Nonparametric regression of spatio-temporal data using infinite-dimensional covariates

Subhrajyoty Roy

Department of Statistics and Data Science, Washington University at St. Louis, USA

and

Soudeep Deb

Decision Sciences Area, Indian Institute of Management Bangalore, India

and

Sayar Karmakar

Department of Statistics, University of Florida, USA

and

Rishideep Roy

School of Mathematics, Statistics and Actuarial Science, University of Essex, UK

March 9, 2026

Abstract

In spatio-temporal analysis, we often record data at specific time intervals but with varying spatial locations between these timepoints. We propose a conditional model to analyze such spatio-temporal data that accommodates the dependencies alongside second-order stationary explanatory variables, which may be infinite-dimensional and accommodate spatio-temporal covariates. Because of the absence of a mixing-type dependence condition in this case, which is typically required by the existing studies, we consider a weaker polynomially decaying moment contraction (PMC) condition on the covariates. In this paper, we obtain nonparametric point estimates of the mean and covariate functions of such a regression model, which we then show to be statistically consistent. We also obtain a simultaneous confidence interval of the mean function using the central limit theorem for the proposed estimator. Such simultaneous inference tools can be used to test for certain specifications of the mean function. Some simulation studies and two real-data analyses have been illustrated to corroborate the findings.

Keywords: Simultaneous confidence bands; Spatio-temporal process; Kernel estimation; Infinite dimensional regression

1 Introduction

Spatiotemporal datasets increasingly arise from monitoring systems whose sampling designs evolve over time. At each timepoint, measurements are recorded at a collection of spatial sites that can change as sensors fail, new sensors are deployed, or access constraints vary. At the same time, analysts often observe explanatory information aggregated at the temporal level (e.g., meteorological fields, land-use surfaces, remote-sensing products, GPS tracking, or historical records), whose natural representation is functional or otherwise high or possibly infinite-dimensional. These features, namely irregular time-stamps, time-varying spatial sampling locations, and infinite-dimensional temporal-level covariates, create a setting in which classical spatiotemporal methods and existing functional regression approaches do not directly deliver nonparametric inference on the conditional mean and volatility surfaces.

To better understand this framework, as the first practical use-case, we showcase our inference strategy on an air pollution dataset from New Delhi, the capital of India. This dataset exhibits irregularly spaced observations of PM2.5 (airborne particles with a diameter smaller than 2.5 micrometers), with missing data across time due to inactivity of sensors: see Figure 1 which shows this missingness. Naturally, it demands an infinite-dimensional covariate process due to the growing nature of historical observations and factors like measurements of other pollutants, weather patterns observed in a different set but nearby locations, etc., thus highlighting the need for a model that allows for infinite-dimensional covariates.

Figure 1: A heatplot indicating the missingness of the PM2.5 observations across locations (columns) and timepoints (rows).

For the second motivating example, we turn our attention to a very different context that stems from the analysis of shots in soccer. In Figure 2, the locations of all shots taken in the matches against a specific team (Chelsea) by three different teams (Arsenal, Liverpool, and Manchester City) are illustrated. It is clear that the shots are irregularly distributed over the entire playing field, with no fixed spatial design. One may aim to analyze the quality of a shot, commonly measured through expected goals or xG (Mead et al. 2023), based on various contextual variables over the season. These covariates are naturally indexed over time, and accumulate as the season progresses, yielding a high (or infinite) dimensional structure. Together, the irregular spatial sampling of shot locations and the

complex temporal evolution of contextual information pose challenges for standard spatial or functional regression methods.
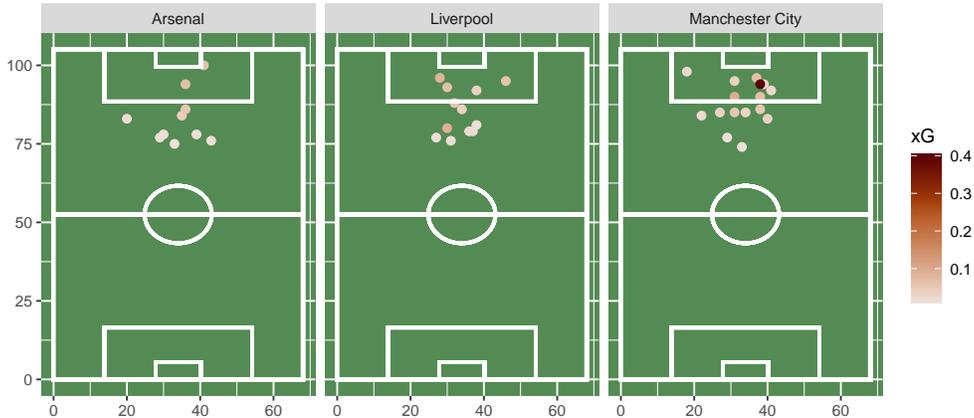

Figure 2: Position and expected goals (xG) of shots taken against Chelsea during the league 2014-15 by the players of Arsenal, Liverpool, and Manchester City.

To define the mathematical framework for similar applications, let $Y_t(s) \in \mathbb{R}$ denote the observed response at time t and location $s \in \mathcal{S}$, and let $\boldsymbol{X}_t$ be a second-order stationary covariate taking values in an abstract space $\chi \subseteq \mathbb{R}^\infty$. While we only keep the time index in the subscript to focus on the temporal evolution, we emphasize that $\boldsymbol{X}_t$, being infinite-dimensional, can accommodate spatio-temporal covariates and lagged responses by including explanatory variables across all locations. We formulate the conditional regression model

$$Y_t(s) = \mu(\boldsymbol{X}_t, s) + \sigma(\boldsymbol{X}_t, s)\, \epsilon_t(s). \tag{1}$$

where the timepoints t index the temporal horizon $\mathcal{T}$ and locations s index the spatial horizon $\mathcal{S}$ which is considered to be a compact subset of $\mathbb{R}^d$. In real-life problems, $\mathcal{S}$ is often a compact subset of $\mathbb{R}^2$ or $\mathbb{R}^3$, depending on the dataset at hand. Without loss of generality, following the principles of in-fill asymptotics, we can assume $\mathcal{T} = [0, 1]$ and $\mathcal{S} = [0, 1]^d$ for a fixed d after suitably scaling the temporal and spatial domains. The mean function $\mu(\cdot, \cdot)$ and the covariance function $\sigma(\cdot, \cdot)$ are both defined on $\chi \times \mathcal{S}$. They are assumed to be measurable and sufficiently smooth. The error component $\epsilon_t(s)$ in (1) captures idiosyncratic variation with a nontrivial spatial correlation. Specifically, as temporal dependence is entirely captured by the covariates $\boldsymbol{X}_t$, we consider $\epsilon_t(s)$ to be temporally independent and identically distributed (iid) random variables satisfying $\mathbb{E}(\epsilon_t(s)) = 0$. They are also independent of the past covariates $\boldsymbol{X}_{t'}$ for any $t' \leqslant t$ and for all $s \in \mathcal{S}$. These will be discussed in detail in Section 2.1.

Although the framework (1) models the response variable $Y_t(s)$ for all timepoints $t \in \mathcal{T}$ and spatial points $s \in \mathcal{S}$, in practice, the continuous curves are not typically observed. We assume that the data are collected at some irregularly spaced timepoints and at some specific spatial locations that may differ from one timepoint to another. Keeping that in view, we shall represent the dataset with observations from a total of n timepoints as

$$\mathcal{D} = \{(t_i, s_{t_i j}, \boldsymbol{X}_{t_i}, Y_{t_i}(s_{t_i j})) : 0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1, j \in \{1, 2, \ldots, n_i\}\}, \tag{2}$$

where the ith timepoint t_i observes data from n_i locations. The goal is to draw nonparametric inference about $\mu(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ given $\mathcal{D}$, and consequently, use the model (1) to make predictions of the response function at unknown locations or at a future timepoint.

Our model framework can be regarded as a significant generalization of the conventional stochastic time series regression model (see, e.g., Zhao & Wu 2008) in two ways. First, we incorporate the infinite-dimensional covariate $\boldsymbol{X}_t$ into the model. Second, we extend the time series framework to a spatiotemporal framework, allowing a non-trivial spatial dependency structure. Of course, it provides an extremely general setting, and includes the popular autoregressive setups (e.g. AR, ARCH) by letting $\boldsymbol{X}_t$ to be a spatially aggregated function of the lagged response values. It is easy to note that allowing $\boldsymbol{X}_t$ to be infinite-dimensional automatically includes the finite-dimensional case of the covariates $\boldsymbol{X}_t \in \mathbb{R}^d$ for some fixed $d > 0$, by choosing the coordinates of $\boldsymbol{X}_t$ to be equal to 0 beyond the fixed dimension d.

It is imperative to note that the existing literature addresses only fragments of our setting. Functional regression in an iid setting may accommodate infinite-dimensional covariates but not temporal dependence; functional time series methods accommodate dependence but typically assume regular time grids and do not treat spatially indexed responses observed on time-varying location sets; and nonparametric spatiotemporal regression focuses on finite-dimensional covariates and regular grids. We bridge this research gap and develop frequentist nonparametric estimation and inference for the mean and volatility surfaces in (1) under the irregular design (2) with possibly infinite-dimensional temporal-level covariates. Our key contribution in this regard is discussed in more detail next, in light of some existing literature.

1.1   Existing literature and our contribution

In functional data analysis, models similar to (1) have been studied to some extent, where an infinite-dimensional covariate process is used to model the response variable. The readers are referred to Ferraty & Vieu (2003, 2004) for a general treatment of univariate responses, and Xiang et al. (2013), Omar & Wang (2019) for a multivariate response setup. However, all these approaches consider each datapoint to be iid without any regard to the temporal dependence. Recently, Li & Yang (2023) addressed this gap by considering a temporally dependent functional data setup, but the observations are presented at regular intervals of time and space. We relax that by allowing the dataset (2) to have irregularly spaced points in time and varying locations in space.

Nonlinear spatiotemporal regression model is far less studied compared to the extensive literature on the nonlinear analysis of time series or the literature on spatial models (Cressie & Wikle 2011). Much success of spatiotemporal modeling has instead come from a Bayesian point of view (see Haining & Li 2020, for a detailed review). Among the frequentist viewpoint through nonparametric estimation, Wang & Wang (2009) solved a related problem of mean (or trend) estimation using a local linear estimation technique. A completely nonparametric estimation was attempted by Yang & Qiu (2018) for modeling data on incidence rates of diseases. While both studies considered regularly spaced timepoints, the case for irregularly spaced timepoints was investigated by Al-Sulami et al. (2017), who proposed a semi-parametric approach. However, these studies did not allow the covariate process to be infinite-dimensional. On that note, perhaps the closest article to our approach is the work by Hong & Linton (2020), who considered infinite-dimensional covariates but only in a time series setup.

While the existing works on nonparametric modeling of spatiotemporal data and functional data analysis with infinite-order regression have grown into independent sets of literature, our proposed model (1) aims to bring a novel perspective by combining these two ideas. By bridging these, it allows us to efficiently estimate the trend function using even irregularly spaced observations in space and time, which had only limited attention in the past. Furthermore, we completely avoid any mixing-type assumption, which is a standard practice to model data dependence across space and time. In fact, Hong & Linton (2020) illustrated that, since $\boldsymbol{X}_t$ is infinite-dimensional, it is quite possible that it contains infinite order lags of some exogenous variable, say $\boldsymbol{Z}_t$ such that $\boldsymbol{X}_t = \sum_{j=0}^{\infty} a_j \boldsymbol{Z}_{t-j}$ for some scalars $a_j \in \mathbb{R}$. This is not α-mixing in general. To avoid such scenarios, they considered a near-epoch dependence (NED) condition on the covariates $\boldsymbol{X}_t$, but still require an α-mixing condition on $(Y_t, \boldsymbol{Z}_t)$ jointly. However, we show that it is possible to completely avoid any mixing-type conditions and obtain valid inference under a polynomially-decaying moment contraction (PMC) assumption. It is much weaker than the standard geometric moment contraction (GMC) condition, and encompasses various processes which are not α or β-mixing (see Chen et al. 2016, for a counter-example). Our requirement is more general than NED conditions, and can be easier to work with due to lesser restrictions on the moments (Remark 5.1, Shao & Wu 2007). As a result, in addition to allowing stronger than geometric dependence structure, we are able to derive asymptotic validity results of our proposed estimator under only a second-order stationarity condition on the covariate $\boldsymbol{X}_t$. This improves upon the requirement of the existence of $(2+\delta)$-order moment $(\delta > 0)$ for Y_t and fourth order moment condition for $\boldsymbol{X}_t$ required by the analysis of Hong & Linton (2020). In summary, the major contributions of our work are threefold:

(a) We propose a first-of-its-kind nonlinear regression model with infinite-dimensional covariates for spatiotemporally dependent data and its nonparametric estimation from irregularly sampled observations;

(b) We replace the mixing-type dependence condition with a functional dependence assumption and allow for significantly slow decay of dependence through PMC assumption.

(c) We reduce the higher-order moment assumptions, which, in turn, allows for a slightly heavier-tailed distribution of covariates.

The methodology and theoretical results are presented in Section 2 and Section 3 below. In Section 4, we demonstrate the use of the proposed methodology in the real life dataset of air pollution from India. The paper concludes with some important remarks in Section 5. Additional theoretical discussions, proofs, another real life example with soccer data, and a simulation study are deferred to the supplementary material in the interest of space.

1.2 Notations

For better understanding of the technical discussions, we collect some notations that will be used throughout the paper. The sets of real numbers, natural numbers, and integers are denoted respectively by $\mathbb{R}, \mathbb{N}, \mathbb{Z}$. For two sequences a_n and b_n, $a_n = \mathcal{O}(b_n)$ denotes that for all sufficiently large n, a_n is bounded between cb_n and Cb_n for two positive real constants $c < C < \infty$. The notation $a_n = o(b_n)$ indicates that $\limsup_{n\to\infty}(a_n/b_n) = 0$. Correspondingly, $\mathcal{O}_{\mathbb{P}}(\cdot)$

and $o_{\mathbb{P}}(\cdot)$ denote the probabilistic versions, where the analytical convergence and boundedness are respectively replaced by convergence in probability and stochastic boundedness.

2 Nonparametric estimation

2.1 Mathematical Framework

Before we discuss the proposed estimator, it is critical to define all components and conditions of model (1). We emphasize, and shall explicate properly as required later, that these conditions do not affect the generality of the spatiotemporal model to a large extent.

As the covariate process $\boldsymbol{X}_t$ is allowed to be infinite-dimensional, the abstract space of its range χ must be well-structured to allow necessary mathematical operations to be performed. We assume that χ is an affine subspace of a weighted L^2 Banach space, namely, for any $\boldsymbol{x}, \boldsymbol{x}' \in \chi$, the difference $(\boldsymbol{x} - \boldsymbol{x}') \in \chi$, and, $\chi \subseteq \boldsymbol{D}\mathcal{L}^2$, where

$$\boldsymbol{D}\mathcal{L}^2 = \left\{ (z_1, z_2, \dots) \in \mathbb{R}^\infty : \sum_{i=1}^\infty \zeta_i^{-2} z_i^2 < \infty \right\} \tag{3}$$

for some fixed choice of sequence $\{\zeta_i\}_{i=1}^\infty$. Let us use $\boldsymbol{D}$ to denote an infinite-dimensional linear operator that multiplies the ith coordinate by ζ_i for all i. Then, the above restriction allows one to formally define scaled norms for the infinite-dimensional covariates in $\mathbb{R}^\infty$, and induce a scaled L^2 metric $\left\| \boldsymbol{D}^{-1}(\boldsymbol{x} - \boldsymbol{x}') \right\|$ for any $\boldsymbol{x}, \boldsymbol{x}' \in \chi$. In addition, we consider a PMC condition to model the dependence structure of the covariate process $\{\boldsymbol{X}_t\}_{t \in \mathcal{T}}$. Specifically, there exists iid random variables $\{Z_i\}_{i \in \mathbb{Z}}$ and measurable functions $G_t : \mathbb{R}^\infty \to \chi$ satisfying

$$\boldsymbol{X}_t = G_t\left(\dots, Z_{t-1}, Z_t \right).$$

For an independent and iid copy Z_i^* for $i \in \mathbb{Z}$, assume that the quantity

$$\theta_2(m) = \sup_t \left\| \boldsymbol{D}^{-1} \left(\boldsymbol{X}_t - G_t\left(\dots, Z_{t-m-1}, Z_{t-m}^*, Z_{t-m+1}, \dots, Z_t \right) \right) \right\|_2 = \mathcal{O}(m^{-\tau}), \tag{4}$$

for some $\tau > 2$. Equation (4) controls the temporal dependence structure through the widely used notion of functional dependence measures (see Wu 2005, 2011, Wu & Zhou 2011, Karmakar & Wu 2020). The PMC condition also implies that $\Delta_2(m) = \mathcal{O}(m^{-\tau+1})$, where

$$\Delta_2(m) = \sup_t \left\| \boldsymbol{D}^{-1} \left(\boldsymbol{X}_t - G_t(\dots, Z_{t-m-2}^*, Z_{t-m-1}^*, Z_{t-m}^*, Z_{t-m+1}, \dots, Z_t) \right) \right\|.$$

This is a generalization of $\theta_2(m)$ that considers the changes in $\boldsymbol{X}_t$ due to the change in all historical values up to a time lag of m.

Next, assume that the error terms satisfy a temporal homogeneity condition

$$\text{Cov}(\epsilon_{t_1}(s_1), \epsilon_{t_2}(s_2)) = \rho(s_1, s_2) = \mathcal{O}\left(\|s_1 - s_2\|^{-(d+\delta)} \right), \tag{5}$$

for all $s_1, s_2 \in \mathcal{S}$, where $\rho(\cdot, \cdot)$ is a symmetric function satisfying $\rho(s, s) = 1$ for any $s \in \mathcal{S}$ and d is the dimension of the spatial horizon $\mathcal{S}$. $\delta > 0$ is some small positive constant.

Typically, $\rho(s_1, s_2) = \rho^*(\|s_1 - s_2\|)$ with $\rho^*(0) = 1$, and the decay rate in (5) simply restricts the ρ^* to be a polynomially decaying function of its argument. This connection immediately simplifies the conditional variance structure of the data $Y_t(s)$ by separating it into two parts: one involving temporal correlation governed by the covariate $\boldsymbol{X}_t$ and another involving spatial correlation governed by the $\rho(\cdot, \cdot)$ function, i.e., we can write $\mathrm{Cov}(Y_{t_1}(s_1), Y_{t_2}(s_2)) = \sigma(\boldsymbol{X}_{t_1}, s_1)\sigma(\boldsymbol{X}_{t_2}, s_2)\rho(s_1, s_2)$.

We have also indicated that in model (1), the mean and covariance functions are sufficiently smooth. To add some mathematical formalism, we assume that for each $\boldsymbol{x} \in \chi$, both $\mu(\boldsymbol{x}, \cdot)$ and $\sigma(\boldsymbol{x}, \cdot)$ are elements of a Hilbert space $\mathcal{H}$ of functions from $\mathcal{S}$ to $\mathbb{R}$, equipped with the inner product $\langle f, g \rangle = \int_{\mathcal{S}} f(s)g(s)ds$. The space $\mathcal{H}$ can be represented as

$$\mathcal{H} = \left\{ \sum_{i=1}^{\infty} a_i b_i(s) : a_i \in \mathbb{R}, \text{ and } \sum_{i=1}^{\infty} |a_i| < \infty \right\},$$

where $\{(b_i : \mathcal{S} \to \mathbb{R})\}_{i \in \mathbb{N}}$ is a set of countable orthonormal basis functions of $\mathcal{H}$. This suggests a representation of the mean and covariance functions in the form

$$\mu(\boldsymbol{x}, s) = \sum_{k=1}^{\infty} \mu_k(\boldsymbol{x})b_k(s), \ \sigma(\boldsymbol{x}, s) = \sum_{k=1}^{\infty} \sigma_k(\boldsymbol{x})b_k(s), \tag{6}$$

such that $\sum_{k=1}^{\infty} |\mu_k(\boldsymbol{x})|$ and $\sum_{k=1}^{\infty} |\sigma_k(\boldsymbol{x})|$ both exist and are finite for each $\boldsymbol{x} \in \chi$. Such representation has appeared before for spatiotemporal modeling (Wikle & Cressie 1999), and it is standard in the functional PCA literature due to Karhunen-Loéve expansion (Karhunen 1946, Loève 1946), where the basis functions are chosen to correspond with the eigenfunctions of covariance-operator (Yao et al. 2005). The basis functions in (6) have also been extensively used in dependent data modeling, in the form of bi-square functions (Cressie & Johannesson 2008), splines (Goldsmith et al. 2012), wavelets (Chen & Christensen 2015), etc. We refer to Cressie et al. (2022) and the references therein for a comprehensive exposition. The particular choice of the basis function often has to be paired with the choice of the hyperparameter K_ϵ as in Theorem 2, to maintain a consistent level of accuracy. In a practical setting, one may use any orthogonal set of basis functions for computational convenience, and the corresponding K_ϵ is picked out via a cross-validation approach; see Section 4 for our specific choice.

While this takes care of a smoothness restriction over the spatial locations $\mathcal{S}$, to ensure further smoothness over the set χ, we assume that the component functions $\mu_k(\boldsymbol{x})$ and $\sigma_k(\boldsymbol{x})$, for any $k \in \mathbb{N}$, are continuous on χ with respect to the scaled distance metric introduced in (3). Moreover, these conditions imposed on $\mu(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ in (6) are not usually very restrictive. Since $\mathcal{S} = [0, 1]^d$ is a Hausdorff topological space, and is compact, there exist choices of universal kernel functions such that their linear span is able to uniformly approximate any continuous function (Micchelli et al. 2006). Hence, by choosing the basis of the corresponding Hilbert space spanned by those universal kernels, it follows that the assumption of existence of such representations given in (6) becomes simply a rephrasing of a continuity-type smoothness condition for $\mu(\boldsymbol{x}, \cdot)$ and $\sigma(\boldsymbol{x}, \cdot)$.

2.2 Estimation methodology

We begin with the nonparametric estimation of the mean component for our nonlinear regression model: $Y_t(s) = \mu(\boldsymbol{X}_t, s) + \sigma(\boldsymbol{X}_t, s)\,\epsilon_t(s)$. Because the mean function has a decomposition as shown in (6), it is enough to estimate the component functions $\mu_k(\boldsymbol{x})$ for each $k \in \mathbb{N}$. In order to proceed further, we shall need two assumptions about the boundedness of the expected covariate and the error components, given as follows.

(A1) For each $s \in \mathcal{S}$, define $\tilde{\mu}(s) = \mathbb{E}(|\mu(\boldsymbol{X}_0, s)|)$ and $\tilde{\sigma}^2(s) = \mathbb{E}(\sigma^2(\boldsymbol{X}_0, s))$ where the expectations are taken over the marginal distribution of the covariate $\boldsymbol{X}_0$. Assume that both $\tilde{\mu}(s)$ and $\tilde{\sigma}^2(s)$ are continuous and are integrable on $\mathcal{S}$.

