$$\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_n(\chi) - m(\chi)| = O_{a.co.} \left(h_n^\beta \right) + O_{a.co.} \left(\sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}} \right), \tag{9}$$

where $\epsilon = \frac{\log n}{n}$.

Proof of Lemma 3. In order to simplify the proof, we introduce some notations in this article. For $\forall \chi \in S_{\mathfrak{F}}$, let $k(\chi) = \arg \min_{k=1,2,\dots,N_\epsilon(S_{\mathfrak{F}})} d(\chi, \chi_k)$, $s_n^2 = \max \{s_{n,1}^2, s_{n,2}^2, s_{n,3}^2, s_{n,4}^2\}$ be the mixed operator covariance,

$$s_{n,1}^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(Y_i u_i, Y_j u_j)|, s_{n,2}^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(v_i, v_j)|,$$

$$s_{n,3}^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(u_i, u_j)|, s_{n,4}^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(w_i, w_j)|,$$

where $u_i = I_{B(\chi_{k(\chi)}, C_2 h_n(\chi) + \epsilon)}$, $0 < h_n \rightarrow 0$,

$$v_i = \frac{Y_i K(h_n(\chi_k)^{-1} d(\chi_k, \chi_i))}{EK(h_n(\chi_k)^{-1} d(\chi_k, \chi_1))} - \frac{E(Y_i K(h_n(\chi_k)^{-1} d(\chi_k, \chi_i)))}{EK(h_n(\chi_k)^{-1} d(\chi_k, \chi_1))},$$

$$w_i = \frac{K(h_n(\chi_k)^{-1} d(\chi_k, \chi_i))}{EK(h_n(\chi_k)^{-1} d(\chi_k, \chi_1))} - \frac{EK(h_n(\chi_k)^{-1} d(\chi_k, \chi_i))}{EK(h_n(\chi_k)^{-1} d(\chi_k, \chi_1))}.$$

For the fixed $\chi \in S_{\mathfrak{F}}$ in model (3), we have the decomposition as follows:

$$\begin{aligned} \hat{m}_n(\chi) - m(\chi) &= \frac{\hat{m}_{2n}(\chi)}{\hat{m}_{1n}(\chi)} - m_n(\chi) \\ &= \frac{1}{\hat{m}_{1n}(\chi)} [\hat{m}_{2n}(\chi) - E\hat{m}_n(\chi)] + \frac{1}{\hat{m}_{1n}(\chi)} [E\hat{m}_n(\chi) - m_n(\chi)] \\ &\quad + \frac{m_n(\chi)}{\hat{m}_{1n}(\chi)} [1 - \hat{m}_{1n}(\chi)]. \end{aligned}$$

where:

$$\hat{m}_{1n}(\chi) = \frac{\sum_{i=1}^n K(h_n(\chi)^{-1} d(\chi, \chi_i))}{nEK(h_n(\chi)^{-1} d(\chi, \chi_1))}, \hat{m}_{2n}(\chi) = \frac{\sum_{i=1}^n Y_i K(h_n(\chi)^{-1} d(\chi, \chi_i))}{nEK(h_n(\chi)^{-1} d(\chi, \chi_1))}.$$

It suffices to prove the three following results in order to establish (9),

$$\sup_{\chi \in S_{\mathfrak{F}}} |E\hat{m}_{2n}(\chi) - m(\chi)| = O_{a.co.} \left(h_n^\beta \right), \tag{10}$$

$$\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{2n}(\chi) - E\hat{m}_{2n}(\chi)| = O_{a.co.} \left(\sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}} \right), \tag{11}$$

$$\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{1n}(\chi) - 1| = \sup_{\chi \in S_F} |\hat{m}_{1n}(\chi) - E\hat{m}_{1n}(\chi)| = O_{a.co.} \left(\sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}} \right). \tag{12}$$

As to the Equation (10). For $\chi \in S_{\mathfrak{F}}$, by the Equation (6) and Assumption 4, it follows that:

$$\begin{aligned} |E\hat{m}_{2n}(\chi) - m(\chi)| &= \left| E \frac{\sum_{i=1}^n Y_i K(h_n(\chi)^{-1}d(\chi, \chi_i))}{nEK(h_n(\chi)^{-1}d(\chi, \chi_1))} - m(\chi) \right| \\ &= \left| \frac{EY_1 K(h_n(\chi)^{-1}d(\chi, \chi_1))}{EK(h_n(\chi)^{-1}d(\chi, \chi_1))} - m(\chi) \right| \\ &= \left| \frac{m(\chi_1)EK(h_n(\chi)^{-1}d(\chi, \chi_1)) - m(\chi)EK(h_n(\chi)^{-1}d(\chi, \chi_1))}{EK(h_n(\chi)^{-1}d(\chi, \chi_1))} \right| \\ &= |m(\chi_1) - m(\chi)| = O_{a.co.} (h_n^\beta). \end{aligned}$$

Then, we need to show the Equation (11). In fact, we have the decomposition as follows:

$$\begin{aligned} &\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{2n}(\chi) - E\hat{m}_{2n}(\chi)| \\ &\leq \sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{2n}(\chi) - \hat{m}_{2n}(\chi_{k(\chi)})| + \sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{2n}(\chi_{k(\chi)}) - E\hat{m}_{2n}(\chi_{k(\chi)})| \\ &+ \sup_{\chi \in S_{\mathfrak{F}}} |E\hat{m}_{2n}(\chi_{k(\chi)}) - E\hat{m}_{2n}(\chi)| =: I_1 + I_2 + I_3. \end{aligned}$$

For I_1 , by Assumption 3, it is easily seen that:

$$0 < C_1 < EK(h_n(\chi)^{-1}d(\chi, \chi_i)) < C_2 < \infty,$$

thus,

$$\begin{aligned} I_1 &= \sup_{\chi \in S_{\mathfrak{F}}} \left| \frac{\sum_{i=1}^n Y_i K(h_n(\chi)^{-1}d(\chi, \chi_i))}{nEK(h_n(\chi)^{-1}d(\chi, \chi_1))} - \frac{\sum_{i=1}^n Y_i K(h_n(\chi_{k(\chi)})^{-1}d(\chi_{k(\chi)}, \chi_i))}{nEK(h_n(\chi_{k(\chi)})^{-1}d(\chi_{k(\chi)}, \chi_1))} \right| \\ &= \sup_{\chi \in S_{\mathfrak{F}}} \frac{1}{n} \sum_{i=1}^n Y_i \left| \frac{K(h_n(\chi)^{-1}d(\chi, \chi_i))}{EK(h_n(\chi)^{-1}d(\chi, \chi_1))} - \frac{K(h_n(\chi_{k(\chi)})^{-1}d(\chi_{k(\chi)}, \chi_i))}{EK(h_n(\chi_{k(\chi)})^{-1}d(\chi_{k(\chi)}, \chi_1))} \right| \\ &\leq C \sup_{\chi \in S_{\mathfrak{F}}} \frac{1}{n} \sum_{i=1}^n Y_i \left| K\left(\frac{d(\chi, \chi_i)}{h_n(\chi)}\right) - K\left(\frac{d(\chi_{k(\chi)}, \chi_i)}{h_n(\chi_{k(\chi)})}\right) \right| I_{B(\chi, h_n(\chi)) \cup B(\chi_{k(\chi)}, h_n(\chi))}(\chi_i) \\ &\leq C \sup_{\chi \in S_{\mathfrak{F}}} \frac{1}{n} \sum_{i=1}^n Y_i I_{B(\chi_{k(\chi)}, C_2 h_n(\chi) + \epsilon)}(\chi_i), \end{aligned}$$

for $\forall \eta > 0$, we have:

$$\begin{aligned} & P\left(I_1 > \eta \sqrt{\frac{s_{n,1}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right) \\ & \leq P\left(C \sup_{\chi \in S_{\mathfrak{F}}} \frac{1}{n} \sum_{i=1}^n Y_i I_{B(\chi_{k(\chi)}, C_2 h_n(\chi) + \epsilon)}(\chi_i) > \eta \sqrt{\frac{s_{n,1}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right) \\ & \leq CN_{\epsilon}(S_{\mathfrak{F}}) \max_{k \in \{\chi_1, \chi_2, \dots, \chi_{N_{\epsilon}(S_{\mathfrak{F}})}\}} P\left(\sum_{i=1}^n |Y_i| I_{B(\chi_{k(\chi)}, C_2 h_n(\chi) + \epsilon)}(\chi_i) > \eta \sqrt{s_{n,1}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}\right). \end{aligned}$$

According to (4) in Lemma 2 and Assumption 6, we have:

$$\begin{aligned} & P\left(\sum_{i=1}^n |Y_i| I_{B(\chi_{k(\chi)}, C_2 h_n(\chi) + \epsilon)}(\chi_i) > \eta \sqrt{s_{n,1}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}\right) \\ & \leq \exp\left\{\frac{-\eta s_{n,1}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{2 \sum_{i=1}^n c_i^2}\right\} \leq \exp\left\{\left(1 - \frac{\eta s_{n,1}^2}{2 \sum_{i=1}^n c_i^2}\right) - \Psi_{S_{\mathfrak{F}}}(\epsilon)\right\} < \infty. \end{aligned}$$

Hence, it follows that:

$$I_1 = O_{a.co.}\left(\sqrt{\frac{s_{n,1}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right). \tag{13}$$

For I_2 , similar to the proof of I_1 , for $\forall \eta > 0$, we have:

$$\begin{aligned} & P\left(I_2 > \eta \sqrt{\frac{s_{n,2}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right) \\ & = P\left(\sup_{\chi \in S_{\mathfrak{F}}} \left| \frac{\sum_{i=1}^n Y_i K\left(\frac{d(\chi_{k(\chi)}, \chi_i)}{h_n(\chi_{k(\chi)})}\right)}{nEK\left(\frac{d(\chi_{k(\chi)}, \chi_1)}{h_n(\chi_{k(\chi)})}\right)} - \frac{E \sum_{i=1}^n Y_i K\left(\frac{d(\chi_{k(\chi)}, \chi_i)}{h_n(\chi_{k(\chi)})}\right)}{nEK\left(\frac{d(\chi_{k(\chi)}, \chi_1)}{h_n(\chi_{k(\chi)})}\right)} \right| > \eta \sqrt{\frac{s_{n,2}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right) \\ & = P\left(\sup_{\chi \in S_{\mathfrak{F}}} \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i K\left(\frac{d(\chi_{k(\chi)}, \chi_i)}{h_n(\chi_{k(\chi)})}\right)}{EK\left(\frac{d(\chi_{k(\chi)}, \chi_1)}{h_n(\chi_{k(\chi)})}\right)} - \frac{EY_i K\left(\frac{d(\chi_{k(\chi)}, \chi_i)}{h_n(\chi_{k(\chi)})}\right)}{EK\left(\frac{d(\chi_{k(\chi)}, \chi_1)}{h_n(\chi_{k(\chi)})}\right)} \right| > \eta \sqrt{\frac{s_{n,2}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right) \\ & \leq N_{\epsilon}(S_{\mathfrak{F}}) \max_{k \in \{\chi_1, \chi_2, \dots, \chi_{N_{\epsilon}(S_{\mathfrak{F}})}\}} P\left(\frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i K\left(\frac{d(\chi_{k(\chi)}, \chi_i)}{h_n(\chi_{k(\chi)})}\right)}{EK\left(\frac{d(\chi_{k(\chi)}, \chi_1)}{h_n(\chi_{k(\chi)})}\right)} - \frac{EY_i K\left(\frac{d(\chi_{k(\chi)}, \chi_i)}{h_n(\chi_{k(\chi)})}\right)}{EK\left(\frac{d(\chi_{k(\chi)}, \chi_1)}{h_n(\chi_{k(\chi)})}\right)} \right| > \eta \sqrt{\frac{s_{n,2}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right) \\ & \leq CN_{\epsilon}(S_{\mathfrak{F}}) \max_{k \in \{\chi_1, \chi_2, \dots, \chi_{N_{\epsilon}(S_{\mathfrak{F}})}\}} P\left(\frac{1}{n} \sum_{i=1}^n |Y_i| I_{B(\chi_{k(\chi)}, C_2 h_n(\chi) + \epsilon)}(\chi_i) > \eta \sqrt{\frac{s_{n,2}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right). \end{aligned}$$