(A2) The first and second order moments of the errors $\epsilon_t(s)$ are uniformly bounded, i.e., $\sup_{t \in [0,1], s \in \mathcal{S}} \mathbb{E}(\epsilon_t^2(s)) < \infty$.

Assumption (A2) is a technical assumption that controls the behaviour of the errors and allows us to perform interchangeability of integral and expectation operators later on. It is usually very weak: for example, if the second order moment of $\epsilon_t(s)$ exists and $\mathbb{E}(\epsilon_t^2(s))$ is a continuous function of t and s, then by compactness of $\mathcal{T} = [0,1]$ and $\mathcal{S}$, the assumption (A2) follows naturally. As a consequence of Assumption (A2), we also have $\sup_{t \in [0,1], s \in \mathcal{S}} \mathbb{E}(|\epsilon_t(s)|) < \infty$. Additionally, note that since $\tilde{\mu}(s)$ and $\tilde{\sigma}^2(s)$ as in Assumption (A1) are continuous on the compact set $\mathcal{S}$, they are also uniformly bounded. For notational convenience, let us indicate M as a generic constant which serves as the uniform bound for all these quantities. Therefore, it follows by an application of Cauchy-Schwarz inequality that

$$\int_{\mathcal{S}} \mathbb{E}\left[|\sigma(\boldsymbol{X}_t, s)\epsilon_t(s)b_k(s)|\right] ds \leqslant \int_{\mathcal{S}} M \mathbb{E}\left[|\sigma(\boldsymbol{X}_t, s)b_k(s)|\right] ds \leqslant M \left(\int_{\mathcal{S}} b_k^2(s)ds\right) \left(\int_{\mathcal{S}} \mathbb{E}(\sigma^2(\boldsymbol{X}_t, s))ds\right).$$

The first term $\int_{\mathcal{S}} b_k^2(s)ds$ is equal to 1 as $b_k(s)$ is an orthonormal basis function of $\mathcal{H}$. For the second term, note that due to the second-order stationarity of $\boldsymbol{X}_t$, it is equal to $\int_{\mathcal{S}} \mathbb{E}(\sigma^2(\boldsymbol{X}_0, s))ds$, which is also bounded due to Assumption (A1).

Now, starting with the decomposition of the mean function as in (6), an application of Fubini's theorem yields that

$$\int_{\mathcal{S}} Y_t(s)b_k(s)\, ds = \int_{\mathcal{S}} \sum_{i=1}^{\infty} \mu_i(\boldsymbol{X}_t)b_i(s)b_k(s)\, ds + \int_{\mathcal{S}} \sigma(\boldsymbol{X}_t, s)\epsilon_t(s)b_k(s)\, ds.$$

Following the orthonormality of the basis functions, the above can be simplified as

$$\int_{\mathcal{S}} Y_t(s)b_k(s)ds = \mu_k(\boldsymbol{X}_t) + \eta_{tk}, \tag{7}$$

where $\eta_{tk} = \int_{\mathcal{S}} \sigma(\boldsymbol{X}_t, s)\epsilon_t(s)b_k(s)\, ds$. Subsequently, we may write

$$\mathbb{E}(\eta_{tk}) = \mathbb{E}\left(\int_{\mathcal{S}} \sigma(\boldsymbol{X}_t, s)\epsilon_t(s)b_k(s)\, ds\right) = \int_{\mathcal{S}} b_k(s)\mathbb{E}(\sigma(\boldsymbol{X}_t, s))\mathbb{E}(\epsilon_t(s))\, ds = 0,$$

where we use the independence of $\epsilon_t(s)$ and $\boldsymbol{X}_t$. Also, by another application of Fubini's

theorem (valid due to Assumption (A2)), we get

$$\text{Cov}(\eta_{t_1 k}, \eta_{t_2 k}) = \mathbb{E}(\eta_{t_1 k} \eta_{t_2 k}) = \mathbb{E}\left[\int_{\mathcal{S}} \int_{\mathcal{S}} \sigma(\boldsymbol{X}_{t_1}, s)\sigma(\boldsymbol{X}_{t_2}, s')\epsilon_{t_1}(s)\epsilon_{t_2}(s')b_k(s)b_k(s')\, ds\, ds'\right]$$

$$= \int_{\mathcal{S}^2} \mathbb{E}\left[\sigma(\boldsymbol{X}_{t_1}, s)\sigma(\boldsymbol{X}_{t_2}, s')\right]\rho(s, s')b_k(s)b_k(s')\, ds\, ds'.$$

Hence, for every $k \in \mathbb{N}$, we get a reduced temporal model

$$Y_{tk}^* = \int_{\mathcal{S}} Y_t(s)b_k(s)\, ds = \mu_k(\boldsymbol{X}_t) + \eta_{tk}, \tag{8}$$

where η_{tk} is a zero-mean second-order error process. Now, fix any $k \in \mathbb{N}$. The reduced model shown in (8) is a nonlinear regression model for dependent data, but with infinite-dimensional covariates $\boldsymbol{X}_t$. Hence, ideally, one may estimate the mean component functions $\mu_k(\boldsymbol{x})$ using local constant estimators, by suitably replacing the aggregated response Y_{tk}^* (which is unobserved) by an estimate based on the dataset $\mathcal{D}$ as detailed below.

It is worthwhile to note that, although the errors $\epsilon_t(s)$ are independent of the covariates $\boldsymbol{X}_t$, that is not the case for the aggregated error η_{tk}, as its variance depends on $\boldsymbol{X}_t$ through the $\sigma(\cdot, \cdot)$ function. However, it still follows that

$$\mathbb{E}\left(\eta_{tk} \mid \{\boldsymbol{X}_{t'} : t' \leqslant t\}\right) = \int_{\mathcal{S}} b_k(s)\sigma(\boldsymbol{X}_t, s)\mathbb{E}\left(\epsilon_t(s) \mid \{\boldsymbol{X}_{t'} : t' \leqslant t\}\right) ds = 0,$$

for any $t \in \mathcal{T}$. As we shall show below, this conditional mean-zero property is usually enough to ensure desirable theoretical properties of the proposed nonparametric estimators. We next focus on estimating each coefficient $\mu_k(\cdot)$ in (8) using standard nonparametric methods. For that, a kernel function of the following form is introduced:

$$\mathcal{K}^*\left(\boldsymbol{X}_t, \boldsymbol{x}\right) = \mathcal{K}\left(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\right) = K\left(\|\boldsymbol{V}^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\|/h_n\right),$$

where $\boldsymbol{H}_n$ is a bandwidth matrix of the form $h_n \boldsymbol{D}$, $\boldsymbol{D}$ being a diagonal matrix in line with the linear operator defined in (3). Here, we also make use the scaled distance function as in (3) to ensure that the above expressions are well-defined. Therefore, it is possible to use a univariate kernel function $K(\cdot)$ that acts upon the norm of the scaled distance vector scaled by an appropriate choice of the bandwidth h_n. The specific conditions on the choice of the kernel function $K(\cdot)$ are discussed in Section 2.3.

We can now provide an estimator of $\mu_k(x)$ as

$$\widehat{\mu}_k(x) = \frac{\sum_{i=1}^n \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{X}_{t_i}))\widehat{Y}_{t_i k}^*}{\sum_{i=1}^n \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{X}_{t_i}))} = \frac{\sum_{i=1}^n K\left(\left\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{X}_{t_i})\right\|\right)\widehat{Y}_{t_i k}^*}{\sum_{i=1}^n K\left(\left\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{X}_{t_i})\right\|\right)} \tag{9}$$

where $\widehat{Y}_{t_i k}^*$ is an estimate of $Y_{t_i k}^*$ based on dataset $\mathcal{D}$. Although the usual Monte Carlo method can be used to estimate the aggregated response as $Y_{t_i k}^* \approx \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{t_i}(s_{t_i j})b_k(s_{t_i j})$, it has one serious drawback. If we assume that the choice of the spatial points $\{s_{t_i 1}, \ldots, s_{t_i n_i}\}$ are deterministic, then it is possible that they may concentrate on one region of space for any particular time period. The simple Monte Carlo procedure in this case will provide a biased

estimate. On the other hand, if we assume that the choices of spatial points are uniformly at random, then the covariance of the Y-values in these locations does not decay sufficiently fast with the distance between two spatial points. Due to the presence of this covariance, the Monte Carlo estimate may end up having infinite variance, thus leading to non-consistency of the estimator. Gathering insights from the functional data analysis, we subdivide the spatial horizon $\mathcal{S}$ into small grids, and choose a representative point on each hypercube of the grid. Then, the Monte-Carlo procedure is applied for this reduced set of representative points instead of all the points. This ensures a uniform distribution of the considered spatial points and also allow us to control the covariance between representative points of two distant grids. The final algorithm to obtain $\widehat{Y}^*_{t_i k}$ is illustrated in Algorithm 1. A correctness guarantee that this procedure produces a consistent estimate of $Y^*_{t_i k}$ is given in Proposition 2.

Algorithm 1: Modified Monte-Carlo procedure for estimating $Y^*_{t_i,k}$

Input: $t_i \in \mathcal{T}$, $\{s_{t_{ij}}\}_{j=1}^{n_i} \subseteq \mathcal{S}$, Basis index $k \in \mathbb{Z}^+$ and response $\{Y_{t_i}(s_{t_{ij}})\}_{j=1}^{n_i}$

1 Let $\epsilon^*_{t_i} \leftarrow \inf \left\{\epsilon > 0, \text{ and } \mathcal{S} \subseteq \cup_{j=1}^{n_i} B(s_{t_{ij}}, \epsilon)\right\}$, the "effective spatial resolution".

2 Let $H_1, H_2, \ldots, H_{r_i}$ be hypercubes from the grid of $[0,1]^d$ with diameters equal to $\epsilon^*_{t_i}$

3 Let $s^c_l \leftarrow \text{Center}(H_l)$ for each $l = 1, \ldots, r_i = (\epsilon^*_{t_i}/\sqrt{d})^{-d}$.

4 Let $s_{t_i,j_l} \leftarrow \arg\min_j \|s^c_l - s_{t_{ij}}\|$ for each $l = 1, 2, \ldots, r_i$.

5 Obtain the final estimate as

$$\widehat{Y}^*_{t_i k} = \frac{1}{r_i} \sum_{l=1}^{r_i} Y_{t_i}(s_{t_i j_l}) b_k(s_{t_i j_l}) \tag{10}$$

2.3 Choice of the Kernel and the Bandwidth

Before diving into technical results on the aforementioned estimation procedure, it is meaningful to state the underlying assumptions concerning the choice of the kernel and the bandwidth parameters. We restrict our choice to the type-I kernels as motivated by Hong & Linton (2020). The kernel function $K : \mathbb{R}^+ \to \mathbb{R}^+$ is bounded and has bounded support, i.e.,

$$\int K(u)du = 1, \; C_1 \mathbf{1}_{[0,\lambda]}(u) \leqslant K(u) \leqslant C_2 \mathbf{1}_{[0,\lambda]}(u), \tag{11}$$

for some positive finite constants C_1, C_2 and λ. The uniform kernel is one such example of a type-I kernel. Due to this, we can now represent the small ball probability using

$$\phi_{\boldsymbol{x}}(h_n u) = \mathbb{P}\left(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\| \leqslant u\right) = P(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\| \leqslant h_n u) = P(\boldsymbol{X}_t \in B(\boldsymbol{x}, h_n u)), \tag{12}$$

where $B(\boldsymbol{x}, h_n u)$ is the infinite-dimensional hyper ellipsoid centered at $\boldsymbol{x} \in \chi \subseteq \mathbb{R}^\infty$ with different axis lengths $2h_n \eta_j u$ at the jth direction. Note that, due to the stationarity of $\boldsymbol{X}_t$, the quantity $\phi_{\boldsymbol{x}}(h_n \lambda)$ is free of the time index t. Furthermore, regarding the choice of the bandwidth, we make the following assumption.

(A3) The sequence of bandwidth parameter h_n satisfies $h_n \to 0$, $\phi_{\boldsymbol{x}}(h_n u) \to 0$, and $n h_n^2 \phi_{\boldsymbol{x}}^2(h_n u) \to \infty$ as $n \to \infty$ for any fixed $u > 0$ where $\phi_{\boldsymbol{x}}(h_n u)$ is the small-ball probability defined in (12). Additionally, the joint small-ball probability defined below in (13) satisfies $\phi_{\boldsymbol{x},|t-t'|}(h_n u, h_n u) = \mathcal{O}(\phi_{\boldsymbol{x}}^2(h_n u))$ for any $t \neq s$.

Assumption (A3) ensures that while the size of the neighborhood and the small-ball probability $\phi_{\boldsymbol{x}}(h_n u)$ diminishes asymptotically, the number of observations in this neighborhood $B(\boldsymbol{x}, h_n u)$ increases at a rate more than $n^{1/2} h_n^{-1}$. In the second part of Assumption (A3), we consider the joint small-ball probability for any $t \neq t'$,

$$\phi_{\boldsymbol{x},|t-t'|}(h_n u, h_n v) = \mathbb{P}\left(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\| \leqslant u, \|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x})\| \leqslant v\right), \tag{13}$$

which again, by the second-order stationarity of $\boldsymbol{X}_t$, depends only on the difference in time indices $|t - t'|$. By bounding this joint small-ball probability, Assumption (A3) ensures only a localized dependence in the joint distribution of $(\boldsymbol{X}_t, \boldsymbol{X}_{t'})$ for $t \neq t'$. This assumption is the same as Assumption B.5 of Hong & Linton (2020), which, as we show below, helps to control the cross-product moments of the kernel functions. An obvious consequence of Assumption (A3) is that as $n \to \infty$,

$$n\phi_{\boldsymbol{x}}(h_n u) = \sqrt{n h_n^2 \phi_{\boldsymbol{x}}^2(h_n u)} \times \frac{\sqrt{n}}{h_n} \to \infty. \tag{14}$$

The condition (A3) along with type-I kernel is key to obtaining various insights on the asymptotic nature of the kernel function. Starting with the expected value of the kernel function, note that $E(\mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x}))) = \mathbb{E}(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\|)) = \int K(u)\, d\phi_{\boldsymbol{x}}(h_n u)$, which we denote as $\psi_{\boldsymbol{x},1}(h_n)$ below. It is also easy to see that $\mathbb{E}\left(\mathcal{K}^2(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x}))\right) = \int K^2(u)\, d\phi_{\boldsymbol{x}}(h_n u)$, denoted as $\psi_{\boldsymbol{x},2}(h_n, 0)$. Due to the boundedness of the kernel function, the quantity $\psi_{\boldsymbol{x},1}(h_n)$ decays exactly at the order of $\phi_{\boldsymbol{x}}^j(h_n \lambda)$ for $j = 1, 2$. Therefore, we have,

$$\frac{1}{\phi_{\boldsymbol{x}}(h_n \lambda)} \mathbb{E}\left[\mathcal{K}^j(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x}))\right] \to \xi_j, \tag{15}$$

as $n \to \infty$ for some $\xi_1 > 0, \xi_2 > 0$ (cf. Corollary 1 of Hong & Linton (2020)).

Moving on to the joint distributions, we consider the cross moment of kernel functions

$$\mathbb{E}\left(\mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x}))\mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x}))\right) = \mathbb{E}\left(K\left(\left\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\right\|\right) K\left(\left\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x})\right\|\right)\right)$$

$$= \int K(u)K(v)d\phi_{\boldsymbol{x},|t-t'|}(h_n u, h_n v) = \psi_{\boldsymbol{x},2}(h_n, |t - t'|) \leqslant C_2^2 C \phi_{\boldsymbol{x}}^2(h_n \lambda) \to 0, \tag{16}$$

as $n \to \infty$ due to Assumption (A3). Here, $C > 0$ is some arbitrary constant free of $\boldsymbol{x}$. While the above derivation provides a uniform rate of decay for the cross-order moments of the kernel functions, we expect the decay rate to be $\xi_1^2 \phi_{\boldsymbol{x}}^2(h_n \lambda) + g(|t - t'|)$ for some function $g(\cdot)$ that controls the covariance between the kernel function evaluated at $(\boldsymbol{X}_t - x)$ and $(\boldsymbol{X}_{t'} - \boldsymbol{x})$ as a function of the lag $|t - t'|$. In view of the PMC dependence structure as in (4), it turns out that a quantitative upper bound to a Martingale difference of any measurable Lipschitz function of $\boldsymbol{X}_t$ is possible, as illustrated in the following result.

Proposition 1. *Let $f : \chi \to \mathbb{R}$ be a function such that it is measurable and Lipschitz (under the scaled norm given in (3)) with a constant C, i.e., $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leqslant C\|\boldsymbol{D}^{-1}(\boldsymbol{x} - \boldsymbol{y})\|$ for any $\boldsymbol{x}, \boldsymbol{y} \in \chi$. Then, with $\mathcal{F}_t$ denoting the σ-field generated by $\{\boldsymbol{X}_{t'} : t' \leqslant t\}$,*

$$|\mathbb{E}\left(f(\boldsymbol{X}_{t_i}) \mid \mathcal{F}_{t_i - m}\right) - \mathbb{E}(f(\boldsymbol{X}_{t_i}))| \leqslant C\Delta_2(m)/\sqrt{2}. \tag{17}$$

We note that, to make use of Proposition 1, we need to apply Lipschitz-type smoothness restrictions on the kernel function. This is formalized below.

(A4) For each $\boldsymbol{x} \in \chi$, the function $f_{\boldsymbol{x}} : \chi \to \mathbb{R}$ given by $f_{\boldsymbol{x}}(\boldsymbol{y}) = \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{y}))\mu_k(\boldsymbol{y})$ is Lipschitz over χ in the scaled L^2-metric given in (3), where μ_k is the component functions of μ as given in (6). The Lipschitz constant is also uniform over $\boldsymbol{x} \in \chi$. The same holds for the kernel-weighted variance components, i.e., for $g_{\boldsymbol{x}}(\boldsymbol{y}) = \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{y}))\sigma_k(\boldsymbol{y})$ for any fixed $k \in \mathbb{Z}^+$ and uniformly over any $\boldsymbol{x} \in \chi$.

This results in an improved bound on the autocovariance between the kernel functions, as illustrated in the following corollary, whose proof is provided in the supplementary material.

Corollary 1. *For sufficiently large n, for some constant $C > 0$ and C_2 as in (11), we have*

$$Cov\left(\mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})), \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x}))\right) \leqslant \frac{\sqrt{2}CC_2^2}{9h_n^2\lambda^2}\Delta_2(|t - t'|)(1 - \mathbb{E}\left(\phi_{\boldsymbol{X}_0}(3h_n\lambda)\right))$$
$$+ C_2^2(1 + C)\phi_{\boldsymbol{X}}^2(h_n\lambda)\mathbb{E}\left(\phi_{\boldsymbol{X}_0}(3h_n\lambda)\right). \quad (18)$$

A related inequality can be found in inequality (34) of Xiao & Wu (2011). We shall make use of (15) and (18) repeatedly while deriving the asymptotic properties of the nonparametric estimators of $\mu(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$.

3 Asymptotic properties

3.1 Consistency of the estimator of mean

The first step in establishing various asymptotic properties of the proposed estimator would be to ensure that it is consistent. In this subsection, we derive two key theorems related to the consistency of the mean estimator. However, before proceeding with them, we present a short result that provides correctness guarantees for the Modified Monte-Carlo procedure illustrated in Algorithm 1 and provides a quantitative bound on the error.

Proposition 2. *Fix any $t_i \in \mathcal{T}, k \in \mathbb{Z}^+$. Under assumptions (A1), we have*

$$\mathbb{E}\left|\widehat{Y}_{t_ik}^* - Y_{t_ik}^*\right| = \mathcal{O}\left(\max\left\{(\epsilon_{t_i}^*)^d, \sup_{\|s-s'\|\leqslant\epsilon_{t_i}^*}|b_k(s) - b_k(s')|, \sup_{\|s-s'\|\leqslant\epsilon_{t_i}^*}|\widetilde{\mu}(s) - \widetilde{\mu}(s')|\right\}\right), \quad (19)$$

where $\epsilon_{t_i}^*$ and $\widehat{Y}_{t_ik}^*$ is the estimated spatially-aggregated response $Y_{t_ik}^*$ using the modified Monte-Carlo procedure (see (10)), and $\widetilde{\mu}(s)$ is as defined in Assumption (A1). Let us denote the right-hand side of (19) as $\delta_k(\epsilon_{t_i}^*)$.

It is clear that, in addition to Assumption (A1), if we consider $\epsilon_{t_i}^* \to 0$ (which is a proxy for the average gap of data-points in the spatial horizon at timepoint t_i) in the asymptotic regime, then Proposition 2 establishes L^1-type convergence for the modified Monte-Carlo estimate $\widehat{Y}_{t_ik}^*$, from which the consistency of the Algorithm 1 readily follows.

Moving on, the first key result presented in this section, establishes the consistency of the estimate $\widehat{\mu}_k(\boldsymbol{x})$ of each component of the mean function to the corresponding component $\mu_k(\boldsymbol{x})$ as in (6), pointwise at each $\boldsymbol{x} \in \chi$.

Theorem 1. *Along with Assumptions (A1)-(A4), assume that the effective spatial resolution $\epsilon_{t_i}^*$ as defined in Algorithm 1 decays to 0 uniformly over $t_i \in \mathcal{T}$. Then, the estimate of the mean components, i.e., $\widehat{\mu}_k(x)$ as in (9) is pointwise consistent for $\mu_k(x)$ as $n \to \infty$. Mathematically, it means that for any fixed $k \in \mathbb{Z}^+$ and any $\boldsymbol{x} \in \chi$, $\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}) = o_{\mathbb{P}}(1)$.*

While Theorem 1 establishes the pointwise consistency of $\widehat{\mu}_k(\boldsymbol{x})$ for the component functions $\mu_k(\boldsymbol{x})$, it has limited applicability in practice. Since the unknown mean function $\mu(\boldsymbol{x}, s)$ may consist of infinitely many component functions as illustrated in (6), with a finite amount of computational power, we can estimate it by a truncated series consisting of a large, but finitely many terms. This motivates the following theorem.