Thus,

$$I_2 = O_{a.co.}\left(\sqrt{\frac{s_{n,2}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}}\right). \tag{14}$$

Finally, for I_3 , we can get $I_3 \leq E\left(\sup_{\chi \in S_{\mathfrak{F}}} \left| \hat{m}_{2n}(\chi_{k(\chi)}) - \hat{m}_{2n}(\chi) \right| \right)$. The proof process is similar to I_1 , and we can obtain:

$$I_3 = O_{a.co.} \left(\sqrt{\frac{s_{n,1}^2 \Psi_{S_{\mathfrak{F}}}(\epsilon)}{n^2}} \right). \tag{15}$$

Therefore, combining the Equations (13)–(15), the Equation (11) can be established. Similarly, we may prove the Equation (12). Hence, the proof of Lemma 3 is completed. \square

Proof of Theorem 1. According to Lemma 1, let $S_{\Omega} = S_{\mathfrak{F}}$, $A_i = \chi_i$, $B_i = Y_i$, $G(t, (\chi, A_i)) = K(t^{-1}d(\chi, \chi_i))$, $D_n(\chi) = H_{n,k}(\chi)$, $c_{n,\chi}(\chi) = \hat{m}_{kNN}(\chi)$, $c(\chi) = \hat{m}_n(\chi)$. Let $\beta_n \in (0, 1)$ be an increasing sequence such that $\beta_n - 1 = O(u_n)$, where $u_n = \phi^{-1}\left(\frac{k_n}{n}\right)^{\beta} + \sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}}\left(\frac{\log n}{n}\right)}{n^2}}$ is a decreasing positive real sequence such that $\lim_{n \rightarrow \infty} u_n = 0$ and $h_n = \phi^{-1}\left(\frac{k_n}{n}\right)^{\beta}$. Let $\{D_n^-(\beta_n, \chi)\}_{n \in \mathbb{N}}$ and $\{D_n^+(\beta_n, \chi)\}_{n \in \mathbb{N}}$ be two real random sequences such that:

$$\varphi_{\chi}(D_n^-(\beta_n, \chi)) = \varphi_{\chi}(h_n(\chi))\beta_n^{\frac{1}{2}}, \tag{16}$$

$$\varphi_{\chi}(D_n^+(\beta_n, \chi)) = \varphi_{\chi}(h_n(\chi))\beta_n^{-\frac{1}{2}}, \tag{17}$$

Firstly, we verify the conditions (L_4) and (L_5) in Lemma 1. By $\varphi_{\chi}(D_n^-(\beta_n, \chi))$ and $\beta_n - 1 = O(u_n)$, it is easy to follow that the local bandwidth $D_n^-(\beta_n, \chi)$ satisfies the condition (5). Combining $h_n = \phi^{-1}\left(\frac{k_n}{n}\right)^{\beta}$ with Assumption 2, it follows that $h_n(\chi)$ satisfies the condition (6). Let $k_n = n\phi(h_n)$, from Assumption 2(i) we obtain that $\frac{k_n}{n} = \frac{n\phi(h_n)}{n} = \phi(h_n)$ is satisfied. Hence, according to the conditions of the Theorem, the Equations (7) and (8) in Lemma 3 hold. Thus, by Lemma 3, we have:

$$\sup_{\chi \in S_{\mathfrak{F}}} |c_{n,\chi}(D_n^-(\beta_n, \chi)) - c(\chi)| = O_{a.co.} \left(\phi^{-1}\left(\frac{k_n}{n}\right)^{\beta} + \sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}}\left(\frac{\log n}{n}\right)}{n^2}} \right) = O_{a.co.}(u_n).$$