Theorem 2. *Suppose that the assumptions of Theorem 1 hold. In addition, assume that there exists a function $b_\infty : \mathcal{S} \to (0, \infty)$ such that $|b_k(s)| \leqslant b_\infty(s)$ for all $s \in \mathcal{S}$. Then for any fixed $x \in \chi$, $s \in \mathcal{S}$ and $\epsilon > 0$, there exists sufficiently large K_ϵ such that*

$$\left(\sum_{k=1}^{K_\epsilon} \widehat{\mu}_k(\boldsymbol{x}) b_k(s) - \mu(\boldsymbol{x}, s) \right) = \mathcal{O}_{\mathbb{P}}(\epsilon),$$

as $n \to \infty$. Moreover, if $b_\infty(s)$ is continuous, then as the sample size $n \to \infty$,

$$\sup_{s \in \mathcal{S}} \left| \sum_{k=1}^{K_\epsilon} \widehat{\mu}_k(\boldsymbol{x}) b_k(s) - \mu(\boldsymbol{x}, s) \right| = \mathcal{O}_{\mathbb{P}}(\epsilon).$$

3.2 Asymptotic distribution of the estimate of mean

A natural next step is to establish a non-degenerate asymptotic limit of the estimator, after proper normalization. As before, fix $k \in \mathbb{Z}^+$. Then, for any $\boldsymbol{x} \in \chi$, we aim to understand the asymptotics for $v_{nk}(\boldsymbol{x})^{-1/2}(\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}) - \tilde{b}_{nk}(\boldsymbol{x}))$, where $\tilde{b}_{nk}(\boldsymbol{x})$ is a bias component and $v_{nk}(\boldsymbol{x})$ is the suitable normalization factor to obtain a non-degenerate limit, which turns out to be equal to $(n\phi_{\boldsymbol{x}}(h_n\lambda))^{-1}$. However, before describing the result in detail, we present an additional assumption that strengthens the bandwidth restrictions of Assumption (A3) by connecting it with both the spatial and temporal dependency patterns.

(A5) Let τ be the exponent in PMC condition (4) and $\delta_k(\epsilon_{t_i}^*)$ be as given in Proposition 2. Assume that there exists $0 < (\tau-1)^{-1} < \beta < \alpha < 1$ such that the sequence of bandwidth parameter h_n satisfies $n^{1-2\alpha+2\beta}\phi_{\boldsymbol{x}}(h_n\lambda) \to 0$ and $\sqrt{n}\phi_{\boldsymbol{x}}^{-1/2}(h_n\lambda)(\sup_{t_i} \delta_k(\epsilon_{t_i}^*)) \to 0$.

Assumption (A5) can be thought of as an extension of Assumption (A3) that determines the precise rate at which the small-ball probability needs to decay. For example, if $\chi = \mathbb{R}^d$ and $h_n = n^{-d/2}$, then the small-ball probability decays at the rate of $\mathcal{O}(h_n^d) = \mathcal{O}(n^{-1/2})$. If $\tau = 3$, one may choose $\beta = x + 1/2$ and $\alpha = x + 5/6$ for some $x \in (0, 1/6)$. In this case, $\phi_{\boldsymbol{x}}(h_n\lambda)n^{1-2(\alpha-\beta)} = \mathcal{O}(n^{-1/6})$ which decays to 0, satisfying the first part of Assumption (A5). As $n \to \infty$ and one has dense samples over the unit square, $\epsilon_{t_i}^* \to 0$, which along with

continuity of mean and basis functions also imply $\delta_k(\epsilon_{t_i}^*) \to 0$. With a precise interplay between the spatial resolution of the sampling and the small-ball probability, the second part of Assumption (A5) can be satisfied. The final result on the asymptotic normality of the suitably centered and scaled estimator is obtained below.

Theorem 3. *Let ξ_1 and ξ_2 be as defined in* (15). *Under Assumptions (A1)-(A5), as $n \to \infty$,*

$$\frac{\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}\xi_1}{\sqrt{\xi_2\sigma_{kk}(\boldsymbol{x})}} \left(\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}) - \tilde{b}_{nk}(\boldsymbol{x})\right), \tag{20}$$

is asymptotically distributed as a standard normal random variable, where

$$\sigma_{kl}(\boldsymbol{x}) = \sum_{u=1}^{\infty} \sum_{v=1}^{\infty} c_{u,v}(k,l)\sigma_u(\boldsymbol{x})\sigma_v(\boldsymbol{x}), \quad \text{for any } k, l = 1, 2, \ldots, \tag{21}$$

$$c_{u,v}(k,l) = \int_{\mathcal{S}^2} \rho(s,s')b_k(s)b_u(s)b_l(s')b_v(s')dsds'. \tag{22}$$

and $\sigma_u(\boldsymbol{x})$s are the component functions of $\sigma(\boldsymbol{x},s)$ as in (6). *The bias $\tilde{b}_{nk}(\boldsymbol{x})$ is given by*

$$\tilde{b}_{nk}(\boldsymbol{x}) = \frac{1}{n\mathbb{E}(K_0)} \sum_{i=1}^{n} \mathbb{E}(K_{t_i}\mu_k(\boldsymbol{X}_{t_i})) - \mu_k(\boldsymbol{x}). \tag{23}$$

It is clear from Assumption (A4) that $\tilde{b}_{nk}(\boldsymbol{x}) = O(h_n)$, where h_n is the bandwidth. While $h_n \to 0$ asymptotically, justifying the pointwise consistency of the mean estimator, in many practical purposes, we may need a bias-correction procedure to ensure a lower-order bias term. For this, one may consider a jackknife-type procedure and obtain a revised estimator as

$$\widehat{\mu}_k^{*,(h_n)}(\boldsymbol{x}) = 2\widehat{\mu}_k^{(h_n)}(\boldsymbol{x}) - \widehat{\mu}_k^{(2h_n)}(\boldsymbol{x}), \tag{24}$$

where $\widehat{\mu}_k^{(h_n)}(\boldsymbol{x})$ is the estimator of the kth mean component with bandwidth h_n. It is now easy to see that the bias of the corrected estimate $\widehat{\mu}_k^{*,(h_n)}(\boldsymbol{x})$ is $O(h_n^2)$. Another point to note is that the quantity $\phi_{\boldsymbol{x}}(h_n\lambda)$ appearing in normalizing constant is usually unknown, but one may estimate the entire normalizing constant $n^{1/2}\phi_{\boldsymbol{x}}^{1/2}(h_n\lambda)$ by a plug-in estimate of the square root of the number of points lying in the support of the kernel function $K(\cdot, \boldsymbol{x})$ centered at $\boldsymbol{x} \in \chi$. Finally, if we were to assume a Karhunen-Loéve-type expansion for the Y-value at a location s, when covariate is $\boldsymbol{x}$, as $Y(\boldsymbol{x}) = \sum_{k=1}^{\infty} a_k(\boldsymbol{x})b_k(s)$, then the quantity $\sigma_{kl}(\boldsymbol{x})$ can be viewed as a covariance operator acting on these random coefficients $a_k(\boldsymbol{x})$.

The proof of Theorem 3 follows by considering a decomposition of the normalized estimated given in (20) as in Masry (2005) and then follows a big-block small-block decomposition to establish the result. While Masry (2005) makes use of the α-mixing property to establish asymptotic independence between the big blocks to apply central limit theorems, we achieve this through applying the PMC condition on $\boldsymbol{X}_t$ as in Assumption 4. The technical details of the proof are outlined in the supplementary material. One may also establish a more general result that considers the estimation of multiple components of the mean function simultaneously. We present this as the following corollary.

Corollary 2. *Under same assumptions as in Theorem 3, for any finite set of distinct indices $k_1, k_2, \ldots, k_m \in \mathbb{Z}^+$ and any fixed $\boldsymbol{x} \in \chi$, as $n \to \infty$,*

$$\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)} \begin{bmatrix} \widehat{\mu}_{k_1}(\boldsymbol{x}) - \mu_{k_1}(\boldsymbol{x}) - \tilde{b}_{nk_1}(\boldsymbol{x}) \\ \vdots \\ \widehat{\mu}_{k_m}(\boldsymbol{x}) - \mu_{k_m}(\boldsymbol{x}) - \tilde{b}_{nk_m}(\boldsymbol{x}) \end{bmatrix} \xrightarrow{d} N \left(\boldsymbol{0}, \frac{\xi_2}{\xi_1^2} \begin{bmatrix} \sigma_{k_1 k_1}(\boldsymbol{x}) & \ldots & \sigma_{k_1 k_m}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k_1 k_m}(\boldsymbol{x}) & \ldots & \sigma_{k_m k_m}(\boldsymbol{x}) \end{bmatrix} \right).$$

(25)

Additionally, under the setting of Theorem 2, for any $\epsilon > 0$ there exists a sufficiently large K_ϵ such that, for a standard normal random variable Z,

$$\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)} \frac{\sum_{k=1}^{K_\epsilon} \left(\widehat{\mu}_k(\boldsymbol{x}) - \tilde{b}_{nk}(\boldsymbol{x}) \right) b_k(s) - \mu(\boldsymbol{x}, s)}{\sqrt{\frac{\xi_2^2}{\xi_1^2} \sum_{k=1}^{K_\epsilon} \sum_{l=1}^{K_\epsilon} \sigma_{kl}(\boldsymbol{x}) b_k(s) b_l(s)}} = Z + \mathcal{O}_{\mathbb{P}}(\epsilon).$$

(26)

The proof of this is analogous to the proof of Theorem 3 and hence is omitted for brevity.

3.3 Simultaneous Confidence Interval Estimation

Extending on the point estimators for the mean component, we can obtain an interval estimator to quantify the uncertainty of the estimation. From a direct usage of the asymptotic normality present in (26), one can obtain an asymptotic confidence interval for $\mu_k(\boldsymbol{x})$, provided a valid consistent estimator is available for $\sigma_{kk}(\boldsymbol{x})$. To this end, consider the covariance between $\eta_{t_1 k}$ and $\eta_{t_2 l}$ for any $k, l \in \mathbb{Z}^+$ as given in (8).

$$\begin{aligned} \text{Cov}(\eta_{t_1 k}, \eta_{t_2 l}) &= \int_{\mathcal{S}^2} \mathbb{E}\left[\sigma(\boldsymbol{X}_{t_1}, s) \sigma(\boldsymbol{X}_{t_2}, s') \right] \rho(s, s') b_k(s) b_l(s') ds ds' \\ &= \sum_{u,v} \mathbb{E}\left[\sigma_u(\boldsymbol{X}_{t_1}) \sigma_v(\boldsymbol{X}_{t_2}) \right] \int_{\mathcal{S}^2} \rho(s, s') b_k(s) b_u(s) b_l(s') b_v(s') ds ds' \\ &= \sum_{u,v} c_{u,v}(k, l) \mathbb{E}\left[\sigma_u(\boldsymbol{X}_0) \sigma_v(\boldsymbol{X}_{|t_1 - t_2|}) \right], \end{aligned}$$

where the last line follows from stationarity of $\boldsymbol{X}_t$ and $c_{u,v}(k, l)$ is as given in (22). Here, we also apply Fubini's theorem to exchange the expectation and the integral operator, which can be easily validated due to the uniform boundedness of the basis functions and $\tilde{\sigma}(s)$, as given in Assumption (A1). By choosing $t_1 = t_2$, we get $\text{Cov}(\eta_{t_1 k}, \eta_{t_1 l}) = \sum_{u,v} c_{u,v}(k, l) \mathbb{E}\left[\sigma_u(\boldsymbol{X}_0) \sigma_v(\boldsymbol{X}_0) \right]$, or more precisely, the conditional covariance will be

$$\text{Cov}(\eta_{t_i k}, \eta_{t_i l} \mid \mathcal{F}_{t_i}) = \sum_{u,v} c_{u,v}(k, l) \sigma_u(\boldsymbol{X}_{t_i}) \sigma_v(\boldsymbol{X}_{t_i}).$$

(27)

Connecting the above expression, namely (27), with its population counterpart $\sigma_{k,l}(\boldsymbol{x})$ as in (21), a natural nonparametric estimate of $\sigma_{kl}(\boldsymbol{x})$ is given by

$$\widehat{\sigma}_{kl}(\boldsymbol{x}) = \frac{\sum_{i=1}^n K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{X}_{t_i})\|)(\widehat{Y}_{t_i k}^* - \widehat{\mu}_k(\boldsymbol{x}))(\widehat{Y}_{t_i l}^* - \widehat{\mu}_l(\boldsymbol{x}))}{\sum_{i=1}^n K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{X}_{t_i})\|)}.$$

(28)

As established in the following theorem, it provides a reasonably valid estimator for $\sigma_{kl}(\boldsymbol{x})$.

Theorem 4. *Under the same set of assumptions as in Theorem 1, for any fixed $k, l \in \mathbb{Z}^+$ and a fixed $\boldsymbol{x} \in \chi$, we have $\widehat{\sigma}_{kl}(\boldsymbol{x}) - \sigma_{kl}(\boldsymbol{x}) = o_{\mathbb{P}}(1)$ as $n \to \infty$.*

Guided by Theorems 3 and 4, for any fixed $k \in \mathbb{Z}^+$ and $\boldsymbol{x} \in \chi$, we can now construct a Wald-style $100(1-\alpha)\%$ asymptotic confidence interval for $\mu_k(\boldsymbol{x})$ as illustrated in the following Corollary.

Corollary 3. *Under the assumptions of Theorems 3 and 4, for any fixed $k \in \mathbb{Z}^+$ and $\boldsymbol{x} \in \chi$,*

$$\lim_{n \to \infty} \mathbb{P}\left(\left| \widehat{\mu}_k(\boldsymbol{x}) - \tilde{b}_{nk}(\boldsymbol{x}) - \mu_k(\boldsymbol{x}) \right| \leq \frac{z_{1-\alpha/2}\sqrt{\xi_2 \widehat{\sigma}_{k,k}(\boldsymbol{x})}}{\xi_1 \sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}} \right) = (1-\alpha),$$

$z_{1-\alpha/2}$ is the $(1-\alpha/2)^{th}$ quantile of the standard normal distribution, and the estimates $\widehat{\mu}_k(\boldsymbol{x})$ and $\widehat{\sigma}_{kk}(\boldsymbol{x})$ are as given by (9) and (28) respectively. The same conclusion holds if one considers a bias-corrected version $\widehat{\mu}_k^{*,(h_n)}(\boldsymbol{x})$ as in (24).

However, to produce a confidence interval for the entire mean function $\mu(\boldsymbol{x}, s)$, not just the individual components, we need to additionally consider the existence of a dominating function $b_\infty(s)$ on the basis functions, as in Theorem 2. Then, starting with any $\epsilon > 0$, we obtain a sufficiently large $K_\epsilon \in \mathbb{Z}^+$ as guaranteed by Theorem 2, which by means of (26) produces a $100(1-\alpha)\%$ confidence interval for $\mu(\boldsymbol{x}, s)$ with the endpoints given by

$$\widehat{\mu}_{1:K_\epsilon}^*(\boldsymbol{x}, s) \pm \left(\frac{z_{1-\alpha/2}\sqrt{\xi_2 Q_{K_\epsilon}(\boldsymbol{x}, s)}}{\xi_1 \sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}} + \epsilon b_\infty(s) \right),$$

where $b_\infty(\cdot)$ is as given in Theorem 2 and

$$\widehat{\mu}_{1:K}^*(\boldsymbol{x}, s) = \sum_{k=1}^K \widehat{\mu}_k^{*,(h_n)}(\boldsymbol{x}) b_k(s), \ \boldsymbol{b}_{1:K}(s) = (b_1(s), \ldots, b_K(s))^\intercal,$$

$$\widehat{\boldsymbol{\Sigma}}_{1:K,1:K}(\boldsymbol{x}) = ((\widehat{\sigma}_{kl}(\boldsymbol{x})))_{k,l=1}^K, \ Q_K(\boldsymbol{x}, s) = \boldsymbol{b}_{1:K}^\intercal(s) \widehat{\boldsymbol{\Sigma}}_{1:K,1:K}(\boldsymbol{x}) \boldsymbol{b}_{1:K}(s).$$

It is possible to extend the above pointwise confidence interval for each fixed $\boldsymbol{x} \in \chi$ to a simultaneous confidence interval over a dense subset χ_n of χ given by $\{\boldsymbol{x} : \boldsymbol{x} \in \chi, \text{ and } B(\boldsymbol{x}, 2h_n\lambda) \cap \chi = \phi\}$, where $B(\boldsymbol{x}, 2h_n\lambda)$ denotes the infinite-dimensional ball centered at $\boldsymbol{x}$ with radius $2h_n\lambda$ with respect to the scaled distance metric. This means, for any $\boldsymbol{x} \neq \boldsymbol{x}' \in \chi_n$, the scaled norm $\|\boldsymbol{D}^{-1}(\boldsymbol{x} - \boldsymbol{x}')\|_2 > 2h_n\lambda$. The following theorem establishes this uniform probabilistic bound over all $x \in \chi_n$ but for a fixed $s \in \mathcal{S}$.

Theorem 5. *Fix any $s \in \mathcal{S}$. Suppose that the Assumptions (A1)-(A5) hold. Let χ_n be a countable, bounded subset of χ such that for any $\boldsymbol{x}_0 \in \chi_n$, there is a local neighborhood $B(\boldsymbol{x}_0, \delta)$ for which the infinite series $\sum_{k=1}^\infty \mu_k(\boldsymbol{x})$ converges uniformly over all $\boldsymbol{x} \in B(\boldsymbol{x}_0, \delta)$ with δ independent of n. In addition, assume that there is a dominating function $b_\infty(s)$ such*

that $|b_k(s)| \leqslant b_\infty(s)$ for this $s \in \mathcal{S}$. Then for any $\epsilon > 0$, there exists $K_\epsilon \in \mathbb{Z}^+$ such that

$$\lim_{n \to \infty} \mathbb{P}\left(\sup_{\boldsymbol{x} \in \chi_n} \frac{\xi_1\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}}{\sqrt{\xi_2 Q_{K_\epsilon}(\boldsymbol{x}, s)}} \left| \left|\hat{\mu}^*_{1:K_\epsilon}(\boldsymbol{x}, s) - \mu(\boldsymbol{x}, s)\right| - \epsilon b_\infty(s)\right| < B_{m_n}(z)\right) \geqslant e^{-2e^{-z}} \qquad (29)$$

for any fixed $z > 0$ and $s \in \mathcal{S}$. Here,

$$B_{m_n}(z) = \sqrt{2\log(m_n)} - \frac{1}{\sqrt{2\log(m_n)}}\left[\frac{1}{2}\log(\log(m_n)) + \log(2\sqrt{\pi})\right] + \frac{z}{\sqrt{2\log(m_n)}}, \ m_n \geqslant 2,$$

and $m_n = |\chi_n|$, the number of elements in the set χ_n.

The size of the confidence set is controlled by m_n, which depends on the properties of the covariate domain χ. For example, if χ is a subset of Euclidean vector space, then m_n can be approximated by typical covering number metrics, while if χ is a smooth-function class over a field $\mathfrak{F}$, m_n can be approximated by the covering number of $\mathfrak{F}$ multiplied by appropriate functions of VC-dimension or Dudley's entropy integral for the function class. We find it imperative to point out that such a Gumbel distribution-based simultaneous confidence interval arises in many situations: see Wu & Zhao (2007), Zhou & Wu (2010), Ma et al. (2012), Deb et al. (2024) for some useful applications.

4 Real-life example: Analyzing air pollution in Delhi

We consider a dataset on air pollution in India, extracted from the government website of the Central Pollution Control Board (CPCB: https://cpcb.nic.in/). The dataset consists of measurements of various pollutants (e.g. PM2.5, PM10, NO_2, etc.), collected from stations irregularly distributed across different regions of India, with many not actively monitored at the moment. Keeping in view the continuously worsening pollution situation in Delhi (Dutta & Jinsart 2022), we focus on the 38 stations in the national capital region for our analysis. Figure 3 shows the active status and locations of these measurement stations, most of which are concentrated around the union territory of Delhi and nearby regions.

The pollutant measurements are collected at an hourly rate between January, 2015 to June, 2020; though these ranges differ from one station to another. Empirically, all pollutant measurements display a positively skewed pattern. Thus, we apply the log-transformation to all of them for the main analysis, and use log(PM2.5) as our main response variable $Y_t(s)$, for station s and timepoint t. In general, we shall use $\boldsymbol{Z}_t(s)$ to denote the collection of all pollutant measurements (including PM2.5) available at station s and timepoint t. As an illustration, the temporal variation of the log-transformed PM2.5 measurements are presented in Figure 4, with a few selected stations highlighted. All 38 stations demonstrate a consistent pattern, where the PM2.5 values are usually higher in winter, show a decreasing trend till the months of late summer, and remain low during the monsoon months (Roy et al. 2020).
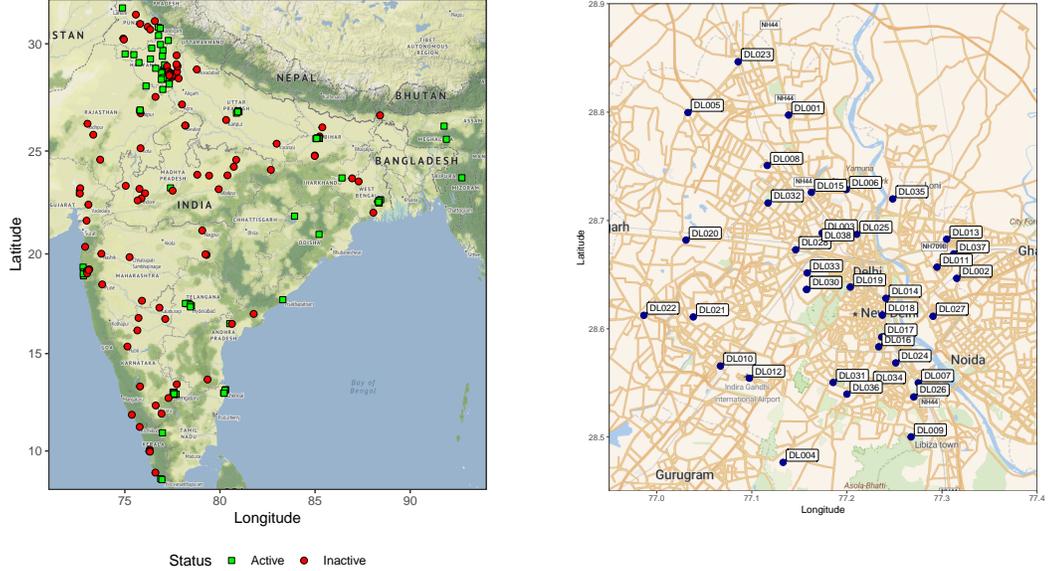
Figure 3: The locations of the measurement stations across India and their status (Left), and in the union territory of Delhi region (Right)

4.1 Application of the proposed method

To demonstrate the applicability of our estimator described in Section 2.2, we model $Y_t(s)$ using $\{\boldsymbol{Z}_j(s) : s \in \mathcal{S}, \ j < t\}$ (this set of covariates is denoted as $\boldsymbol{X}_t$). As it includes measurements of different pollutants in vastly different scales, we also perform a normalization step for each. Although our theoretical analysis considers $\boldsymbol{X}_t$ to be infinite-dimensional, the number of historical observations is always finite in practice. Hence, we take these finite-dimensional covariates $\boldsymbol{X}_t' = (X_{t1}, X_{t2}, \ldots, X_{tp})$ and augment infinitely many zeros (i.e., $\boldsymbol{X}_t' = (X_{t1}, \ldots, X_{tp}, 0, 0, \ldots)$) to make it infinite-dimensional and coherent with the theories derived above. As a distance metric, we choose a discounted $\mathcal{L}_2$-metric that discounts past observations with lag l by ϕ^l for some $\phi \in (0, 1)$. We choose $\phi = 0.9$ by looking at the partial autocorrelation function of the log-transformed pollutant measurements; however, as discussed below, results remain fairly stable for any $\phi \geqslant 0.6$. Additionally, we use orthogonal Legendre polynomials on $[0, 1]^2$ as basis functions, and make a judicious choice of the number of basis functions via a leave-one-location-out cross-validation method. As illustrated in the introduction, the missingness of the observations brings challenges for standard spatiotemporal techniques with the subset of only complete cases or techniques involving regularly-spaced time series observations. Our approach circumvents these issues by modeling through irregularly-spaced time series data and using infinite-dimensional covariates spanning across multiple locations.

To study the efficacy of our proposed estimator, it may be useful to look at the two different plots given in Figures 5 and 6. The first one primarily illustrates the spatial aspect of the fitted model by depicting the estimated level of the response variable across the entire region. The map shows that the PM2.5 levels are higher in central Delhi, corresponding to the increasing population density of the city, while the levels are lower in the suburbs. Also,
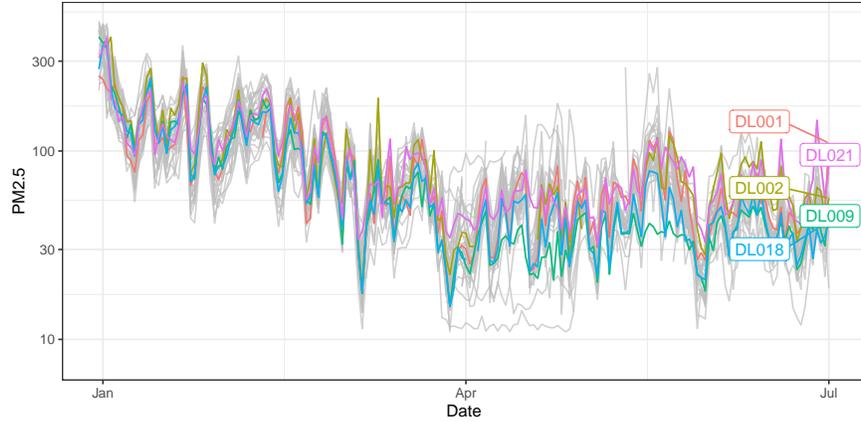
Figure 4: Hourly measurements of PM2.5 in different measurement stations in Delhi for the year 2020; some representative locations are highlighted by colour (y-axis is in log-scale).

the spatial structure is non-isotropic: the PM2.5 levels are higher in the north-west direction compared to the south-east part of Delhi. We also depict this spatial variation across the first day of three different months and for two times of the day (Morning 10 AM, and evening 8 PM). The intensity of the difference in pollution levels between central Delhi and suburbs are most prominent during colder months and decays as the weather gets warmer.
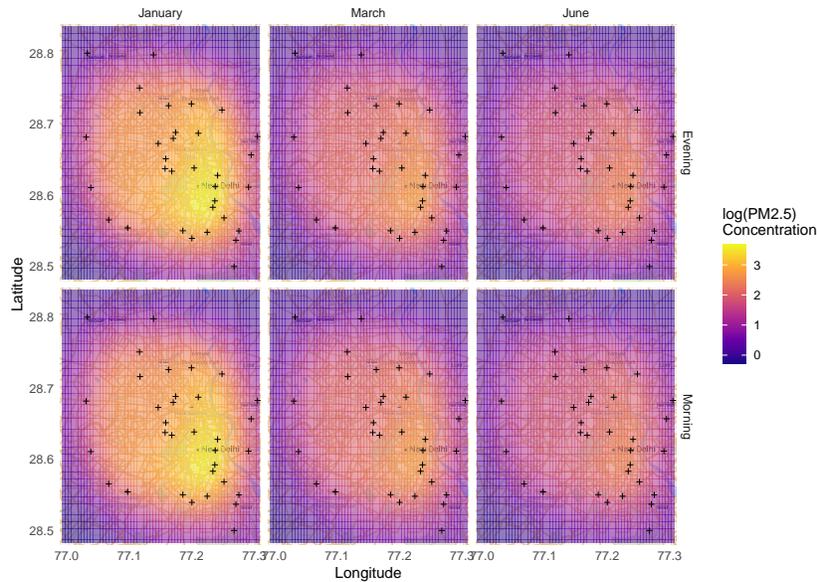

Figure 5: Estimated mean levels of PM2.5-concentration across the entire Delhi region for three different months and during two specific times of the day (Morning 10 AM and evening 8 PM). The black plus points indicate the locations of the measurement stations.

Next, in Figure 6, we depict the predicted mean response for the month of June, 2020 across four randomly selected stations DL001, DL002, DL018 and DL021 using the pollutant measurements of their closest 3 locations (including self) as the covariates. For ease of

19

visualization, we plot the response and estimates aggregated every 4 hours. This figure presents the temporal aspect captured by the fitted model. Although the predicted response matches the general pattern of the observed datapoints, the performance of the estimation depends on the specific location considered, and possibly each location requires a different number of neighboring locations' data as covariates, depending on several external factors such as the terrain heights, the density of vegetation, and the primary usage of the land area in those regions.
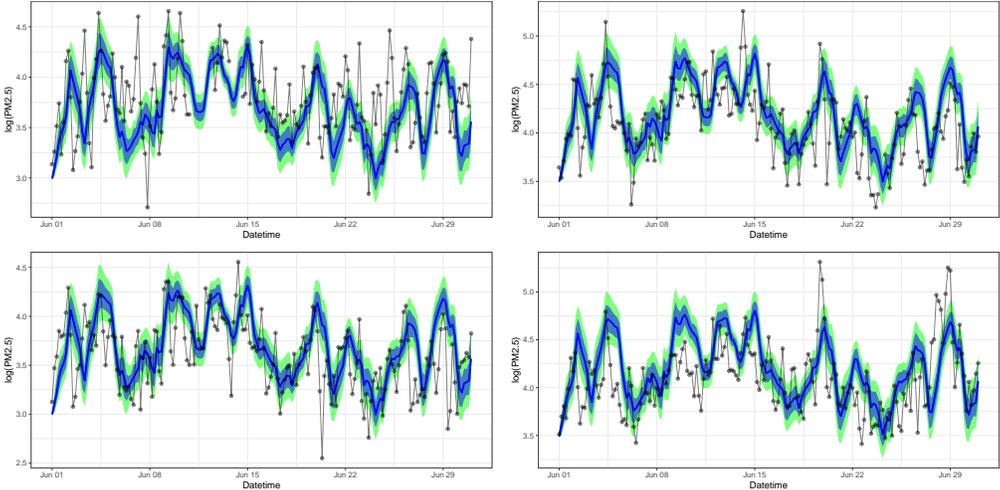

Figure 6: The predicted mean response (log of PM2.5 measurements) for four different measurement stations DL001 (top-left), DL002 (top-right), DL018 (bottom-left), DL021 (bottom-right). The blue band depicts the pointwise confidence interval, and the green band depicts the simultaneous confidence band across all timepoints in June 2020.

To investigate further, we perform estimation for each of the 38 measurement stations for the month of June, 2020 by considering different numbers of covariates taken from its nearest r locations, where r is varied as $r = 1, \ldots, 38$. For example, $r = 1$ indicates that only the past observations of the particular location is considered as covariate, and no spatial effect from other locations is considered in the modeling. With this experiment, for each location, we obtain the optimal number of nearest locations to be taken for covariates by minimizing the RMSE metric between the predicted values and the true log-transformed PM2.5 observations. In Figure 7, we describe this optimal number of locations in the unit of kilometers around each location as a circle, while superimposing the data on the purpose of land usage.

As a general pattern from Figure 7, we see that the stations in the central Delhi require much less information for forecasting compared to the measurement stations in the outskirts. This is possibly because the locations of central Delhi are homogeneous in terms of land usage, residential and economic activities, and hence measurements from each additional nearby station in that region have less novel information to account for. On the other hand, the three stations on the east of the Yamuna river require information on many locations as its nearby industrial land usage and related activities are very different from the residential activities on the other side of the river. We believe that a more detailed analysis of the correlation between various factors and the forecasts from our model may provide useful insights about
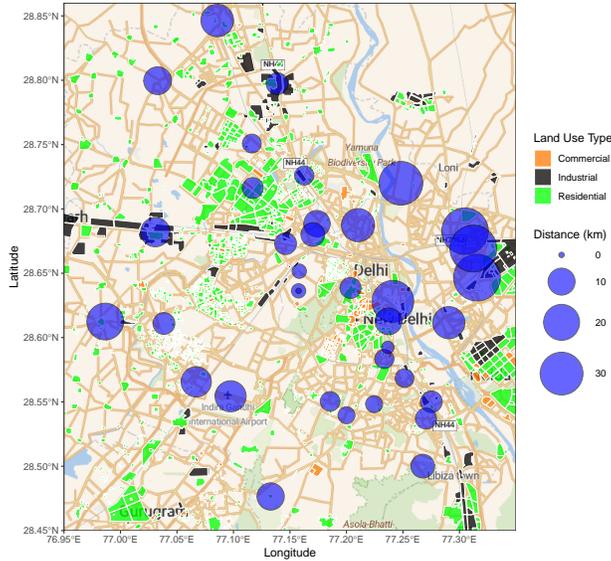
Figure 7: The best choice of the number of spatial covariates (in km) in predicting the PM2.5 measurements for June 2020 in every station at Delhi, along with landuse data.

the optimal and strategic placements of these measurement stations in the future.

4.2 Sensitivity analysis

There are a few hyperparameters of the proposed nonparametric regression method. It is important to understand to what effect these hyperparameters affect the quality of the final prediction. To this end, we consider two different exercises. For the first one, we aim to analyze the degradation in prediction performance between in-training vs out-of-training locations. Let $s_0 \in \mathcal{S}$ be a fixed location. Then, we consider two separate fitted models $\mathcal{M}_{\text{in}}$ and $\mathcal{M}_{\text{out}}$ trained respectively on two separate datasets $\mathcal{D}_{\text{in}}$ and $\mathcal{D}_{\text{out}}$ given by

$$\mathcal{D}_{\text{out}} = \left\{ (Y_t(s), \boldsymbol{X}_t(s)) : t \in \mathcal{T}, s \in \mathcal{S} \setminus \{s_0\} \right\}, \ \mathcal{D}_{\text{in}} = \mathcal{D}_{\text{out}} \cup \left\{ (Y_t(s_0), \boldsymbol{X}_t(s_0)) : t \in \mathcal{T} \right\}.$$

The one-step ahead forecasts for the location s_0 from these two models are compared against each other using Diebold-Mariano (DM) test (Diebold 2015). The results of the DM test, along with the prediction error metric MAPE are summarized in Table 1. It can be seen that for most of the locations (30 out of 35), the null hypothesis (i.e., the two forecasts are same) is not rejected at the nominal level of $\alpha = 0.05$. Also, the error in prediction remains within the range of $10\% - 15\%$ at most of the locations. This serves as an additional verification of the fact that our proposed nonparametric method performs equally well to forecast the levels of PM2.5 at an unknown location not present in the training data.

21

Table 1: MAPE of the two forecast models where the prediction location is included (and excluded) in the training set, along with the results of DM test for their forecasts. Locations "DL006" and "DL011" are removed due to lack of data.

Location	In	Out	DM (p-value)	Location	In	Out	DM (p-value)	Location	In	Out	DM (p-value)
DL001	12.720	12.680	$0.235\,(0.407)$	DL015	11.040	11.080	$-7.534\,(1.000)$	DL027	12.990	13.070	$-3.474\,(1.000)$
DL002	10.370	9.960	$4.425\,(0.000)$	DL016	12.450	12.510	$-10.750\,(1.000)$	DL028	12.420	12.450	$-7.601\,(1.000)$
DL003	12.110	12.120	$-1.100\,(0.864)$	DL017	22.850	22.840	$-4.135\,(1.000)$	DL029	12.000	11.930	$1.657\,(0.049)$
DL004	26.670	29.070	$-19.625\,(1.000)$	DL018	10.690	10.720	$-5.654\,(1.000)$	DL030	20.560	20.210	$11.728\,(0.000)$
DL005	14.610	17.620	$-20.867\,(1.000)$	DL019	12.110	12.060	$3.376\,(0.000)$	DL031	15.650	15.740	$-7.595\,(1.000)$
DL007	13.310	13.400	$-4.598\,(1.000)$	DL020	13.440	13.650	$-11.650\,(1.000)$	DL032	11.130	11.150	$-3.238\,(0.999)$
DL008	11.450	11.440	$1.368\,(0.086)$	DL021	7.410	7.550	$-14.008\,(1.000)$	DL033	15.410	15.740	$-19.082\,(1.000)$
DL009	11.340	12.150	$-14.883\,(1.000)$	DL022	18.070	22.420	$-20.891\,(1.000)$	DL034	10.660	10.660	$1.082\,(0.140)$
DL010	12.060	12.160	$-7.580\,(1.000)$	DL023	23.940	23.360	$6.566\,(0.000)$	DL035	11.020	11.110	$-11.765\,(1.000)$
DL012	15.690	15.740	$-3.727\,(1.000)$	DL024	11.850	11.930	$-11.361\,(1.000)$	DL036	10.990	11.050	$-9.657\,(1.000)$
DL013	16.190	16.980	$-18.435\,(1.000)$	DL025	19.750	19.780	$-0.937\,(0.825)$	DL037	12.500	12.760	$-10.086\,(1.000)$
DL014	9.960	10.070	$-10.466\,(1.000)$	DL026	10.720	10.810	$-5.158\,(1.000)$				

5   Conclusion

The main contribution of this work is a nonparametric regression framework for irregularly sampled spatio-temporal observations with infinite-dimensional covariates. Our work establishes a full asymptotic theory, including simultaneous confidence bands, under a polynomially decaying moment contraction (PMC) condition rather than classical mixing assumptions. We start by establishing the consistency of our estimator of the mean part of the model, which also involves a truncated series approximation of the mean surface. Next, we provide asymptotic normality of the properly scaled and centered mean estimator. We finally end with consistency of variance estimators, which yields the confidence interval and simultaneous confidence band constructions. We address the spatial irregularity through a grid approximation, while the infinite dimensionality of the problem is addressed through small-ball probability assumptions.

Some future directions one can consider are to strengthen the inferential theory from the mean to the covariance structure, by deriving asymptotic distributions and uniform confidence statements for $\sigma(x, s)$, complementing the consistency results already obtained for covariance estimation. Another possible direction is to relax the stationarity assumptions on $\{\boldsymbol{X}_t\}$ and develop an analogous theory under local stationarity or structural breaks, while preserving the PMC-based framework. One could potentially extend the asymptotic results, which are currently obtained pointwise in x, and obtain functional central theorems, which could substantially broaden the scope of simultaneous inference.

Data Availability Statement

The data used in the paper are sourced from the official website of the Central Pollution Control Board in India (CPCB: https://cpcb.nic.in/). The cleaned dataset and R implementation codes are available at the following link: https://curated-webrepo.s3.ap-south-1.amazonaws.com/regression-code.zip.

References

Al-Sulami, D., Jiang, Z., Lu, Z. & Zhu, J. (2017), 'Estimation for semiparametric nonlinear regression of irregularly located spatial time-series data', *Econometrics and Statistics* **2**, 22–35.

Chen, X. & Christensen, T. M. (2015), 'Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions', *Journal of Econometrics* **188**(2), 447–465.

Chen, X., Shao, Q.-M., Wu, W. B. & Xu, L. (2016), 'Self-normalized Cramér-type moderate deviations under dependence', *The Annals of Statistics* **44**(4), 1593 – 1617.

Cressie, N. & Johannesson, G. (2008), 'Fixed rank kriging for very large spatial data sets', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **70**(1), 209–226.

Cressie, N., Sainsbury-Dale, M. & Zammit-Mangion, A. (2022), 'Basis-function models in spatial statistics', *Annual Review of Statistics and Its Application* **9**(1), 373–400.

Cressie, N. & Wikle, C. K. (2011), *Statistics for spatio-temporal data*, John Wiley & Sons.

Deb, S., Neves, C. & Roy, S. (2024), 'Nonparametric quantile regression for spatio-temporal processes', *arXiv preprint arXiv:2405.13783* .

Diebold, F. X. (2015), 'Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests', *Journal of Business & Economic Statistics* **33**(1), 1–1.

Dutta, A. & Jinsart, W. (2022), 'Air pollution in delhi, india: It's status and association with respiratory diseases', *Plos one* **17**(9), e0274444.

Ferraty, F. & Vieu, P. (2003), 'Curves discrimination: A nonparametric functional approach', *Computational Statistics & Data Analysis* **44**(1), 161–173. Special Issue in Honour of Stan Azen: a Birthday Celebration.

Ferraty, F. & Vieu, P. (2004), 'Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination', *Journal of Nonparametric Statistics* **16**(1-2), 111–125.

Goldsmith, J., Crainiceanu, C. M., Caffo, B. & Reich, D. (2012), 'Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements', *Journal of the Royal Statistical Society Series C: Applied Statistics* **61**(3), 453–469.

Haining, R. & Li, G. (2020), *Modelling Spatial and Spatial-Temporal Data: A Bayesian Approach*, Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences.

Hong, S. Y. & Linton, O. (2020), 'Nonparametric estimation of infinite order regression and its application to the risk-return tradeoff', *Journal of Econometrics* **219**(2), 389–424.

Karhunen, K. (1946), 'Zur Spektraltheorie stochastischer Prozesse', *Annales Academiae Scientiarum Fennicae. Series A. I. Mathematica-Physica* **1946**(34), 7.

Karmakar, S. & Wu, W. B. (2020), 'Optimal gaussian approximation for multiple time series', *Statistica Sinica* **30**(3), 1399–1417.

Li, J. & Yang, L. (2023), 'Statistical inference for functional time series', *Statistica Sinica* **33**, 519–549.

Loève, M. (1946), 'Fonctions aléatoires à décomposition orthogonale exponentielle', *La Revue Scientifique* **84**, 159–162.

Ma, S., Yang, L. & Carroll, R. J. (2012), 'A simultaneous confidence band for sparse longitudinal regression', *Statistica Sinica* **22**, 95.

Masry, E. (2005), 'Nonparametric regression estimation for dependent functional data: asymptotic normality', *Stochastic Processes and their Applications* **115**(1), 155–177.

Mead, J., O'Hare, A. & McMenemy, P. (2023), 'Expected goals in football: Improving model performance and demonstrating value', *Plos one* **18**(4), e0282295.

Micchelli, C. A., Xu, Y. & Zhang, H. (2006), 'Universal kernels', *Journal of Machine Learning Research* **7**(95), 2651–2667.

Omar, K. M. T. & Wang, B. (2019), 'Nonparametric regression method with functional covariates and multivariate response', *Communications in Statistics - Theory and Methods* **48**(2), 368–380.

Roy, S., Sengupta, D., Rudra, K. & Saha, U. S. (2020), 'Analysis of pollution patterns in regions of kolkata', *Calcutta Statistical Association Bulletin* **72**(2), 133–170.

Shao, X. & Wu, W. B. (2007), 'Asymptotic spectral theory for nonlinear time series', *The Annals of Statistics* **35**(4), 1773 – 1801.

Wang, H. & Wang, J. (2009), 'Estimation of the trend function for spatio-temporal models', *Journal of Nonparametric Statistics* **21**(5), 567–588.

Wikle, C. K. & Cressie, N. (1999), 'A dimension-reduced approach to space-time kalman filtering', *Biometrika* **86**(4), 815–829.

Wu, W. B. (2005), 'Nonlinear system theory: Another look at dependence', *Proceedings of the National Academy of Sciences* **102**(40), 14150–14154.

Wu, W. B. (2011), 'Asymptotic theory for stationary processes', *Stat. Interface* **4**(2), 207–226.

Wu, W. B. & Zhao, Z. (2007), 'Inference of trends in time series', *Journal of the Royal Statistical Society Series B* **69**(3), 391–410.

Wu, W. B. & Zhou, Z. (2011), 'Gaussian approximations for non-stationary multiple time series', *Statistica Sinica* pp. 1397–1413.

Xiang, D., Qiu, P. & Pu, X. (2013), 'Nonparametric regression analysis of multivariate longitudinal data', *Statistica Sinica* **23**(2), 769–789.

Xiao, H. & Wu, W. B. (2011), 'Asymptotic inference of autocovariances of stationary processes', *arXiv preprint arXiv:1105.3423* .

Yang, K. & Qiu, P. (2018), 'Spatiotemporal incidence rate data analysis by nonparametric regression', *Statistics in medicine* **37**(13), 2094–2107.

Yao, F., Müller, H.-G. & Wang, J.-L. (2005), 'Functional data analysis for sparse longitudinal data', *Journal of the American Statistical Association* **100**(470), 577–590.

Zhao, Z. & Wu, W. B. (2008), 'Confidence bands in nonparametric time series regression', *The Annals of Statistics* **36**(4), 1854–1878.

Zhou, Z. & Wu, W. B. (2010), 'Simultaneous inference of linear models with time varying coefficients', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **72**(4), 513–531.

Supplementary material to "Nonparametric regression of spatio-temporal data using infinite-dimensional covariates"

S.1 Inference of the Variance Component

In the main paper, we have tackled the problem of estimating the mean function $\mu(\boldsymbol{x}, s)$, and the component-wise covariance operator $\sigma_{kl}(\boldsymbol{x})$, but a detailed exposition on the estimation of the scale component $\sigma(\boldsymbol{x}, s)$ in the model (1) mentioned in the main manuscript was not included. This section establishes the estimation procedure and the asymptotic properties of the estimator for the variance component. Note that all cross-references in this document may pertain to the contents in the supplement (whenever it has the prefix S.) or the main manuscript (if there is no additional prefix).

S.1.1 Estimation Algorithm

Akin to the case with the mean function, we again start with the reduced model established in equation (8) of the main paper and use the decomposition given in (6) to estimate the component functions $\sigma_u(\boldsymbol{x})$ for each $u \in \mathbb{Z}^+$. Note that, by definition of $\sigma_{kl}(\boldsymbol{x})$, we know that

$$\sigma_{kl}(\boldsymbol{x}) = \sum_{u=1}^{\infty} \sum_{v=1}^{\infty} c_{u,v}(k,l)\sigma_u(\boldsymbol{x})\sigma_v(\boldsymbol{x}), \quad \text{for any } k, l \in \mathbb{Z}^+. \tag{S.1}$$

If we substitute $\sigma_{kl}(\boldsymbol{x})$ by its estimate $\widehat{\sigma}_{k,l}(\boldsymbol{x})$ as in (28) for each $k, l \in \mathbb{Z}^+$, we have a countable set of equations involving the components $\sigma_u(\boldsymbol{x})$. If such a system is solvable, then it would let us recover an estimate of the variance component as $\widehat{\sigma}(\boldsymbol{x}, s) = \sum_{u=1}^{\infty} \widehat{\sigma}_u(\boldsymbol{x}) b_u(s)$ where $\widehat{\sigma}_u(\boldsymbol{x})$ is a solution of these equations.