Similarly, for $D_n^+(\beta_n, \chi)$, we can also get:

$$\sup_{\chi \in S_{\mathfrak{F}}} |c_{n,\chi}(D_n^+(\beta_n, \chi)) - c(\chi)| = O_{a.co.} \left(\phi^{-1}\left(\frac{k_n}{n}\right)^{\beta} + \sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}}\left(\frac{\log n}{n}\right)}{n^2}} \right) = O_{a.co.}(u_n).$$

Secondly, checking the conditions (L_1) and (L_2) in Lemma 1, and combining (16) and (17) with $\beta_n \in (0, 1)$, it is clearly followed that:

$$\varphi_{\chi}(D_n^-(\beta_n, \chi)) \leq \varphi_{\chi}(h_n(\chi)) \leq \varphi_{\chi}(D_n^+(\beta_n, \chi)), \tag{18}$$

By Assumption 1 we get:

$$D_n^-(\beta_n, \chi) \leq h_n(\chi) \leq D_n^+(\beta_n, \chi).$$

According to (5) and (18), for $n \rightarrow \infty$, we have:

$$\varphi_{\chi}(D_n^-(\beta_n, \chi)) \leq \varphi_{\chi}(H_{n,\chi}(\chi)) \leq \varphi_{\chi}(D_n^+(\beta_n, \chi)),$$

That is:

$$\varphi_{\chi}(D_n^-(\beta_n, \chi)) \leq \varphi_{\chi}(D_n(\chi)) \leq \varphi_{\chi}(D_n^+(\beta_n, \chi)),$$

Therefore, by Assumption 1 we can get:

$$D_n^-(\beta_n, \chi) \leq D_n(\chi) \leq D_n^+(\beta_n, \chi),$$

Thus,

$$I_{\{D_n^-(\beta_n, \chi) \leq D_n(\beta_n, \chi) \leq D_n^+(\beta_n, \chi), \forall \chi \in S_{\mathfrak{F}}\}} \rightarrow 1, a.co. n \rightarrow \infty.$$

(L₂) is checked.

Finally, we establish the condition (L₃) in Lemma 1. Similar to Kudraszow and Vieu [19], we denote:

$$f^*(\chi, h_n(\chi)) = EK(h_n(\chi)^{-1}d(\chi, \chi_1)), \forall \chi \in S_{\mathfrak{F}}.$$

and let:

$$F_1 = \frac{f^*(\chi, D_n^-(\beta_n, \chi))}{f^*(\chi, D_n^+(\beta_n, \chi))}, F_2 = \frac{\hat{m}_{1n}(\chi, D_n^-(\beta_n, \chi))}{\hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi))} - 1, F_3 = \frac{f^*(\chi, D_n^+(\beta_n, \chi))}{f^*(\chi, D_n^-(\beta_n, \chi))} \beta_n - 1.$$