One approach to solve this system is as follows: Following the decomposition of $\sigma(\boldsymbol{x}, s)$ as in (6), we must have $\sum_{u=1}^{\infty} \sigma_u(\boldsymbol{x})$ to be finitely summable, and hence for any given $\epsilon > 0$, there exists a $U_\epsilon \in \mathbb{Z}^+$ such that $\sum_{u=U_\epsilon+1}^{\infty} \sigma_u(\boldsymbol{x}) < \epsilon$ for each $\boldsymbol{x} \in \chi$. Additionally, as $|\rho(s, s')| \leqslant 1$, by an application of Cauchy-Schwartz inequality, it follows that $c_{u,v}(k,l) \leqslant 1$ for all $u, v, k, l \in \mathbb{Z}^+$. Therefore, we have

$$\sum_{\max(u,v)>U_\epsilon} |c_{u,v}(k,l)\sigma_u(x)\sigma_v(x)| \leqslant \sum_{u=U_\epsilon+1}^{\infty} \sum_{v=1}^{\infty} \sigma_u(x)\sigma_v(x) + \sum_{u=1}^{\infty} \sum_{v=U_\epsilon+1}^{\infty} \sigma_u(x)\sigma_v(x) = \mathcal{O}(\epsilon).$$

As a result, we can truncate the infinite sums present in the right-hand side of (S.1) with an error of at most $\mathcal{O}(\epsilon)$. At this point, let us denote the U_ϵ-length vector comprising of $\{\sigma_u(\boldsymbol{x})\}_{u=1}^{U_\epsilon}$ as $\boldsymbol{\sigma}_{1:U_\epsilon}(\boldsymbol{x})$. Let us denote the corresponding vector of the estimates as $\widehat{\boldsymbol{\sigma}}_{1:U_\epsilon}(\boldsymbol{x})$. Let us also denote $\boldsymbol{C}_{k,l}$ as a $U_\epsilon \times U_\epsilon$-matrix comprising of the elements $\{c_{u,v}(k,l)\}_{u,v=1}^{U_\epsilon}$. Using these notations and the truncation argument, we can then rewrite the system of quadratic equations to be solved as a system of linear equations in the unknown coordinates of $\text{vec}(\boldsymbol{\sigma}_{1:U_\epsilon})$,

which yields

$$\widehat{\boldsymbol{\sigma}}_{1:U_\epsilon}(\boldsymbol{x})\widehat{\boldsymbol{\sigma}}_{1:U_\epsilon}(\boldsymbol{x})^\intercal = \text{vec}^{-1}\left([\text{vec}(\boldsymbol{C}_{11}): \cdots : \text{vec}(\boldsymbol{C}_{U_\epsilon, U_\epsilon})]^{-1}\text{vec}(\widehat{\boldsymbol{\Sigma}}(\boldsymbol{x}))\right) = \boldsymbol{A}, \qquad \text{(S.2)}$$

where $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{x})$ is the $U_\epsilon \times U_\epsilon$-size matrix with entries $\sigma_{kl}(\boldsymbol{x})$ for $k, l = 1, 2, \ldots, U_\epsilon$, and $\text{vec}(\cdot)$ denotes the vectorization operation of a matrix. However, the solution of the right-hand side of (S.2) may not result in a rank-one matrix, hence we obtain the estimate $\widehat{\boldsymbol{\sigma}}_{1:U_\epsilon}(\boldsymbol{x})$ by considering the partial eigendecomposition of the output matrix. In summary, we obtain the estimate of the variance component as expressed in Algorithm 1.

Algorithm 1: Algorithm for estimation of $\sigma(\boldsymbol{x}, s)$.

Input: $t_i \in \mathcal{T}$, $\{s_{t_{ij}}\}_{j=1}^{n_i} \subseteq \mathcal{S}$, Response observations $\{Y_{t_i}(s_{t_{ij}})\}_{j=1}^{n_i}$

1 for $k = 1, 2, \ldots, U_\epsilon$ **do**

2 $\quad$ Compute $\widehat{\sigma}_{kl}(\boldsymbol{x})$ using (28)

3 Calculate $\boldsymbol{A}$ as the RHS of (S.2)

4 $\boldsymbol{A}^* \leftarrow \arg\min_{\boldsymbol{B} \text{ is p.d.}} \|\boldsymbol{A} - \boldsymbol{B}\|_F^2$

5 $\lambda_1(\boldsymbol{A}^*) \leftarrow$ maximum eigenvalue of $\boldsymbol{A}^*$ and $e_1(\boldsymbol{A}^*)$ be corresponding eigenvector

6 for $u = 1, 2, \ldots, U_\epsilon$ **do**

7 $\quad$ $\widehat{\sigma}_u(\boldsymbol{x}) \leftarrow \sqrt{\lambda_1(\boldsymbol{A}^*)}|e_{1u}(\boldsymbol{A}^*)|$

One caveat of the above approach is that to compute the matrix $\boldsymbol{C}_{kl}$ and its elements $c_{u,v}(k, l)$, one requires the knowledge of the spatial correlation function $\rho(s, s')$. If it is known, one can directly use the above algorithm. Often in different contexts, the functional form of $\rho(s, s')$ may be known, e.g., $\rho(s, s') = e^{-\rho\|s-s\|^2}$ for some unknown parameter $\rho > 0$. In this case, we can estimate ρ by the choice that minimizes the discrepancy $\|A - \widehat{\boldsymbol{\sigma}}_{1:U_\epsilon}(\boldsymbol{x})\widehat{\boldsymbol{\sigma}}_{1:U_\epsilon}(\boldsymbol{x})^\intercal\|_2$ in the above estimation process.

In a completely nonparametric setup, the form of $\rho(s, s')$ is unknown. In this case, one can employ an iterative scheme as follows. Given an estimate of $\widehat{\rho}^{(m)}(s, s')$ at m^{th} step of the iteration, we can employ the above procedure to obtain an estimate of $\widehat{\sigma}^{(m)}(\boldsymbol{x}, s)$ for any $\boldsymbol{x} \in \chi$ and $s \in \mathcal{S}$. However, using (5), we can obtain a revised estimate of the $\rho(\cdot, \cdot)$ function as

$$\widehat{\rho}^{(m+1)}(s, s') = \frac{\sum_{i=1}^n K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{x})\|)Y_{t_i}(s)Y_{t_i}(s')}{\sum_{i=1}^n K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{x})\|)\widehat{\sigma}^{(m)}(\boldsymbol{x}, s)\widehat{\sigma}^{(m)}(\boldsymbol{x}, s')}. \qquad \text{(S.3)}$$

The final estimates of $\rho(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ are given by their corresponding converged values.

S.1.2 Asymptotic Results

From Theorem 4, we see that for any fixed $k, l \in \mathbb{Z}^+$, the quantity $\sigma_{kl}(\boldsymbol{x})$ can be estimated consistently. Let us denote $\boldsymbol{\Sigma}_{1:K,1:K}(\boldsymbol{x})$ as the matrix with entries $\sigma_{kl}(\boldsymbol{x})$ for $k, l \in \{1, 2, \ldots, K\}$, and denote the estimated matrix as $\widehat{\boldsymbol{\Sigma}}_{1:K,1:K}(\boldsymbol{x})$ with corresponding entries $\widehat{\sigma}_{kl}(\boldsymbol{x})$. Although Theorem 4 shows that each element of the matrix $\boldsymbol{\Sigma}_{1:U_\epsilon,1:U_\epsilon}(\boldsymbol{x}) - \widehat{\boldsymbol{\Sigma}}_{1:U_\epsilon,1:U_\epsilon}(\boldsymbol{x})$ is small, the errors can accumulate based on the choice of ϵ through a considerable increase in U_ϵ. Let us denote a sequence of $U_\epsilon \times U_\epsilon$-matrices $\boldsymbol{C}_{k,l}$ for each $k, l \in \{1, 2, \ldots, U_\epsilon\}$ comprising of the

elements $c_{k,l}(u, v)$ as defined in (22). Then, based on the algorithm illustrated in Section S.1.1 and an application of Weyl's inequality, it is easy to see that the estimation error of the variance component is bounded by $o_{\mathbb{P}}(\sqrt{2U_\epsilon} \sup_{k,l} \mathrm{cond}(\boldsymbol{C}_{k,l}))$. Here, $\mathrm{cond}(\boldsymbol{A})$ denotes the condition number of a matrix $\boldsymbol{A}$. Based on the above discussion, we now state a consistency theorem for the estimation of $\sigma(\boldsymbol{x}, s)$ without a formal proof.

Proposition 1. *Assume the same conditions as in Theorem 4. Given a fixed $x \in \chi$, for any $\epsilon > 0$, find $U_\epsilon \in \mathbb{Z}^+$ such that $\sum_{u=U_\epsilon}^\infty \sigma_u(\boldsymbol{x}) < \epsilon$. Then, for any $u \in \{1, 2, \ldots, U_\epsilon\}$,*

$$|\hat{\sigma}_u(\boldsymbol{x}) - \sigma_u(\boldsymbol{x})| = \mathcal{O}(\epsilon) + o_{\mathbb{P}}\left(\sqrt{2U_\epsilon} \sup_{k,l} cond(\boldsymbol{C}_{k,l})\right).$$

S.2 Proofs of the Results

S.2.1 Proof of Proposition 1

Let $\boldsymbol{X}'_{t_i}$ be another independent copy of $\boldsymbol{X}_{t_i}$. Then, consider the chain of equality,

$$
\begin{aligned}
(\mathbb{E}(f(\boldsymbol{X}_{t_i})) - \mathbb{E}(f(\boldsymbol{X}_{t_i}) \mid \mathcal{F}_{t_i-m}))^2 &= (\mathbb{E}\left(f(\boldsymbol{X}_{t_i}) - \mathbb{E}(f(\boldsymbol{X}_{t_i})) \mid \mathcal{F}_{t_i-m}\right))^2 \\
&\leqslant \mathbb{E}\left((f(\boldsymbol{X}_{t_i}) - \mathbb{E}(f(\boldsymbol{X}_{t_i})))^2 \mid \mathcal{F}_{t_i-m}\right) \\
&= \frac{1}{2}\mathbb{E}\left((f(\boldsymbol{X}_{t_i}) - f(\boldsymbol{X}'_{t_i}))^2 \mid \mathcal{F}_{t_i-m}\right) \\
&\leqslant \frac{C^2}{2}\mathbb{E}\left(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{X}'_{t_i})\|^2 \mid \mathcal{F}_{t_i-m}\right), \text{ as } f \text{ is Lipschitz,} \\
&= \frac{C^2}{2}\mathbb{E}\left(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{X}'_{t_i})\|^2 \mid \ldots, \xi_{t_i-m-1}, \xi_{t_i-m}\right) \\
&\leqslant \frac{C^2}{2}\Delta_2^2(m),
\end{aligned}
$$

where the last line follows from the definition of $\Delta_2(m)$. Taking square root to both sides now yields (17).

S.2.2 Proof of Corollary 1

Let us fix any $\boldsymbol{x}_0 \in \chi$ and $\delta > 0$, and consider any two distinct points $\boldsymbol{x}, \boldsymbol{y} \in \chi$ such that $\|\boldsymbol{D}^{-1}(\boldsymbol{x} - \boldsymbol{y})\| > \delta$. Since $h_n \to 0$ as $n \to \infty$ as in Assumption (A3), for sufficiently large n we must have $\delta > 2h_n\lambda$. As a result, by applying triangle inequality, we obtain that either $\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{x}_0)\|$ or $\|\boldsymbol{H}_n^{-1}(\boldsymbol{y} - \boldsymbol{x}_0)\|$ must be greater than λ, where $\boldsymbol{H}_n = h_n\boldsymbol{D}$. Since the kernel has a bounded support $[0, \lambda]$, either $K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{x}_0)\|) = 0$ or $K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{y} - \boldsymbol{x}_0)\|) = 0$. Therefore, the following Lipschitz property holds for any $\boldsymbol{x}_0 \in \chi$:

$$|K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{x}_0)\|) - K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{y} - \boldsymbol{x}_0)\|)| \leqslant C_2 \leqslant (C_2/\delta)\|\boldsymbol{D}^{-1}(\boldsymbol{x} - \boldsymbol{y})\|.$$

Now, we can decompose the covariance into two terms: When $\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\| > \delta$, we make use of the Lipschitz property along with Proposition 1, and when $\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\| \leqslant \delta$,

we make use of the bound given in (16). Since $\mathbb{E}\left[K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\|)\right]$ remains same for any t due to the stationarity of $\boldsymbol{X}_t$, let us choose c to denote this expectation. For any $t' < t$, this now yields

$$
\begin{aligned}
&\text{Cov}\left(\mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})), \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x}))\right) \\
&= \mathbb{E}\left[\left(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\|) - c\right)\left(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x})\|) - c\right)\right] \\
&= \mathbb{E}\left[\left(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\|) - c\right)\left(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x})\|) - c\right)\mathbf{1}_{\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\|>\delta}\right] + \\
&\quad \mathbb{E}\left[\left(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\|) - c\right)\left(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x})\|) - c\right)\mathbf{1}_{\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\|\leqslant\delta}\right] \\
&\leqslant \mathbb{E}\left[\left(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x})\|) - c\right)\mathbb{E}\left[\left(K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\|) - c\right)\mathbf{1}_{\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\|>\delta} \mid \mathcal{F}_{t'}\right]\right] + \\
&\quad C_2^2 C \phi_{\boldsymbol{x}}(h_n\lambda)\mathbb{P}(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\| \leqslant \delta), \text{ by (16)} \\
&\leqslant \frac{\sqrt{2}C_2^2 C}{\delta}\Delta_2(|t - t'|)\mathbb{P}(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\| > \delta) + C_2^2 C \phi_{\boldsymbol{x}}(h_n\lambda)\mathbb{P}(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\| \leqslant \delta)
\end{aligned}
$$

The above inequality holds for any $\delta > 0$ and for all sufficiently large n. By rewriting $\mathbb{P}\left(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\| \leqslant \delta\right)$ as a conditional expectation and using the stationarity of $\boldsymbol{X}_t$, we obtain

$$
\mathbb{P}\left(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_{t'})\| \leqslant \delta\right) = \mathbb{E}\left(\mathbb{P}\left(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_t - \boldsymbol{X}_s)\| \leqslant \delta\right) \mid \boldsymbol{X}_t\right) = \mathbb{E}\left(\phi_{\boldsymbol{X}_t}(\delta)\right) = \mathbb{E}\left(\phi_{\boldsymbol{X}_0}(\delta)\right).
$$

Since we simply require $\delta > 2h_n\lambda$ for sufficiently large n, we can choose $\delta = 3h_n\lambda$. Therefore,

$$
\begin{aligned}
\text{Cov}\left(\mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})), \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t'} - \boldsymbol{x}))\right) &\leqslant \frac{\sqrt{2}CC_2^2}{9h_n^2\lambda^2}\Delta_2(|t - t'|)(1 - \mathbb{E}\left(\phi_{\boldsymbol{X}_0}(3h_n\lambda)\right)) \\
&\quad + C_2^2(1 + C)\phi_{\boldsymbol{x}}^2(h_n\lambda)\mathbb{E}\left(\phi_{\boldsymbol{X}_0}(3h_n\lambda)\right).
\end{aligned}
$$

This is same as (18) that we wanted to show.

S.2.3 Proof of Proposition 2

Let us fix $t_i \in \mathcal{T}$ and $k \in \mathbb{Z}^+$. With $\mathcal{S} = [0, 1]^d$, let us denote H_l as the lth hypercube with diameter $\epsilon_{t_i}^*$, with its center at s_l^c, as described in Algorithm 1. Also note that since $\mathcal{S}$ is compact and $b_k(s)$ is continuous, $\sup_{s \in \mathcal{S}} |b_k(s)|$ exists and is finite. Similarly, by Assumption (A1), we know that the same conclusion holds for $\tilde{\mu}(s) = \mathbb{E}(\mu(\boldsymbol{X}_0, s))$ and $\tilde{\sigma}(s) = \mathbb{E}(\sigma(\boldsymbol{X}_0, s))$. Let us take M to a uniform upper bound of all these three functions.

Now, the expected error due to the modified Monte-Carlo procedure is

$$
\begin{aligned}
\mathbb{E}\left(\widehat{Y}_{t_ik}^* - Y_{t_ik}^*\right) &= \mathbb{E}\left(\frac{1}{r_i}\sum_{l=1}^{r_i} Y_{t_i}(s_{t_ij_l})b_k(s_{t_ij_l}) - \int_{[0,1]^d} Y_{t_i}(s)b_k(s)ds\right) \\
&= \frac{1}{r_i}\sum_{l=1}^{r_i}\mathbb{E}\left(Y_{t_i}(s_{t_ij_l})b_k(s_{t_ij_l}) - \int_{H_l} Y_{t_i}(s)b_k(s)ds\right) \\
&= \frac{1}{r_i}\sum_{l=1}^{r_i}\mathbb{E}\left(Y_{t_i}(s_{t_ij_l})b_k(s_{t_ij_l}) - Y_{t_i}(s_l^*)b_k(s_l^*)\right),
\end{aligned}
$$

4

for some $s_l^* \in H_l$ by an application of Mean Value Theorem. Continuing,

$$
\begin{aligned}
\mathbb{E}\left(\widehat{Y}_{t_i k}^* - Y_{t_i k}^*\right) &= \frac{1}{r_i} \sum_{l=1}^{r_i} \mathbb{E}\left(\mu(\boldsymbol{X}_{t_i}, s_{t_i j_l}) b_k(s_{t_i j_l}) - \mu(\boldsymbol{X}_{t_i}, s_l^*) b_k(s_l^*)\right) \\
&= \frac{1}{r_i} \sum_{l=1}^{r_i} \mathbb{E}\left(\mu(\boldsymbol{X}_0, s_{t_i j_l}) b_k(s_{t_i j_l}) - \mu(\boldsymbol{X}_0, s_l^*) b_k(s_l^*)\right), \text{ by stationarity} \\
&\leqslant M \sup_{\|s-s'\| \leqslant \epsilon_{t_i}^*} |b_k(s) - b_k(s')| + M \sup_{\|s-s'\| \leqslant \epsilon_{t_i}^*} |\mathbb{E}(\mu(\boldsymbol{X}_0, s) - \mu(\boldsymbol{X}_0, s'))|,
\end{aligned}
$$

which takes care of the bias term of the error. Here, the constant M is the generic uniform bound on the basis function $b_k(s)$ and the expected mean function $\tilde{\mu}(s)$; see the discussions following Assumption (A2). Now, turning our attention to the variance, we obtain

$$
\begin{aligned}
\operatorname{Var}\left(\widehat{Y}_{t_i k}^*\right) &= \operatorname{Var}\left(\frac{1}{r_i} \sum_{l=1}^{r_i} Y_{t_i}(s_{t_i j_l}) b_k(s_{t_i j_l})\right) \\
&= \frac{1}{r_i^2}\left[\sum_{l=1}^{r_i} b_k^2(s_{t_i j_l}) \operatorname{Var}\left(Y_{t_i}(s_{t_i j_l})\right) + \sum_{l_1, l_2} b_k(s_{t_i j_{l_1}}) b_k(s_{t_i j_{l_2}}) \operatorname{Cov}\left(Y_{t_i}(s_{t_i j_{l_1}}), Y_{t_i}(s_{t_i j_{l_2}})\right)\right] \\
&\leqslant \frac{M}{r_i^2} \sum_{l=1}^{r_i} \mathbb{E}(\sigma^2(\boldsymbol{X}_{t_i}, s_{t_i j_l})) + \frac{M'}{r_i^2} \sum_{l_1 \neq l_2}^{r_i} \mathbb{E}(\sigma(\boldsymbol{X}_{t_i}, s_{t_i j_{l_1}}) \sigma(\boldsymbol{X}_{t_i}, s_{t_i j_{l_2}})) \rho(s_{t_i j_{l_1}}, s_{t_i j_{l_2}}) \\
&= \frac{M}{r_i} \mathbb{E}(\sigma^2(\boldsymbol{X}_0, s_{t_i j_l})) + \frac{M'}{r_i^2} \sum_{l_1 \neq l_2}^{r_i} \mathbb{E}(\sigma(\boldsymbol{X}_0, s_{t_i j_{l_1}}) \sigma(\boldsymbol{X}_0, s_{t_i j_{l_2}})) \rho(s_{t_i j_{l_1}}, s_{t_i j_{l_2}}), \\
&\leqslant \frac{M''}{r_i} + \frac{M''}{r_i^2} \sum_{l_1 \neq l_2}^{r_i} \rho(s_{t_i j_{l_1}}, s_{t_i j_{l_2}})
\end{aligned}
$$

Here, $M, M',$ and M'' are some generic constants which serve as the uniform bound for the basis function $b_k(s)$ and the expected mean $\tilde{\mu}(s)$ and variance $\tilde{\sigma}^2(s)$ functions.

We have the first term $\mathcal{O}((\epsilon_{t_i}^*)^d)$. For the second quantity, let us look at the distances between two points $s_{t_i j_{l_1}}$ and $s_{t_i j_{l_2}}$ for each pair of choices. For a fixed j_{l_1}, there are $(2k+1)^d - (2k-1)^d$ hypercubes which are k units distant from $s_{t_i j_{l_1}}$ in L^1 distance. Correspondingly, there are $(2k+1)^d - (2k-1)^d = \mathcal{O}(k^{d-1})$ many choices of l_2 such that the Euclidean distance between $s_{t_i, j_{l_1}}$ and $s_{t_i, j_{l_2}}$ is at least $k\epsilon_{t_i}^*/\sqrt{d}$. It now follows that due to the spatial dependence structure given in (5), we obtain an upper bound of the second term

$$
\begin{aligned}
\frac{M^4}{r_i^2} \sum_{l_1 \neq l_2}^{r_i} \rho(s_{t_i j_{l_1}}, s_{t_i j_{l_2}}) &\leqslant \frac{M^4}{r_i^2} \sum_{l_1=1}^{r_i} \sum_k \left((2k+1)^d - (2k-1)^d\right) (k\epsilon_{t_i}^*/\sqrt{d})^{-(d+\delta)} \\
&= \frac{M^4}{r_i} \sum_k \left((2k+1)^d - (2k-1)^d\right) (k\epsilon_{t_i}^*/\sqrt{d})^{-(d+\delta)} \\
&\leqslant \frac{M^4}{r_i} \sum_{k=1}^{\infty} \mathcal{O}(k^{-(1+\delta)}) = \mathcal{O}(1/r_i),
\end{aligned}
$$

since $\sum_{k=1}^{\infty} \mathcal{O}(k^{-(1+\delta)}) < \infty$ for $\delta > 0$.

5

Combining all of the above and using the fact that $r_i(\epsilon_{t_i}^*/\sqrt{d})^d = 1$, we get

$$
\mathbb{E}^2|\widehat{Y}_{t_ik}^* - Y_{t_ik}^*| \leqslant \mathbb{E}\left[\left(\widehat{Y}_{t_ik}^* - Y_{t_ik}^*\right)^2\right]
$$

$$
= \mathcal{O}\left(\max\left\{(\epsilon_{t_i}^*)^d, \sup_{\|s-s'\|\leqslant\epsilon_{t_i}^*}|b_k(s) - b_k(s')|, \sup_{\|s-s'\|\leqslant\epsilon_{t_i}^*}|\widetilde{\mu}(s) - \widetilde{\mu}(s'))|\right\}^2\right).
$$

Taking square root on both sides now completes the proof.

S.2.4   Proof of Theorem 1

We start by fixing any $k \in \mathbb{Z}^+$ and any $\boldsymbol{x} \in \chi$. Denoting $K_t = K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_t - \boldsymbol{x})\|)$, we rewrite the estimate of k-th component of the mean function as

$$
\widehat{\mu}_k(\boldsymbol{x}) = \frac{\frac{1}{n}\sum_{i=1}^n K_{t_i}\widehat{Y}_{t_ik}^*/\mathbb{E}(K_0)}{\frac{1}{n}\sum_{i=1}^n K_{t_i}/\mathbb{E}(K_0)} = \frac{\widehat{\mu}_{k,2}(\boldsymbol{x})}{\widehat{\mu}_{k,1}(\boldsymbol{x})}, \tag{S.4}
$$

where $\widehat{\mu}_{k,1}(\boldsymbol{x})$ and $\widehat{\mu}_{k,2}(\boldsymbol{x})$ are the quantities it is replacing. Then, the error in estimation can be expressed as

$$
\begin{aligned}
\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}) &= \frac{\widehat{\mu}_{k,2}(\boldsymbol{x})}{\widehat{\mu}_{k,1}(\boldsymbol{x})} - \mu_k(\boldsymbol{x}) \\
&= \frac{\widehat{\mu}_{k,2}(\boldsymbol{x}) - \mu_k(x)\widehat{\mu}_{k,1}(\boldsymbol{x})}{\widehat{\mu}_{k,1}(\boldsymbol{x})} \\
&= \frac{\mathbb{E}\widehat{\mu}_{k,2}(\boldsymbol{x}) - \mu_k(\boldsymbol{x})}{\widehat{\mu}_{k,1}(\boldsymbol{x})} + \frac{\widehat{\mu}_{k,2}(\boldsymbol{x}) - \mathbb{E}\widehat{\mu}_{k,2}(\boldsymbol{x})}{\widehat{\mu}_{k,1}(\boldsymbol{x})} - \mu_k(\boldsymbol{x})\frac{\widehat{\mu}_{k,1}(\boldsymbol{x}) - \mathbb{E}\widehat{\mu}_{k,1}(\boldsymbol{x})}{\widehat{\mu}_{k,1}(\boldsymbol{x})} \\
&= (A_1 + A_2 + A_3)/\widehat{\mu}_{k,1}(\boldsymbol{x}), \tag{S.5}
\end{aligned}
$$

where we use the mean stationarity of $\boldsymbol{X}_t$ to obtain that $\mathbb{E}\widehat{\mu}_{k,1}(\boldsymbol{x}) = 1$.

At this point, the proof follows by showing three things:

1. $\mathbb{E}(\widehat{\mu}_{k,1}(\boldsymbol{x})) = 1$ and $(\widehat{\mu}_{k,1}(\boldsymbol{x}) - 1) = o_{\mathbb{P}}(1)$, hence the denominator converges to 1 and the quantity A_3 converges to 0 in probability. This is demonstrated in Lemma 1.

2. The numerator of the term A_1 converges to 0 in probability, uniformly over all $\boldsymbol{x} \in \chi$. This is demonstrated in Lemma 2.

3. The numerator of the term A_2 converges to 0 in L_1 norm, hence in probability as well. This is illustrated in Lemma 3.

These, together with an application of Slutsky's theorem, will imply that the pointwise convergence $(\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x})) \xrightarrow{P} 0$ holds for any fixed k and $\boldsymbol{x} \in \chi$. More details about this specific argument can be found in Hong & Linton (2020), where the authors use the exact same decomposition.

Now, we will proceed to verify each of these claims through a series of Lemmas.

Lemma 1. *Suppose the Assumptions (A1)-4 hold. Then for any fixed $k \in \mathbb{Z}^+, \boldsymbol{x} \in \chi$, as $n \to \infty$,*

$$\widehat{\mu}_{k,1}(x) - 1 = o_{\mathbb{P}}(1).$$

Proof. Since $\boldsymbol{X}_t$ is second-order stationary, we have

$$\mathbb{E}(\widehat{\mu}_{k,1}(\boldsymbol{x})) = \frac{1}{n\mathbb{E}(K_0)} \sum_{i=1}^{n} \mathbb{E}(K_{t_i}) = 1.$$

Considering the variance, we get

$$
\begin{aligned}
&\operatorname{Var}(\widehat{\mu}_{k,1}(\boldsymbol{x})) \\
=\ & \frac{1}{n^2\mathbb{E}^2(K_0)} \left[\sum_{i=1}^{n} \operatorname{var}(K_{t_i}^2) + \sum_{|i-j|>0} \operatorname{Cov}(K_{t_i}, K_{t_j}) \right] \\
=\ & \frac{1}{n\mathbb{E}^2(K_0)}(\mathbb{E}(K_0^2) - \mathbb{E}^2(K_0)) + \frac{1}{n^2\mathbb{E}^2(K_0)} \sum_{|i-j|>0} \operatorname{Cov}(K_{t_i}, K_{t_j}) \\
\leqslant\ & \frac{(1+o(1))\xi_2}{(1-o(1))n\phi_{\boldsymbol{x}}(h_n\lambda)\xi_1^2} + \frac{1}{(1-o(1))\xi_1^2} \sum_{|i-j|>0} \left[\frac{\sqrt{2}CC_2^2\Delta_2(|t_i - t_j|)}{9n^2h_n^2\lambda^2\phi_{\boldsymbol{x}}^2(h_n\lambda)}p_n + \frac{(1+C)C_2^2}{n^2}\bar{p}_n \right]
\end{aligned}
$$

where the last line follows from (18), where $p_n = 1 - \mathbb{E}(\phi_{\boldsymbol{X}_0}(3h_n\lambda))$ and $\bar{p}_n = 1 - p_n$. Now, for the first term, we apply the fact that $n\phi_{\boldsymbol{x}}(h_n\lambda) \to \infty$ as $n \to \infty$ as shown in (14). For the second term, we note that due to the PMC condition 4, we have $\sum_{l=1}^{n-1} \Delta_2(l) = \mathcal{O}(n^{-\tau+2}) < \infty$, and hence the numerator is $\mathcal{O}(n)$. Therefore, the quantity goes to 0 as $nh_n^2\phi_{\boldsymbol{x}}^2(h_n\lambda) \to \infty$ by the choice of the bandwidth given in Assumption (A3). For the third and the final term, as shown in (18), it is of $\mathcal{O}(\mathbb{E}(\phi_{\boldsymbol{X}_0}(3h_n\lambda)))$. Since, $\phi_{\boldsymbol{x}}(h_nu) \to 0$ for any fixed $\boldsymbol{x} \in \chi$ and any $u > 0$, the third term also becomes asymptotically negligible. Therefore, we have

$$\operatorname{Var}(\widehat{\mu}_{k,1}(x)) \to 0,$$

as $n \to \infty$. An application of Chebyshev's inequality now completes the proof. $\qquad\square$

Lemma 2. *Suppose that the Assumptions (A1)-(A4) hold. Additionally, the effective spatial resolution $\epsilon_{t_i}^*$ decays to 0 uniformly over $t_i \in \mathcal{T}$. Then, for any fixed $k \in \mathbb{Z}^+$, we have $\sup_{\boldsymbol{x}\in\chi} |\mathbb{E}\widehat{\mu}_{k,2}(\boldsymbol{x}) - \mu_k(\boldsymbol{x})| = o_{\mathbb{P}}(1)$ as $n \to \infty$.*

Proof. Fix any $k \in \mathbb{Z}^+$. By using second-order stationarity of $\boldsymbol{X}_t$, we decompose the given quantity as

$$
\begin{aligned}
&\sup_{\boldsymbol{x}\in\chi} |\mathbb{E}\widehat{\mu}_{k,2}(\boldsymbol{x}) - \mu_k(\boldsymbol{x})| \\
=\ & \sup_{\boldsymbol{x}\in\chi} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{E}K_{t_i}\widehat{Y}_{tk}^*}{\mathbb{E}K_0} - \mu_k(\boldsymbol{x}) \right| \\
=\ & \sup_{\boldsymbol{x}\in\chi} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{E}K_{t_i}(\widehat{Y}_{t_ik}^* - Y_{t_ik}^*)}{\mathbb{E}K_0} \right| + \sup_{\boldsymbol{x}\in\chi} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{E}K_{t_i}Y_{t_ik}^*}{\mathbb{E}K_0} - \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{E}K_{t_i}\mu_k(\boldsymbol{x})}{\mathbb{E}K_0} \right|
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant \sup_{\boldsymbol{x}\in\chi} \frac{1}{n}\sum_{i=1}^{n} \frac{\left|\mathbb{E}K_{t_i}(\widehat{Y}^*_{t_ik} - Y^*_{t_ik})\right|}{|\mathbb{E}K_0|} + \sup_{\boldsymbol{x}\in\chi}\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\mathbb{E}(K_{t_i}\mathbb{E}(\mu_k(\boldsymbol{X}_{t_i}) + \eta_{t_ik} - \mu_k(\boldsymbol{x})\mid\mathcal{F}_{t_i}))}{\mathbb{E}K_0}\right| \\
&\leqslant \frac{C_2}{C_1}\sup_{t_i}\mathbb{E}\left|\widehat{Y}^*_{t_ik} - Y^*_{t_ik}\right| + \frac{C_2}{C_1}\sup_{\boldsymbol{x}\in\chi}\sup_{y\in B(\boldsymbol{x},h_n\lambda)}|\mu_k(\boldsymbol{y}) - \mu_k(\boldsymbol{x})|,
\end{aligned}
$$

where the last line follows from the fact that $\mathbb{E}(\eta_{t_ik}\mid\mathcal{F}_{t_i}) = 0$, and the kernel function has bounded support on $[0,\lambda]$ (cf. (11)). Here, $B(\boldsymbol{x}, h_n\lambda)$ denotes the infinite-dimensional hyper-ellipsoid centered at x with axis lengths $2h_n\eta_j\lambda$ in the jth direction. Note that, for the second term, the supremum can also be taken over the choice of any $\boldsymbol{x}$ and $\boldsymbol{y}\in\chi$ such that $\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{y})\| \leqslant \lambda$.

The first term of the above bound is $\mathcal{O}(\delta_k(\epsilon^*_{t_i}))$ where $\delta_k(\epsilon^*_{t_i})$ is as given in Proposition 2. This is asymptotically negligible since $\sup_{t_i}\epsilon^*_{t_i} \to 0$ and that the basis functions $b_k(s)$ is continuous over the spatial horizon $\mathcal{S}$.

For the second term, let us define the notation $K_{\boldsymbol{x}}(\boldsymbol{y}) = \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{y}))$. Then using the triangle inequality and the symmetry of the kernel function, we get that

$$
\begin{aligned}
&|\mu_k(\boldsymbol{x}) - \mu_k(\boldsymbol{y})| \\
&= \frac{1}{K(0) + K_{\boldsymbol{x}}(\boldsymbol{y})}|(K(0) + K_{\boldsymbol{x}}(\boldsymbol{y}))(\mu_k(\boldsymbol{x}) - \mu_k(\boldsymbol{y}))| \\
&\leqslant \frac{1}{K(0) + K_{\boldsymbol{x}}(\boldsymbol{y})}\left(|K_{\boldsymbol{x}}(\boldsymbol{x})\mu_k(\boldsymbol{x}) - K_{\boldsymbol{x}}(\boldsymbol{y})\mu_k(\boldsymbol{y})| + |K_{\boldsymbol{y}}(\boldsymbol{x})\mu_k(\boldsymbol{x}) - K_{\boldsymbol{y}}(\boldsymbol{y})\mu_k(\boldsymbol{y})|\right) \\
&\leqslant \frac{1}{K(0) + C_1}\mathcal{O}(\|\boldsymbol{D}^{-1}(\boldsymbol{x} - \boldsymbol{y})\|),
\end{aligned}
$$

where in the last line, we make use of the Lipschitz property given in Assumption (A4) and the boundedness of the type-I kernel. Since we are considering the supremum when $\|\boldsymbol{H}_n^{-1}(\boldsymbol{x} - \boldsymbol{y})\| < \lambda$, it follows that the above quantity is $\mathcal{O}(h_n)$, which is asymptotically negligible as $n \to 0$, due to the choice of bandwidth as in Assumption (A3). $\qquad\square$

Lemma 3. *Suppose that the Assumptions (A1)-(A4) hold. Additionally, the effective spatial resolution $\epsilon^*_{t_i}$ decays to 0 uniformly over $t_i \in \mathcal{T}$. Fix any $k \in \mathbb{Z}^+$ and $\boldsymbol{x} \in \chi$. Then as $n \to \infty$,*

$$
\mathbb{E}|\widehat{\mu}_{k,2}(\boldsymbol{x}) - \mathbb{E}(\widehat{\mu}_{k,2}(\boldsymbol{x}))| \to 0.
$$

Proof. We start by noting that,

$$
|\widehat{\mu}_{k,2}(\boldsymbol{x}) - \mathbb{E}(\widehat{\mu}_{k,2}(\boldsymbol{x}))| \leqslant \frac{1}{\mathbb{E}K_0}\left|\frac{1}{n}\sum_{i=1}^{n}K_{t_i}\widehat{Y}^*_{t_ik} - \mathbb{E}(K_{t_i}\widehat{Y}^*_{t_ik})\right| = \frac{1}{\mathbb{E}K_0}\left|\frac{1}{n}\sum_{i=1}^{n}U_i\right|. \tag{S.6}
$$

where

$$
U_i = K_{t_i}\widehat{Y}^*_{t_ik} - \mathbb{E}\left(K_{t_i}\widehat{Y}^*_{t_ik}\right).
$$

Note that, $\mathbb{E}(U_i) = 0$ for each $i = 1, 2, \ldots, n$. Although $\{U_i\}_{i=1}^{n}$ are dependent, if we can show a result similar to the law of large numbers for the average of these U_is, then the result will follow. We will use Theorem 1 by Andrews (1988) to show this. In view of this, it suffices to show that there exists an appropriate sequence of real numbers $\{\psi(m)\}_{m=1}^{\infty}$ and a large

8

absolute constant C_0 such that

1. $\mathbb{E}|\mathbb{E}(U_i \mid \mathcal{F}_{t_i-m})| < C_0\psi(m)$ with $\psi(m) \to 0$ as $m \to \infty$.

2. $\mathbb{E}|U_i - \mathbb{E}(U_i \mid \mathcal{F}_{t_i+m})| < C_0\psi(m+1)$ for $m \geq 0$.

3. U_is are uniformly integrable.

Beginning with the first condition, we have

$$
\begin{aligned}
|\mathbb{E}(U_i \mid \mathcal{F}_{t_i-m})| &= \left|\mathbb{E}\left(K_{t_i}\widehat{Y}^*_{t_ik} - \mathbb{E}(K_{t_i}\widehat{Y}^*_{t_ik}) \mid \mathcal{F}_{t_i-m}\right)\right| \\
&= o_{\mathbb{P}}(1) + \left|\mathbb{E}\left(K_{t_i}Y^*_{t_ik} - \mathbb{E}(K_{t_i}Y^*_{t_ik}) \mid \mathcal{F}_{i-m}\right)\right|, \text{as } \sup_{t_i}\mathbb{E}|\widehat{Y}^*_{t_ik} - Y^*_{t_ik}| = o_{\mathbb{P}}(1) \\
&\leqslant o_{\mathbb{P}}(1) + |\mathbb{E}\left(K_{t_i}\mu_k(\boldsymbol{X}_{t_i}) - \mathbb{E}(K_{t_i}\mu_k(\boldsymbol{X}_{t_i})) \mid \mathcal{F}_{i-m}\right)|, \text{ as } \mathbb{E}(\eta_{t_i,k} \mid \mathcal{F}_{t_i-m}) = 0, \\
&\leqslant o_{\mathbb{P}}(1) + Cm^{-\tau},
\end{aligned}
$$

where the last line follows from the fact that the function $f(\boldsymbol{y}) = \mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{y} - \boldsymbol{x}))\mu_k(\boldsymbol{y})$ is Lipschitz due to Assumption (A4), and a direct application of Proposition 1 in addition of the PMC condition (4). Since $\tau > 0$, $\mathbb{E}|\mathbb{E}(Z_i \mid \mathcal{F}_{t_i-m})| < c\psi(m)$ holds with $\psi(m) = m^{-\tau}$ and $C_0 = 2C$.

For the second condition, note that $\boldsymbol{X}_{t_i}$ is a causal process with respect to the filtration $\{\mathcal{F}_{t_i}\}_{i=1}^{\infty}$. Since U_i is a linear combination of only $\{Y_{t_i}(s_{t_ij})\}_{j=1}^{n_i}$ with the weights being a function of $\boldsymbol{X}_{t_i}$, it is known almost surely given the history $\mathcal{F}_{t_i+m}$ for any $m \geq 0$. Hence, $\mathbb{E}(U_i \mid \mathcal{F}_{t_i+m}) = U_i$, and the second condition holds.

For the final condition regarding the uniform integrability of U_i, we consider the decomposition

$$
\begin{aligned}
\sup_i \mathbb{E}|U_i| &\leqslant \sup_{t_i} 2\mathbb{E}|K(t_i)\widehat{Y}^*_{t_ik}|, \\
&= o_{\mathbb{P}}(1) + \sup_{t_i} 2C_2\mathbb{E}\left|\mu_k(\boldsymbol{X}_{t_i}) + \int_{\mathcal{S}}\sigma(\boldsymbol{X}_{t_i}, s)\epsilon_{t_i}(s)b_k(s)ds\right| &\text{(S.7)} \\
&\leqslant o_{\mathbb{P}}(1) + 2C_2\sup_{t_i}\left(\int_{\mathcal{S}}\mathbb{E}(|\mu(\boldsymbol{X}_{t_i}, s)|)b_k(s)ds + \int_{\mathcal{S}}\mathbb{E}\left(\sigma(\boldsymbol{X}_{t_i}, s)\mathbb{E}(|\epsilon_{t_i}(s)|)\right)b_k(s)ds\right) \\
&= o_{\mathbb{P}}(1) + 2C_2\left(\int_{\mathcal{S}}\mathbb{E}(|\mu(\boldsymbol{X}_0, s)|)b_k(s)ds + M\int_{\mathcal{S}}\mathbb{E}\left(\sigma(\boldsymbol{X}_0, s)\right)b_k(s)ds\right), &\text{(S.8)} \\
&\leqslant o_{\mathbb{P}}(1) + 2C_2\left(M'\int_{\mathcal{S}}\widetilde{\mu}(s)ds + M\left(\int_{\mathcal{S}}\widetilde{\sigma}^2(s)ds\right)^{1/2}\left(\int_{\mathcal{S}}b_k^2(s)ds\right)^{1/2}\right) &\text{(S.9)}
\end{aligned}
$$

where M and M' are some generic constants. The equation (S.7) follows from the boundedness of the type-I kernel and Proposition 2, equation (S.8) is a consequence of the stationarity of $\boldsymbol{X}_t$ and boundedness of expected error given in Assumption (A2). The last inequality (S.9) is an application of the Cauchy-Schwarz inequality. Here, we use the notation $\widetilde{\mu}(s) = \mathbb{E}(\mu(\boldsymbol{X}_0, s))$ and $\widetilde{\sigma}^2(s) = \mathbb{E}(\sigma^2(\boldsymbol{X}_0, s))$ as in Assumption (A1). Because of Assumption (A1), both the corresponding integrals are finite, and $\int_{\mathcal{S}}b_k^2(s)ds = 1$ due to the orthonormality of the basis function. Hence $\{U_i : 1 \leqslant i \leqslant n\}$ is uniformly integrable.

Now an application of Theorem 1 by Andrews (1988) implies that $|n^{-1}\sum_{i=1}^n U_i|$ converges to 0 in L_1. We take expectation on both sides of the inequality (S.6). $\qquad\square$

9

S.2.5   Proof of Theorem 2

Fix $\boldsymbol{x} \in \chi$ and any $\epsilon > 0$. As $\mu(\boldsymbol{x}, s)$ has a decomposition as in (6), we have $\sum_{k=1}^{\infty} |\mu_k(\boldsymbol{x})| < \infty$, where $\mu(\boldsymbol{x}, s) = \sum_{k=1}^{\infty} \mu_k(\boldsymbol{x}) b_k(s)$.

Consistency: For the first part of the proof, let us fix $s \in \mathcal{S}$ as well. It is, therefore, possible to find K_ϵ such that the tail series $\sum_{k=K_\epsilon+1}^{\infty} |\mu_k(\boldsymbol{x})| < \epsilon/2b_\infty(s)$, as $b_\infty(s) > 0$ by definition. Hence,

$$\left| \sum_{k=K_\epsilon+1}^{\infty} \mu_k(\boldsymbol{x}) b_k(s) \right| \leqslant \sum_{k=K_\epsilon+1}^{\infty} |\mu_k(\boldsymbol{x})| \, |b_\infty(s)| < \frac{\epsilon}{2}.$$

For the first K_ϵ terms, due to Theorem 1, it is possible to choose sufficiently large n such that for each $k = 1, 2, \ldots, K_\epsilon$, the difference satisfies $(\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x})) = o_{\mathbb{P}}(\epsilon/2K_\epsilon b_\infty(s))$. Therefore,

$$
\begin{aligned}
\left| \sum_{k=1}^{K_\epsilon} \widehat{\mu}_k(\boldsymbol{x}) b_k(s) - \mu(\boldsymbol{x}, s) \right| &= \left| \sum_{k=1}^{K_\epsilon} \widehat{\mu}_k(\boldsymbol{x}) b_k(s) - \sum_{k=1}^{\infty} \mu_k(\boldsymbol{x}) b_k(s) \right| \\
&\leqslant \sum_{k=1}^{K_\epsilon} |\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x})| \, |b_\infty(s)| + \sum_{k=1}^{K_\epsilon+1} |\mu_k(\boldsymbol{x})||b_\infty(s)| \\
&\leqslant \sum_{k=1}^{K_\epsilon} o_{\mathbb{P}}\left(\frac{\epsilon}{2K_\epsilon b_\infty(s)} \right) |b_\infty(s)| + \frac{\epsilon}{2b_\infty(s)} |b_\infty(s)| = o_{\mathbb{P}}(\epsilon).
\end{aligned}
$$

Uniform Consistency: Now we move over to the second part of the proof. As $b_\infty(s)$ is continuous on the compact set $\mathcal{S}$, there exists a constant $B_\infty \in (0, \infty)$ such that the supremum $\sup_{s \in \mathcal{S}} |b_\infty(s)| < B_\infty$. As a result, we have the decomposition

$$\left| \sum_{k=1}^{K_\epsilon} \widehat{\mu}_k(\boldsymbol{x}) b_k(s) - \mu(\boldsymbol{x}, s) \right| \leqslant B_\infty \sum_{k=1}^{K_\epsilon} |\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x})| + B_\infty \sum_{k=K_\epsilon+1}^{\infty} |\mu_k(\boldsymbol{x})|.$$

Since the right-hand side is free of s, the inequality still holds true when the left-hand side of the inequality is replaced by its supremum over $s \in \mathcal{S}$. Finally, we choose K_ϵ such that $\sum_{k=1}^{K_\epsilon+1} |\mu_k(\boldsymbol{x})| < \epsilon/2B_\infty$ and choose sufficiently large n such that $(\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x})) = o_{\mathbb{P}}(\epsilon/2K_\epsilon B_\infty)$ for all $k = 1, 2, \ldots, K_\epsilon$. The proof for uniform consistency now follows the same steps as in the previous part.