Then, (L₃) can be decomposed as follows:

$$\begin{aligned} & \sup_{\chi \in S_{\mathfrak{F}}} \left| \frac{\sum_{i=1}^n G(D_n^-(\beta_n, \chi))}{\sum_{i=1}^n G(D_n^+(\beta_n, \chi))} - \beta_n \right| \\ &= \left| \frac{\hat{m}_{1n}(\chi, D_n^-(\beta_n, \chi))}{\hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi))} \frac{f^*(\chi, D_n^-(\beta_n, \chi))}{f^*(\chi, D_n^+(\beta_n, \chi))} - \frac{f^*(\chi, D_n^+(\beta_n, \chi))}{f^*(\chi, D_n^-(\beta_n, \chi))} \frac{f^*(\chi, D_n^-(\beta_n, \chi))}{f^*(\chi, D_n^+(\beta_n, \chi))} \beta_n \right| \\ &= \left| \frac{f^*(\chi, D_n^-(\beta_n, \chi))}{f^*(\chi, D_n^+(\beta_n, \chi))} \left| \frac{\hat{m}_{1n}(\chi, D_n^-(\beta_n, \chi))}{\hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi))} - \frac{f^*(\chi, D_n^+(\beta_n, \chi))}{f^*(\chi, D_n^-(\beta_n, \chi))} \beta_n \right| \right| \\ &\leq \left| \frac{f^*(\chi, D_n^-(\beta_n, \chi))}{f^*(\chi, D_n^+(\beta_n, \chi))} \right| \left(\left| \frac{\hat{m}_{1n}(\chi, D_n^-(\beta_n, \chi))}{\hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi))} - 1 \right| + \left| \frac{f^*(\chi, D_n^+(\beta_n, \chi))}{f^*(\chi, D_n^-(\beta_n, \chi))} \beta_n - 1 \right| \right) \\ &\triangleq |F_1| |F_2| + |F_1| |F_3|. \end{aligned} \tag{19}$$

By Assumption 3, it is followed that:

$$\sup_{\chi \in S_{\mathfrak{F}}} |F_1| \leq C, \tag{20}$$

and for $\forall \chi \in S_{\mathfrak{F}}, \hat{m}_{1n}(\chi) = \frac{\sum_{i=1}^n K(h_n(\chi)^{-1}d(\chi, \chi_i))}{nEK(h_n(\chi)^{-1}d(\chi, \chi_1))}$, referring to Ferraty et al. [25], we have:

$$\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{1n}(\chi) - 1| = O_{a.co.} \left(\sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right)}{n^2}} \right),$$

Therefore,

$$\begin{aligned} \sup_{\chi \in S_{\mathfrak{F}}} |F_2| &= \sup_{\chi \in S_{\mathfrak{F}}} \left| \frac{\hat{m}_{1n}(\chi, D_n^-(\beta_n, \chi))}{\hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi))} - 1 \right| \\ &= \sup_{\chi \in S_{\mathfrak{F}}} \left| \frac{\hat{m}_{1n}(\chi, D_n^-(\beta_n, \chi)) - 1 + 1 - \hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi))}{\hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi))} \right| \\ &\leq \frac{\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{1n}(\chi, D_n^-(\beta_n, \chi)) - 1| + \sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi)) - 1|}{\inf_{\chi \in S_{\mathfrak{F}}} \hat{m}_{1n}(\chi, D_n^+(\beta_n, \chi))} \quad (21) \\ &= O_{a.co.} \left(\sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right)}{n^2}} \right). \end{aligned}$$

Moreover, for F_3 , according to Lemma 1 in Ezzahrioui and Ould-said [13] and Assumption 2(iii), there exists $\tau > 0$, for $\forall \chi \in S_{\mathfrak{F}}$,

$$f^*(\chi, h_n(\chi)) = \phi(h_n(\chi))\tau f(\chi) + O\left(\phi(h_n(\chi))h_n(\chi)^\beta\right) = \tau\phi(h_n(\chi)) + O\left(\phi(h_n)h_n^\beta\right),$$

by $\frac{\phi(D_n^-(\beta_n, \chi))}{\phi(D_n^+(\beta_n, \chi))} = \beta_n$, $\sup_{\chi \in S_{\mathfrak{F}}} |F_3| = O\left(\phi(h_n)h_n^\beta\right) = O\left(\sqrt{\beta_n}\phi^{-1}\left(\frac{k_n}{n}\right)^\beta\right)$ holds. Hence, for $\beta_n \rightarrow 1$, it follows that

$$\sup_{\chi \in S_{\mathfrak{F}}} |F_3| = O\left(\phi^{-1}\left(\frac{k_n}{n}\right)^\beta\right). \quad (22)$$