S.2.6   Proof of Theorem 3

Following the same approach as in Masry (2005) and Hong & Linton (2020), we decompose the centered estimate of the mean components into

$$\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}) - \tilde{b}_{nk}(\boldsymbol{x}) = \frac{Q_{nk}(\boldsymbol{x}) - \tilde{b}_{nk}(\boldsymbol{x})(\widehat{\mu}_{k,1}(\boldsymbol{x}) - \mathbb{E}(\widehat{\mu}_{k,1}(\boldsymbol{x})))}{\widehat{\mu}_{k,1}(\boldsymbol{x})}$$

where

$$\tilde{b}_{nk}(\boldsymbol{x}) = \frac{\mathbb{E}(\widehat{\mu}_{k,2}(\boldsymbol{x})) - \mu_k(\boldsymbol{x})\mathbb{E}(\widehat{\mu}_{k,1}(\boldsymbol{x}))}{\mathbb{E}(\widehat{\mu}_{k,1}(\boldsymbol{x}))} = \mathbb{E}(\widehat{\mu}_{k,2}(\boldsymbol{x})) - \mu_k(\boldsymbol{x})$$

$$Q_{nk}(\boldsymbol{x}) = (\widehat{\mu}_{k,2}(\boldsymbol{x}) - \mathbb{E}(\widehat{\mu}_{k,2}(\boldsymbol{x}))) - \mu_k(\boldsymbol{x})(\widehat{\mu}_{k,1}(\boldsymbol{x}) - \mathbb{E}(\widehat{\mu}_{k,1}(\boldsymbol{x})))$$

First, we consider the bias term $\tilde{b}_{nk}$. In Lemma 2, we have already established that under suitable assumptions it is asymptotically negligible as $n \to \infty$, hence it is $o_{\mathbb{P}}(1)$. Also, we already know by Lemma 1 that $\widehat{\mu}_{k,1}(\boldsymbol{x})$ converges in probability to $\mathbb{E}(\widehat{\mu}_{k,1}(\boldsymbol{x})) = 1$. Therefore, we can rewrite the above decomposition as

$$\widehat{\mu}_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}) - \tilde{b}_{nk}(\boldsymbol{x}) = \frac{Q_{nk}(\boldsymbol{x})}{\widehat{\mu}_{k,1}(\boldsymbol{x})}(1 + o_{\mathbb{P}}(1)).$$

Hence, the interesting quantity is $Q_{n,k}(\boldsymbol{x})$, whose asymptotic distribution governs the asymptotic limit of the mean estimator. Therefore, we consider the normalized sum

$$\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}Q_{nk}(\boldsymbol{x}) = \frac{\sqrt{\phi_{\boldsymbol{x}}(h_n\lambda)}}{\sqrt{n}\mathbb{E}(K_0)} \sum_{i=1}^{n} \left[K_{t_i}\widehat{Y}_{t_i,k}^* - \mu_k(\boldsymbol{x})K_{t_i} - \mathbb{E}\left(K_{t_i}\widehat{Y}_{t_i,k}^* - \mu_k(\boldsymbol{x})K_{t_i} \right) \right]$$

$$= \tilde{v}_n \sum_{i=1}^{n} U_i + \tilde{v}_n \sum_{i=1}^{n} V_i$$

where

$$U_i = K_{t_i}\left(Y_{t_i,k}^* - \mu_k(\boldsymbol{x}) \right) - \mathbb{E}\left(K_{t_i}\left(Y_{t_ik} - \mu_k(\boldsymbol{x}) \right) \right)$$

$$V_i = K_{t_i}\left(\widehat{Y}_{t_ik}^* - Y_{t_ik}^* \right) - \mathbb{E}\left(K_{t_i}\left(\widehat{Y}_{t_ik}^* - Y_{t_ik}^* \right) \right),$$

$$\text{and, } \tilde{v}_n = n^{-1/2}\phi_{\boldsymbol{x}}^{1/2}(h_n\lambda)/\mathbb{E}(K_0).$$

Note that, since $(\phi_{\boldsymbol{x}}^{-1}(h_n\lambda))\mathbb{E}(K_0) \to \xi_1$ where $\xi_1 \in (0,\infty)$ as $n \to \infty$, so $\tilde{v}_n = (n\phi_{\boldsymbol{x}}(h_n\lambda))^{-1/2}\xi_1^{-1} + o_{\mathbb{P}}(1)$. As shown in (14), we have $n\phi_{\boldsymbol{x}}(h_n\lambda) \to \infty$, hence $\tilde{v}_n \to 0$.

Now, by Lemma 2 and the boundedness of the kernel as in (11), it follows that each $V_i = \mathcal{O}(\delta_k(\epsilon_{t_i}^*))$. Therefore,

$$\tilde{v}_n \sum_{i=1}^{n} V_i = \mathcal{O}\left(\frac{n\sup_{t_i}\delta_k(\epsilon_{t_i}^*)}{\xi_1\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}} \right) = \mathcal{O}\left(\sup_{t_i}\delta_k(\epsilon_{t_i}^*)\sqrt{\frac{n}{\phi_{\boldsymbol{x}}(h_n\lambda)}} \right) = o_{\mathbb{P}}(1)$$

where the last line follows from Assumption (A5). Also, since $\widehat{\mu}_{k,1} \xrightarrow{P} 1$ due to Lemma 1, we have

$$\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}\left(\widehat{\mu}_k(x) - \mu_k(x) - \tilde{b}_{nk}(x) \right) = \tilde{v}_n \sum_{i=1}^{n} U_i \left(\frac{1 + o(1)}{1 - o(1)} \right). \tag{S.10}$$

Therefore, we simply have to restrict our attention to the scaled sum $\tilde{v}_n \sum_{i=1}^{n} U_i$ alone.

Our proof for establishing the asymptotic distribution of the scaled sum $\tilde{v}_n \sum_{i=1}^{n} U_i$ follows similar to the big-block small-block approach as in Masry (2005), but with some modifications

11

to incorporate the PMC-type condition instead of the strong mixing conditions. We begin by picking integers a_n, b_n such that $a_n = \mathcal{O}(n^\beta)$ and $b_n = \mathcal{O}(n^\alpha)$ where α and β are as asserted by Assumption (A5). Let $c_n = (a_n + b_n)$ and $g_n = [n/c_n]$, where $[x]$ denote the largest integer less than or equal to x. Then the sum $\tilde{v}_n \sum_{i=1}^{n} U_i$ can be further decomposed as

$$\tilde{v}_n \sum_{i=1}^{n} U_i = \tilde{v}_n \sum_{j=1}^{g_n} \sum_{i=(j-1)c_n+1}^{(j-1)c_n+a_n} U_i + \tilde{v}_n \sum_{j=1}^{g_n} \sum_{i=(j-1)c_n+(a_n+1)}^{jc_n} U_i + \tilde{v}_n \sum_{i=g_nc_n}^{n} U_i =: S_n + B_n + R_n, \quad \text{(S.11)}$$

where S_n, B_n and R_n are the quantities it is replacing. Note that, since $a_n/b_n \to 0$ as $n \to \infty$, it follows that $b_n/c_n \to 1$ and hence g_n is in the asymptotic order of $n^{1-\alpha}$. Now, to show the proof of the asymptotic normality, the idea is to show that the contribution from the small block sum S_n and the remainder R_n are asymptotically negligible, and the summands of the big-block sum are asymptotically independent, which we will show through the multiplicative decomposition of the characteristic functions of the appropriate random variables. These ideas are presented in precise mathematical terms through the following series of lemmas.

Lemma 4. *Suppose the Assumptions (A1)-(A5) hold. Then as $n \to \infty$, $S_n \xrightarrow{P} 0$, where S_n is as defined in* (S.11).

Proof. Clearly, $\mathbb{E}(S_n) = 0$. In view of Chebyshev's inequality, it is, therefore, enough to show that the variance of the small block sum goes to 0 as $n \to \infty$. To this direction, we have

$$\begin{aligned}
\text{Var}(S_n) &= \tilde{v}_n^2 \text{Var} \left(\sum_{j=1}^{g_n} \sum_{i=(j-1)c_n+1}^{(j-1)c_n+a_n} U_i \right) \\
&= \tilde{v}_n^2 \sum_{j=1}^{g_n} \text{Var} \left(\sum_{i=(j-1)c_n+1}^{(j-1)c_n+a_n} U_i \right) + \tilde{v}_n^2 \sum_{j\neq l}^{g_n} \text{Cov} \left(\sum_{i=(j-1)c_n+1}^{(j-1)c_n+a_n} U_i, \sum_{i=(l-1)c_n+1}^{(l-1)c_n+a_n} U_i \right) \\
&= \tilde{v}_n^2 \sum_{j=1}^{g_n} \left[\sum_{i=(j-1)c_n+1}^{(j-1)c_n+a_n} \text{Var}(U_i) + \sum_{i\neq i'} \text{Cov}(U_i, U_{i'}) \right] + \tilde{v}_n^2 \sum_{j\neq l}^{g_n} \sum_{i\neq i'} \text{Cov}(U_i, U_i')
\end{aligned}$$

where the sums over $i \neq i'$ are taken over the appropriate groups. For the variance term, we have

$$\begin{aligned}
\text{Var}(U_i) &= \text{Var} \left(K_{t_i}(Y_{t_ik}^* - \mu_k(\boldsymbol{x})) \right) \\
&= \mathbb{E} \left[\text{Var} \left(K_{t_i}(Y_{t_ik}^* - \mu_k(\boldsymbol{x})) \mid \mathcal{F}_{t_i} \right) \right] + \text{Var} \left(\mathbb{E} \left[K_{t_i}(Y_{t_ik}^* - \mu_k(\boldsymbol{x})) \mid \mathcal{F}_{t_i} \right] \right) \\
&= \mathbb{E} \left[K_{t_i}^2 \int_{\mathcal{S}^2} \sigma(\boldsymbol{X}_{t_i}, s)\sigma(\boldsymbol{X}_{t_i}, s')\rho(s, s')b_k(s)b_k(s')dsds' \right] + \mathbb{E} \left[K_{t_i}^2(\mu_k(\boldsymbol{X}_{t_i}) - \mu_k(\boldsymbol{x}))^2 \right] \\
&= \mathbb{E} \left[K_0^2 \int_{\mathcal{S}^2} \sigma(\boldsymbol{X}_0, s)\sigma(\boldsymbol{X}_0, s')\rho(s, s')b_k(s)b_k(s')dsds' \right] + \mathbb{E} \left[K_0^2(\mu_k(\boldsymbol{X}_0) - \mu_k(\boldsymbol{x}))^2 \right]
\end{aligned}$$

The last line follows from stationarity of $\boldsymbol{X}_t$. We bound the second term $\mathbb{E}\left[K_0^2(\mu_k(\boldsymbol{X}_0) - \mu_k(\boldsymbol{x}))^2 \right]$ by noting that, the random variable inside the expectation is nonzero only when $\boldsymbol{X}_0$ lies in the infinite-dimensional ball $B(\boldsymbol{x}, h_n\lambda)$. Additionally, when $\boldsymbol{X}_0$ is indeed inside the ball $B(\boldsymbol{x}, h_n\lambda)$, we can apply the Lipschitz condition presented in Assumption (A4) to bound the

12

term. In summary, we get

$$\mathbb{E}\left[K_0^2(\mu_k(\boldsymbol{X}_0) - \mu_k(\boldsymbol{x}))^2\right] \leqslant \mathbb{E}\left[\mathbf{1}_{\{\boldsymbol{X}_0 \in B(\boldsymbol{x}, h_n\lambda)\}}\left(\mathcal{O}\left(\|\boldsymbol{D}^{-1}(\boldsymbol{X}_0 - \boldsymbol{x})\|\right) + (K_0 - K(0))^2\mu_k^2(x)\right)\right]$$
$$\leqslant \mathcal{O}(h_n\lambda) + \mathcal{O}(\phi_{\boldsymbol{x}}(h_n\lambda)) \to 0,$$

as $n \to \infty$. To deal with the first term, note that again it is enough to consider the situation when $\boldsymbol{X}_0$ is within the ball $B(\boldsymbol{x}, h_n\lambda)$. By a similar logic as above and using the Lipschitz continuity of $\mathcal{K}(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_0 - \boldsymbol{x})\|)\sigma_k(\boldsymbol{X}_0)$ as in Assumption (A4), we obtain that $K_0\sigma(\boldsymbol{X}_0, s) = K_0\sigma(\boldsymbol{x}, s) + o_{\mathbb{P}}(1)$ for any fixed $s \in \mathcal{S}$. Since $\mathcal{S}$ is compact, this convergence may be strengthened to a uniform convergence as $\sigma(\boldsymbol{x}, s)$ is continuous over $s \in \mathcal{S}$ due to continuity of the basis functions. Therefore,

$$\int_{\mathcal{S}^2} K_0^2 \left|\sigma(\boldsymbol{X}_0, s)\sigma(\boldsymbol{X}_0, s') - \sigma(\boldsymbol{x}, s)\sigma(\boldsymbol{x}, s')\right| |\rho(s, s')|b_k(s)b_k(s')dsds'$$
$$\leqslant \int_{\mathcal{S}^2} K_0^2 \left(|\sigma(\boldsymbol{X}_0, s)||\sigma(\boldsymbol{X}_0, s') - \sigma(\boldsymbol{x}, s')| + |\sigma(\boldsymbol{x}, s')||\sigma(\boldsymbol{X}_0, s) - \sigma(\boldsymbol{x}, s)|\right) |\rho(s, s')|b_k(s)b_k(s')dsds'$$
$$\leqslant o_{\mathbb{P}}(1)\left[\int_{\mathcal{S}^2} |\sigma(\boldsymbol{X}_0, s)|b_k(s)b_k(s')dsds' + \int_{\mathcal{S}^2} |\sigma(\boldsymbol{x}, s')|b_k(s)b_k(s')dsds'\right] \tag{S.12}$$
$$= o_{\mathbb{P}}(1)\left[\left(\int_{\mathcal{S}} \sigma^2(\boldsymbol{X}_0, s)ds\right)^{1/2}\int_{\mathcal{S}} b_k(s)ds + \left(\int_{\mathcal{S}} \sigma^2(\boldsymbol{x}, s')ds'\right)^{1/2}\int_{\mathcal{S}} b_k(s)ds\right] \tag{S.13}$$

The inequality (S.12) in the third step follows from noting that $|\rho(s, s')| \leqslant 1$. The inequality (S.13) in the fourth step is a simple application of Cauchy-Schwartz inequality and the fact that $\int_{\mathcal{S}} b_k^2(s)ds = 1$. Finally, by taking expectation on both sides, and using Assumption (A1) on the integrability of $\mathbb{E}(\sigma_k(\boldsymbol{X}_0, s))$ as a function of s, we note that the above quantity is also asymptotically negligible. Now, we use the relationship given in (15) to get

$$\mathbb{E}\left[K_0^2 \int_{\mathcal{S}^2} \sigma(\boldsymbol{X}_0, s)\sigma(\boldsymbol{X}_0, s')\rho(s, s')b_k(s)b_k(s')dsds'\right] = \phi_x(h_n\lambda)\xi_2\sigma_{kk}(\boldsymbol{x}) + o_{\mathbb{P}}(1),$$

where $\sigma_{k,k}(\boldsymbol{x})$ is as defined in (28). Together, we have $\text{Var}(U_i) = \phi_x(h_n\lambda)\xi_2\sigma_{kk}(x) + o_{\mathbb{P}}(1)$.

For the covariance term, with $i < i'$, analogus to the above, we obtain

$$\text{Cov}(U_i, U_{i'}) = \text{Cov}\left(K_{t_i}(Y_{t_ik}^* - \mu_k(\boldsymbol{x})), K_{t_{i'}}(Y_{t_{i'}k}^* - \mu_k(\boldsymbol{x}))\right)$$
$$= \mathbb{E}\left[K_{t_i}K_{t_{i'}} \int_{\mathcal{S}^2} \sigma(\boldsymbol{X}_{t_i}, s)\sigma(\boldsymbol{X}_{t_{i'}}, s')\rho(s, s')b_k(s)b_k(s')dsds'\right]$$
$$+ \mathbb{E}\left[K_{t_i}K_{t_{i'}}(\mu_k(\boldsymbol{X}_{t_i}) - \mu_k(\boldsymbol{x}))(\mu_k(\boldsymbol{X}_{t_{i'}}) - \mu_k(\boldsymbol{x}))\right]$$

Note that, both of the terms are nonzero only when both $\boldsymbol{X}_{t_i}$ and $\boldsymbol{X}_{t_{i'}}$ lie in the infinite-dimensional ball $B(\boldsymbol{x}, h_n\lambda)$. By applying the Lipschitz conditions as before and using the bounds on the joint small ball probabilities described in Section 2.3, it is easy to obtain that

$$\mathbb{E}\left[K_{t_i}K_{t_{i'}}(\mu_k(\boldsymbol{X}_{t_i}) - \mu_k(\boldsymbol{x}))(\mu_k(\boldsymbol{X}_{t_{i'}}) - \mu_k(\boldsymbol{x}))\right] = \mathcal{O}(\phi_{\boldsymbol{x}}^2(h_n\lambda)) \to 0,$$
$$\mathbb{E}\left[K_{t_i}K_{t_{i'}} \int_{\mathcal{S}^2} \sigma(\boldsymbol{X}_{t_i}, s)\sigma(\boldsymbol{X}_{t_{i'}}, s')\rho(s, s')b_k(s)b_k(s')dsds'\right] = \mathbb{E}(K_{t_i}K_{t_{i'}})\sigma_{k,k}(\boldsymbol{x}) + o_{\mathbb{P}}(1)$$
$$= \mathcal{O}(\phi_{\boldsymbol{x}}^2(h_n\lambda)) \to 0,$$

under the asymptotic regime considered in the paper. As a result, we have $\text{Cov}(U_i, U_{i'}) \to 0$ as $n \to \infty$ for any $i \neq i'$. Putting everything back together, we obtain that

$$
\begin{aligned}
\text{Var}(S_n) &= \tilde{v}_n^2 \phi_{\boldsymbol{x}}(h_n\lambda)\xi_2 g_n a_n \sigma_{kk}(\boldsymbol{x}) + o(\tilde{v}_n^2 g_n a_n^2 \phi_{\boldsymbol{x}}^2(h_n\lambda)) + \mathcal{O}\left(\tilde{v}_n^2 g_n^2 a_n^2 \phi_{\boldsymbol{x}}^2(h_n\lambda)\right) \\
&= \frac{n^\beta \xi_2}{n^\alpha \xi_1^2} \sigma_{kk}(\boldsymbol{x}) + o\left(\frac{n^{2\beta}\phi_{\boldsymbol{x}}(h_n\lambda)}{n^\alpha \xi_1^2}\right) + \mathcal{O}\left(\frac{n^{2\beta}n^{1-2\alpha}\phi_{\boldsymbol{x}}(h_n\lambda)}{\xi_1^2}\right) \\
&= \frac{n^{\beta-\alpha}\xi_2}{\xi_1^2}\sigma_{kk}(\boldsymbol{x}) + o\left(\frac{n^{2\beta-\alpha}\phi_{\boldsymbol{x}}(h_n\lambda)}{\xi_1^2}\right) + \mathcal{O}\left(\frac{n^{1-2\alpha+2\beta}\phi_{\boldsymbol{x}}(h_n\lambda)}{\xi_1^2}\right)
\end{aligned}
$$

It is now immediate that all the terms are asymptotically negligible due to Assumption (A5) and since $\alpha < 1$. $\square$

Since the number of terms contained in R_n is even smaller than the number of terms in S_n, the contribution of the remainder term R_n is also asymptotically negligible as $n \to \infty$. Now, for the big-block sum B_n, we wish to show that the summands of B_n are asymptotically independent. Then, an application of the central limit theorem as in Masry (2005) can be applied to establish the asymptotic normality of the B_n. To show this asymptotic independence result, our objective will be to demonstrate that the characteristic function of B_n asymptotically splits into a product of characteristic functions of individual summands $\sum_{i=(j-1)c_n+a_n+1}^{jc_n} U_i$ (to be denoted as $B_{n,j}$ for notational convenience), adjusting for the proper normalization factor $\tilde{v}_n$. In view of this, we obtain the relationship $B_n = \tilde{v}_n \sum_{j=1}^{g_n} B_{n,j}$.

Lemma 5. *Suppose that Assumptions (A1)-(A5) hold. Then, as $n \to \infty$,*

$$
\left|\mathbb{E}(e^{itB_n}) - \prod_{j=1}^{g_n}\mathbb{E}(e^{it\tilde{v}_n B_{n,j}})\right| \to 0,
$$

where B_n is the big-block as in (S.11) and $B_{n,j} = \sum_{i=(j-1)c_n+a_n+1}^{jc_n} U_i$ for each $j = 1,\ldots,g_n$.

Proof. It follows that

$$
\left|\mathbb{E}(e^{itB_n}) - \prod_{j=1}^{g_n}\mathbb{E}(e^{it\tilde{v}_n B_{n,j}})\right|
$$

$$
= \left|\mathbb{E}(e^{itB_{g_n c_n}}) - \prod_{j=1}^{g_n}\mathbb{E}(e^{it\tilde{v}_n B_{n,j}})\right|
$$

$$
\leqslant \left|\mathbb{E}(e^{itB_{g_n c_n}}) - \mathbb{E}(e^{itB_{(g_n-1)c_n}})\mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}}\right)\right| + \left|\mathbb{E}(e^{itB_{(g_n-1)c_n}})\mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}}\right) - \prod_{j=1}^{g_n}\mathbb{E}(e^{it\tilde{v}_n B_{n,j}})\right|
$$

$$
= \left|\mathbb{E}(e^{itB_{(g_n-1)c_n}})\mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}} \mid \mathcal{F}_{t_{(g_n-1)c_n}}\right) - \mathbb{E}(e^{itB_{(g_n-1)c_n}})\mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}}\right)\right| +
$$

$$
\left|\mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}}\right)\right|\left|\mathbb{E}(e^{itB_{(g_n-1)c_n}}) - \prod_{j=1}^{g_n-1}\mathbb{E}(e^{it\tilde{v}_n B_{n,j}})\right|
$$

$$
\leqslant \left|\mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}} \mid \mathcal{F}_{t_{(g_n-1)c_n}}\right) - \mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}}\right)\right| + \left|\mathbb{E}(e^{itB_{(g_n-1)c_n}}) - \prod_{j=1}^{g_n-1}\mathbb{E}(e^{it\tilde{v}_n B_{n,j}})\right|
$$

14

where in the last inequality we use the fact that $|\mathbb{E}(e^{itX})| \leqslant \mathbb{E}(|e^{itX}|) = 1$ for any complex-valued random variable X. Note that, this creates an inductive inequality on the factorization provided, namely we provide a bound on the factorization error up to the term $g_n c_n$ using the factorization error till the term $(g_n - 1)c_n$, plus an additional term which takes care of the last term B_{n,g_n} based on a conditional expectation.