Combining (19)–(22), we obtain:

$$\sup_{\chi \in S_{\mathfrak{F}}} \left| \frac{\sum_{i=1}^n G(D_n^-(\beta_n, \chi))}{\sum_{i=1}^n G(D_n^+(\beta_n, \chi))} - \beta_n \right| = O_{a.co.}(u_n).$$

(L_3) is established.

Thus, the conditions (L1)–(L5) in Lemma 1 have been established. By Lemma 1, we can get:

$$\sup_{\chi \in S_{\mathfrak{F}}} |\hat{m}_{KNN}(\chi) - m(\chi)| = O_{a.co.} \left(\phi^{-1}\left(\frac{k_n}{n}\right)^\beta + \sqrt{\frac{s_n^2 \Psi_{S_{\mathfrak{F}}} \left(\frac{\log n}{n} \right)}{n^2}} \right).$$

The proof of the Theorem 1 is completed. \square

5. Conclusions and Future Research

Functional data analysis deals with the analysis and theory of data that are in the form of functions, images and shapes, or more general objects. In a way, correlation is really the heart of data science. The correlation between variables may be complicated, from simply independent to α -mixing or something else, such as negatively associated (NA). The kNN method, as one of the nonparametric methods, is very useful in statistical estimation and machine learning. While regression analysis of functional data under many variable correlated cases, except NA sequences, has been explored. This paper builds a kNN regression estimator of the functional regression model. In particular, we obtain the almost complete convergence rate of kNN estimation. Some simulated experiments and real data analyses illustrate the feasibility and the finite-sample behavior of the method.

Further work includes introducing the kNN machine learning algorithm for functional data analysis and kNN high-dimensional modeling with NA sequences.

Author Contributions: Conceptualization, X.H. and J.W.; methodology, X.H.; software, J.W.; writing—original draft preparation, X.H. and J.W.; writing—review and editing, K.Y. and L.W.; visualization, K.Y.; supervision, X.H.; project administration, K.Y.; funding acquisition, K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Foundation (Grant No. 21BTJ040).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: <https://www.cpc.ncep.noaa.gov/data/indices/> (accessed on 9 January 2022).

Acknowledgments: The authors are most grateful to the Editor and anonymous referee for carefully reading the manuscript and for valuable suggestions which helped in improving an earlier version of this paper. This research was funded by the National Social Science Foundation (Grant No. 21BTJ040).

Conflicts of Interest: The authors declare no conflict of interest in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

NA	Negatively Associated
kNN	k-Nearest Neighbor

References

1. Ramsay, J.; Dalzell, C. Some Tools for Functional Data Analysis. *J. R. Stat. Soc. Ser. B Methodol.* **1991**, *53*, 539–561. [\[CrossRef\]](#)
2. Ramsay, J.; Silverman, B. *Functional Data Analysis*; Springer: New York, NY, USA, 1997.
3. Ramsay, J.; Silverman, B. *Functional Data Analysis*, 2nd ed.; Springer: New York, NY, USA, 2005.
4. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis*; Springer: New York, NY, USA, 2006.
5. Ferraty, F.; Vieu, P. Nonparametric Models for Functional Data, with Application in Regression, Time Series Prediction and Curve Discrimination. *J. Nonparametr. Stat.* **2004**, *16*, 111–125. [\[CrossRef\]](#)
6. Ling, N.X.; Wu, Y.H. Consistency of Modified Kernel Regression Estimation with Functional Data. *Statistics* **2012**, *46*, 149–158. [\[CrossRef\]](#)
7. Baíllo, A.; Grané, A. Local Linear Regression for Functional Predictor and Scalar Response. *J. Multivar. Anal.* **2009**, *100*, 102–111. [\[CrossRef\]](#)
8. Fix, E.; Hodges, J. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *Inter. Stat. Re.* **1989**, *57*, 238–247. [\[CrossRef\]](#)
9. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185.
10. Burba, F.; Ferraty, F.; Vieu, P. k-Nearest Neighbour Method in Functional Nonparametric Regression. *J. Nonparametr. Stat.* **2009**, *21*, 453–469. [\[CrossRef\]](#)
11. Masry, E. Nonparametric Regression Estimation for Dependent Functional Data: Asymptotic Normality. *Stoch. Pro* **2005**, *115*, 155–177. [\[CrossRef\]](#)
12. Laib, N.; Louani, D. Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *J. Stat. Plan. Inference* **2011**, *141*, 359–372. [\[CrossRef\]](#)
13. Ezzahrioui, M.; Ould-Sad, E. Asymptotic Normality of a Nonparametric Estimator of the Conditional Mode Function for Functional Data. *J. Nonparametr. Stat.* **2008**, *20*, 3–18. [\[CrossRef\]](#)
14. Ezzahriouia, M.; Ould-Said, E. Some Asymptotic Results of a Nonparametric Conditional Mode Estimator for Functional Time-Series data. *Stat. Neerl.* **2010**, *64*, 171–201. [\[CrossRef\]](#)
15. Horvath, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer: New York, NY, USA, 2012.
16. Ling, N.X.; Wang, C.; Ling, J. Modified Kernel Regression Estimation with Functional Time Series data. *Stat. Probab. Lett.* **2016**, *114*, 78–85. [\[CrossRef\]](#)
17. Abdelmalek, G.; Abdelhak, C. Strong uniform consistency rates of the local linear estimation of the conditional hazard estimator for functional data. *Int. J. Appl. Math. Stat.* **2020**, *59*, 1–13. [\[CrossRef\]](#)
18. Mustapha, M.; Salim, B.; Ali, L. The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data. *J. Multivar. Anal.* **2021**, *181*.

19. Kudraszow, N.L.; Vieu, P. Uniform Consistency of kNN Regressors for Functional Variables. *Stat. Probab. Lett.* **2013**, *83*, 1863–1870. [[CrossRef](#)]
20. Kara, L.Z.; Laksaci, A.; Rachdi, M.; Vieu, F. Data-driven kNN Estimation in Nonparametric Functional Data Analysis. *J. Multivar. Anal.* **2017**, *153*, 176–188. [[CrossRef](#)]
21. Joag-Dev, K.; Proschan, F. Negative Association of Random Variables with Application. *Ann. Stat.* **1983**, *11*, 286–295. [[CrossRef](#)]
22. Yi, W.; Wang, X.; Sung, S.H. Complete Moment Convergence for Arrays of Rowwise Negatively Associated Random Variables and its Application in Non-parametric Regression Model. *Probab. Eng. Inf. Sci.* **2017**, *32*, 37–57. [[CrossRef](#)]
23. Delsol, L. Advances on Asymptotic Normality in Nonparametric Functional Time Series Analysis. *Statistics* **2009**, *43*, 13–33.
24. Ferraty, F.; Rabhi, A.; Vieu, P. Conditional Quantiles for Dependent Functional Data with Application to the Climatic El Niño Phenomenon. *Sankhyā Indian J. Stat.* **2005**, *67*, 378–398. [[CrossRef](#)]
25. Ferraty, F.; Laksaci, A.; Tadj, A.; Vieu, P. Rate of Uniform Consistency for Nonparametric Estimates with Functional Variables. *J. Stat. Plan. Inference* **2010**, *140*, 335–352.
26. Christofides, T.C.; Hadjikyriakou, M. Exponential Inequalities for N-demimartingales and Negatively Associated Random Variables. *Stat. Probab. Lett.* **2009**, *79*, 2060–2065. [[CrossRef](#)]