To deal with the first term of the right-hand side of the above inequality, consider the difference

$$
\left| \mathbb{E}\left(\cos\left(t\tilde{v}_n B_{n,g_n} \mid \mathcal{F}_{t_{(g_n-1)c_n}} \right) \right) - \mathbb{E}\left(\cos\left(t\tilde{v}_n B_{n,g_n} \right) \right) \right|
$$
$$
= b_n \sup_{(g_n-1)c_n+a_n+1 \leqslant i \leqslant g_n c_n} \left| \mathbb{E}(\cos(t\tilde{v}_n U_i) \mid \mathcal{F}_{t_{(g_n-1)c_n}}) - \mathbb{E}(\cos(t\tilde{v}_n U_i)) \right|. \quad \text{(S.14)}
$$

Now, $\cos(x)$ is Lipschitz with constant 1 and the function $f(y) = K(\|\boldsymbol{H}_n^{-1}(y-x)\|)\mu_k(y)$ is also Lipschitz due to the Assumption (A4). Also, $\mathbb{E}(\sigma(X_{t_i}, s)\epsilon_{t_i}(s)) = 0$ for all $s \in \mathcal{S}$ due to independence of ϵ_{t_i} and $\mathcal{F}_{t_i}$ and $\tilde{v}_n = o(1)$. By combining all of these along with Proposition 1, it follows that the difference in the cosine term as in (S.14) is $\mathcal{O}(\Delta_2(a_n))$ for all $t \in [-M', M']$ for some arbitrarily large but fixed constant $M' > 0$. A similar derivation holds for bounding the discrepancy through the sine function as well, i.e.,

$$
\left| \mathbb{E}\left(\sin\left(t\tilde{v}_n B_{n,g_n} \mid \mathcal{F}_{t_{(g_n-1)c_n}} \right) \right) - \mathbb{E}\left(\sin\left(t\tilde{v}_n B_{n,g_n} \right) \right) \right| = \mathcal{O}(\Delta_2(a_n)). \quad \text{(S.15)}
$$

In particular, by combining the inequalities from (S.14) and (S.15), we obtain that

$$
\left| \mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}} \mid \mathcal{F}_{t_{(g_n-1)c_n}} \right) - \mathbb{E}\left(e^{it\tilde{v}_n B_{n,g_n}} \right) \right| = \mathcal{O}(b_n \Delta_2(a_n)),
$$

for any $t \in [-M, M]$, some interval around 0. As a result, by repeatedly using this bound, we obtain

$$
\left| \mathbb{E}(e^{itB_n}) - \prod_{j=1}^{r} \mathbb{E}(e^{it\tilde{v}_n B_{n,j}}) \right| \leqslant \mathcal{O}(b_n \Delta_2(a_n)) + \left| \mathbb{E}(e^{itB_{(g_n-1)c_n}}) - \prod_{j=1}^{g_n-1} \mathbb{E}(e^{it\tilde{v}_n B_{n,j}}) \right| = \mathcal{O}(g_n b_n \Delta_2(a_n)).
$$

Since $g_n = [n/c_n]$ and $b_n/c_n \to 1$, we have $\mathcal{O}(g_n b_n \Delta_2(a_n)) = \mathcal{O}(na_n^{-\tau+1}) = \mathcal{O}(n^{1-\beta(\tau-1)})$. Again, due to Assumption (A5), this difference tends to 0 asymptotically as $n \to \infty$. $\qquad\square$

Lemma 5 now establishes that the summands of the big blocks are asymptotically independent as $n \to \infty$. As demonstrated before in the proof of Lemma 4, we have $\mathbb{E}(B_n) = 0$ and

$$
\text{Var}(B_n) = \tilde{v}_n^2 g_n b_n \text{Var}(U_i) + \tilde{v}_n^2 g_n b_n^2 \text{Cov}(U_i, U_{i'}) + o(1)
$$
$$
= \frac{\xi_2}{\xi_1^2} \sigma_{kk}(x) + \mathcal{O}\left(\frac{n^\alpha \phi_x(h_n \lambda)}{\xi_1^2} \right)
$$

Note that, here we use the asymptotic independence between U_i and $U_{i'}$ for $i \neq i'$ established in the proof of Lemma 4 to calculate the covariance term. The remainder part can be made small using Assumption (A5). Therefore, an application of the Central Limit Theorem as

15

in Masry (2005) shows that the normalized sum of the big blocks, i.e., B_n is asymptotically normal with mean zero and variance $\xi_2 \xi_1^{-2} \sigma_{kk}(\boldsymbol{x})$.

Finally, combining this result with Lemma 4, we obtain the asymptotic distribution of the normalized estimate of the mean component, completing the proof.

S.2.7 Proof of Theorem 4

Let us fix $\boldsymbol{x} \in \chi$ and denote $K_{t_i} = K(\|\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{x})\|)$ and $w_i = K_{t_i}/\sum_i K_{t_i}$. Then we can decompose the estimate $\widehat{\sigma}_{kl}(\boldsymbol{x})$ as

$$
\begin{aligned}
\widehat{\sigma}_{kl}(\boldsymbol{x}) &= \sum_i w_i (\widehat{Y}_{t_i k}^* - \widehat{\mu}_k(\boldsymbol{x}))(\widehat{Y}_{t_i l}^* - \widehat{\mu}_l(\boldsymbol{x})) \\
&= \sum_i w_i \left(\widehat{Y}_{t_i k}^* - Y_{t_i k}^* - \widehat{\mu}_k(\boldsymbol{x}) + \mu_k(\boldsymbol{x}) + \eta_{t_i k} \right) \left(\widehat{Y}_{t_i l}^* - Y_{t_i l}^* - \widehat{\mu}_l(\boldsymbol{x}) + \mu_l(\boldsymbol{x}) + \eta_{t_i l} \right).
\end{aligned}
$$

First, we apply Proposition 2 and an application of Cauchy-Schwartz inequality to conclude that for any $k, l \in \mathbb{Z}^+$, we have

$$
\mathbb{E} \left(\sum_i w_i \left| \widehat{Y}_{t_i k}^* - Y_{t_i k}^* \right| \left| \widehat{Y}_{t_i, l}^* - Y_{t_i l}^* \right| \right) \leqslant \left(\sup_{t_i} \mathbb{E} \left| \widehat{Y}_{t_i k}^* - Y_{t_i k}^* \right|^2 \right)^{1/2} \left(\sup_{t_i} \mathbb{E} \left| \widehat{Y}_{t_i l}^* - Y_{t_i l}^* \right|^2 \right)^{1/2} = o_{\mathbb{P}}(1).
$$

Due to the asymptotic normality of the estimator of the mean as in Theorem 3, it follows that for any fixed $k \in \mathbb{Z}^+$ and fixed $\boldsymbol{x} \in \chi$,

$$
\mathbb{E}|\mu_k(x) - \widehat{\mu}_k(x)|^2 = \mathbb{E}^2 |\mu_k(x) - \widehat{\mu}_k(x)| + \mathrm{Var}(\widehat{\mu}_k(x)) = o_{\mathbb{P}}(1) + \mathcal{O}((n\phi_x(h_n\lambda))^{-1/2}) = o_{\mathbb{P}}(1),
$$

as $n \to \infty$, where the last equality follows from the Assumption (A3) on the choice of bandwidth. Additionally, Assumption (A2) implies that $\eta_{t_i k}$ is uniformly bounded over $t_i \in [0, 1]$ by a deterministic constant free of n. Therefore, as $n \to \infty$, we obtain

$$
\begin{aligned}
&\mathbb{E}|\widehat{\sigma}_{kl}(\boldsymbol{x}) - \sigma_{k,l}(\boldsymbol{x})| \\
&= o_{\mathbb{P}}(1) + \mathbb{E} \left| \sum_i w_i \eta_{t_i k} \eta_{t_i l} - \sigma_{kl}(\boldsymbol{x}) \right| \\
&= o_{\mathbb{P}}(1) + \frac{1}{\mathbb{E}(K_0)} \mathbb{E} \left| \frac{1}{n} \sum_i K_{t_i}(\eta_{t_i k} \eta_{t_i l} - \sigma_{kl}(\boldsymbol{x})) \right| \\
&\leqslant o_{\mathbb{P}}(1) + \frac{1}{\mathbb{E}(K_0)} \mathbb{E} \left| \sum_i \frac{K_{t_i}}{n}(\eta_{t_i k} \eta_{t_i l} - \mathbb{E}(\eta_{t_i k} \eta_{t_i l} \mid \mathcal{F}_{t_i})) \right| + \frac{1}{\mathbb{E}(K_0)} \mathbb{E} \left| \sum_i \frac{K_{t_i}}{n}(\mathbb{E}(\eta_{t_i k} \eta_{t_i l} \mid \mathcal{F}_{t_i}) - \sigma_{k,l}(\boldsymbol{x})) \right| \\
&\leqslant o_{\mathbb{P}}(1) + \frac{1}{\mathbb{E}(K_0)} \mathbb{E} \left| \frac{1}{n} \sum_i V_i \right| + \frac{1}{\mathbb{E}(K_0)} \mathbb{E} \left| \frac{1}{n} \sum_i K_{t_i}(\sigma_{kl}(\boldsymbol{X}_{t_i}) - \sigma_{kl}(\boldsymbol{x})) \right|,
\end{aligned}
$$

where

$$
V_i = K_{t_i} \eta_{t_i k} \eta_{t_i, l} - \mathbb{E}(K_{t_i} \eta_{t_i k} \eta_{t_i l} \mid \mathcal{F}_{t_i}),
$$

and we make use of the stationarity of $\boldsymbol{X}_t$ and the form of covariance given in (27). We deal with both of these terms separately.

Since $\mathbb{E}(V_i \mid \mathcal{F}_{t_i}) = 0$ for all $i = 1, \ldots, n$, we can apply the same technique shown in the proof of Lemma 3 based on the L^1-type convergence result by Andrews (1988), to show that the first term of the bound is $o_{\mathbb{P}}(1)$ as $n \to \infty$.

For the second term of the bound, choose any $\epsilon > 0$. By the existence of the basis decomposition as in (6), we note that there exists $U_{\epsilon,1}$ such that for the fixed $\boldsymbol{x} \in \chi$, $\sum_{v=1}^{U_{\epsilon,1}} |\sigma_v(\boldsymbol{x})| < \epsilon$. Additionally, because of the decomposition in (21), we must have some $U_{\epsilon,2}$ such that the tail series of the variance decomposition satisfy

$$\sum_{u,v:\max(u,v)>U_{\epsilon,2}} c_{u,v}(k,l)\mathbb{E}(\sigma_u(\boldsymbol{X}_0)\sigma_v(\boldsymbol{X}_0)) < \epsilon.$$

Therefore, by choosing $U_\epsilon = \max(U_{\epsilon,1}, U_{\epsilon,2})$ and using the boundedness of the kernel function as in (11), we can write the second term of the bound as,

$$\mathbb{E}\left|\frac{1}{n}\sum_i K_{t_i}(\sigma_{kl}(\boldsymbol{X}_{t_i}) - \sigma_{kl}(\boldsymbol{x}))\right|$$

$$= \mathbb{E}\left|\frac{1}{n}\sum_i K_{t_i}\sum_{u,v} c_{u,v}(k,l)(\sigma_u(\boldsymbol{X}_{t_i})\sigma_v(\boldsymbol{X}_{t_i}) - \sigma_u(\boldsymbol{x})\sigma_v(\boldsymbol{x}))\right|$$

$$= \mathcal{O}(\epsilon) + \mathbb{E}\left|\frac{1}{n}\sum_i K_{t_i}\sum_{u=1}^{U_\epsilon}\sum_{v=1}^{U_\epsilon} c_{u,v}(k,l)(\sigma_u(\boldsymbol{X}_{t_i})\sigma_v(\boldsymbol{X}_{t_i}) - \sigma_u(\boldsymbol{x})\sigma_v(\boldsymbol{x}))\right|$$

$$\leqslant \mathcal{O}(\epsilon) + \mathbb{E}\left|\frac{1}{n}\sum_{u=1}^{U_\epsilon}\sum_{v=1}^{U_\epsilon} c_{u,v}(k,l)\sum_{i=1}^n K_{t_i}\sigma_u(\boldsymbol{X}_{t_i})(\sigma_v(\boldsymbol{X}_{t_i}) - \sigma_v(\boldsymbol{x}))\right|$$

$$+ \mathbb{E}\left|\frac{1}{n}\sum_{u=1}^{U_\epsilon}\sum_{v=1}^{U_\epsilon} c_{u,v}(k,l)\sum_{i=1}^n K_{t_i}\sigma_v(\boldsymbol{x})(\sigma_u(\boldsymbol{X}_{t_i}) - \sigma_u(\boldsymbol{x}))\right|$$

$$= \mathcal{O}(\epsilon) + 2U_\epsilon^2 C(h_n\lambda),$$

for some constant $C \in (0, \infty)$. Here, we use the fact that $c_{u,v}(k,l) \leqslant 1$ and the function $K_{t_i}\sigma_u(\boldsymbol{x})$ is Lipschitz as assured by Assumption (A4). By choosing sufficiently large n (depending on ϵ), as the bandwidth $h_n \to \infty$, we can ensure that the second term $2U_\epsilon^2 C(h_n\lambda) \leqslant \epsilon$. This makes the entire different $\mathcal{O}(\epsilon)$. Since ϵ is arbitrary, this completes the proof.

S.2.8 Proof of Theorem 5

Due to the specific choice of the χ_n, for any $i = 1, 2, \ldots, n$, we have either $\|\boldsymbol{D}^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{x})\| > h_n\lambda$ or $\|\boldsymbol{D}^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{x}')\| > h_n\lambda$. As the type-I kernel function has bounded support, it follows that

$$\mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{x}))\mathcal{K}(\boldsymbol{H}_n^{-1}(\boldsymbol{X}_{t_i} - \boldsymbol{x}')) = 0,$$

for all $i = 1, 2, \ldots, n$. By the form of the estimates of the mean components as in (9), it follows that the covariance between $\widehat{\mu}_k(\boldsymbol{x})$ and $\widehat{\mu}_l(\boldsymbol{x}')$ is zero for any $k, l \in \mathbb{Z}^+$ and $\boldsymbol{x} \neq \boldsymbol{x}' \in \chi_n$. The same conclusion holds for the variance estimates $\widehat{\sigma}_{kl}(\boldsymbol{x})$ and $\widehat{\sigma}_{k'l'}(\boldsymbol{x}')$ as well for any $k, k', l, l' \in \mathbb{Z}^+$. Therefore, by using the same principle of extreme value theory as in Zhao & Wu (2008), we can obtain the asymptotic probability

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{x}\in\chi_n} \frac{\xi_1\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}}{\sqrt{\xi_2 Q_K(\boldsymbol{x},s)}} \left|\widehat{\mu}_{1:K}(\boldsymbol{x},s) - \mu_{1:K}(\boldsymbol{x},s) - b^*_{n,1:K}(\boldsymbol{x},s)\right| < B_{m_n}(z)\right) = e^{-2e^{-z}},$$

(S.16)

for any fixed $z > 0$, $K \in \mathbb{Z}^+$ and $s \in \mathcal{S}$. This produces a probabilistic bound for the sum of first K components of the mean function. Now, for the given $\delta > 0$, consider an δ-net $\mathcal{N}(\delta)$ of the set χ_n, i.e., for every $\boldsymbol{x} \in \chi_n$, there exists a point $\boldsymbol{x}_0 \in \mathcal{N}(\delta)$ such that $\|\boldsymbol{D}^{-1}(\boldsymbol{x} - \boldsymbol{x}_0)\| < \delta$. Since χ_n is bounded and δ is free of n, it follows that $|\mathcal{N}(\delta)|$ is finite and free of n. Also,

$$\left|\sum_{k=(K+1)}^{\infty} \mu_k(\boldsymbol{x})b_k(s)\right| \leqslant \left|\sum_{k=K+1}^{\infty} (\mu_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}_0))b_k(s)\right| + \left|\sum_{k=K+1}^{\infty} \mu_k(\boldsymbol{x}_0)b_k(s)\right|$$

$$\leqslant b_{\infty}(s) \sup_{\boldsymbol{x}\in N_{\delta}(\boldsymbol{x}_0)} \left|\sum_{k=K+1}^{\infty} (\mu_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}_0))\right| + b_{\infty}(s) \sum_{k=K+1}^{\infty} |\mu_k(\boldsymbol{x}_0)|.$$

Since $\sum_{k=1}^{\infty} \mu_k(\boldsymbol{x})$ is uniformly convergent for all $\boldsymbol{x} \in N_{\delta}(\boldsymbol{x}_0)$, it follows that for any given $\epsilon > 0$, there exists $K_{\epsilon}(\boldsymbol{x}_0)$ such that the above bound is less than $\epsilon b_{\infty}(s)$. Taking supremum over both sides yields,

$$\sup_{\boldsymbol{x}\in\chi_n} \left|\sum_{k=(K+1)}^{\infty} \mu_k(\boldsymbol{x})b_k(s)\right| \leqslant \sup_{\boldsymbol{x}_0\in\mathcal{N}(\delta)} b_{\infty}(s) \left[\sup_{\boldsymbol{x}\in N_{\delta}(\boldsymbol{x}_0)} \left|\sum_{k=K+1}^{\infty} (\mu_k(\boldsymbol{x}) - \mu_k(\boldsymbol{x}_0))\right| + \sum_{k=K+1}^{\infty} |\mu_k(\boldsymbol{x}_0)|\right].$$

Since $\mathcal{N}(\delta)$ has finitely many elements, taking $K_{\epsilon} = \max_{\boldsymbol{x}_0\in\mathcal{N}(\delta)} K_{\epsilon}(\boldsymbol{x}_0)$ we can ensure that the right-hand side of the above quantity is less than $\epsilon b_{\infty}(s)$.

For ease of notation, let us denote

$$\alpha_{n,K} = \frac{\xi_1\sqrt{n\phi_{\boldsymbol{x}}(h_n\lambda)}}{\sqrt{\xi_2 Q_K(\boldsymbol{x},s)}},$$

$$\beta_{n,K} = \widehat{\mu}_{1:K}(\boldsymbol{x},s) - \mu_{1:K}(\boldsymbol{x},s) - b^*_{n,1:K}(\boldsymbol{x},s),$$

$$\gamma_{n,K} = \widehat{\mu}_{1:K}(\boldsymbol{x},s) - \mu(\boldsymbol{x},s) - b^*_{n,1:K}(\boldsymbol{x},s) = \beta_{n,K} - \sum_{k=(K+1)}^{\infty} \mu_k(\boldsymbol{x})b_k(s).$$

By an application of triangle inequality, it follows that

$$\alpha_{n,K_{\epsilon}}(|\beta_{n,K_{\epsilon}}| - \epsilon b_{\infty}(s)) \leqslant \alpha_{n,K_{\epsilon}}|\gamma_{n,K_{\epsilon}}| \leqslant \alpha_{n,K_{\epsilon}}(|\beta_{n,K_{\epsilon}}| + \epsilon b_{\infty}(s)).$$

Combining this with (S.16) now completes the proof.

S.3   Real-life example: Analyzing Soccer data

As another interesting illustration, we apply our proposed methodology to analyze the shot accuracy patterns in the English Premier League (EPL) across multiple seasons. The dataset

is obtained from the `worldfootballR` package (Zivkovic 2022). In a soccer match, various players from both teams take shots from different positions on the field, and as mentioned in Section 1, xG quantifies the quality of each attempt based on the shot characteristics. The reader is referred to Anzer & Bauer (2021) for a detailed exposition on this metric. However, it is important to recognize that xG does not explicitly account for team-specific or player-specific effect. We undertake a small-scaled analysis to uncover how this measure varies spatially across the field and to identify systematic effects associated with different teams and seasons. Specifically, we analyze the attacking patterns of Arsenal, Liverpool and Manchester City for three different seasons – 2017/18, 2021/22, and 2024/25. These three teams are chosen for specific reasons: since 2017/18, only Manchester City and Liverpool have managed to win the EPL whereas during the same period Arsenal have been on a path of steady growth and are at the top of the league table in the ongoing season. In the three seasons we consider, Manchester City were the winners in the first two and finished third in the last one. Liverpool won the league in 2024/25 and finished fourth and second in the other two seasons respectively. Arsenal, on the other hand, finished in the sixth, fifth and second position in the league during these three seasons.

For modeling purposes, the response variable $Y_t(s)$ is taken to be the logit-transformed xG for a shot taken from location s in the tth match. The covariate vector $\boldsymbol{X}_t$ is defined as team-specific profile at the tth match (t can be 1 to 38 during a season), including the proportion of matches already played in the season, win–loss ratios, average goals scored and conceded, mean number of shots, and proportion of shots on target, among others. Figure S.1 displays the estimated mean xG surface – the sigmoid function applied on $\hat{\mu}(\boldsymbol{X}_t, s)$ – for the three representative teams in the three different seasons as mentioned. For illustration, these estimates are obtained under the setting where the covariates are fixed at an average level.

Typically, we observe that the shots are most effective from central locations close to goal, particularly within the penalty area, and this canonical spatial structure is present across all estimated surfaces, serving as a basic validation that the fitted mean function aligns with domain expectations. The key insight is in the team and season-specific deviations from this baseline geometry. For Arsenal, the last season's estimated surface exhibits a comparatively broader region of moderate expected xG outside the penalty box, which is consistent with a shift toward taking (or creating) shots from longer range and wider areas. Although this should not be interpreted as claiming that long shots become high-probability outcomes, we find it imperative to highlight that Arsenal's ranking in the league table improved in 2025. Statistically, our model suggests a relative increase in the conditional expected shot quality associated with those locations, once the match-level context $\boldsymbol{X}_t$ is taken into account.

For Liverpool, we notice a more concentrated pattern of xG in the central locations inside the box. It suggests that the chances created by Liverpool were more impactful in the last season, potentially becoming critical in helping them win the league. One must further observe that it is qualitatively similar to Manchester City's surfaces from the first two seasons under consideration, where they emerged victorious in the end. Generally, it suggests a stable attacking mechanism that consistently produces high-quality chances from the most dangerous zones. In contrast, Manchester City's estimated xG during 2024/25 was less pronounced and can directly be connected to them failing to finish in top two.

Overall, the analysis summarizes how the spatial distribution of shot quality varies across teams and seasons: while all teams exhibit the expected concentration of high xG near goal,
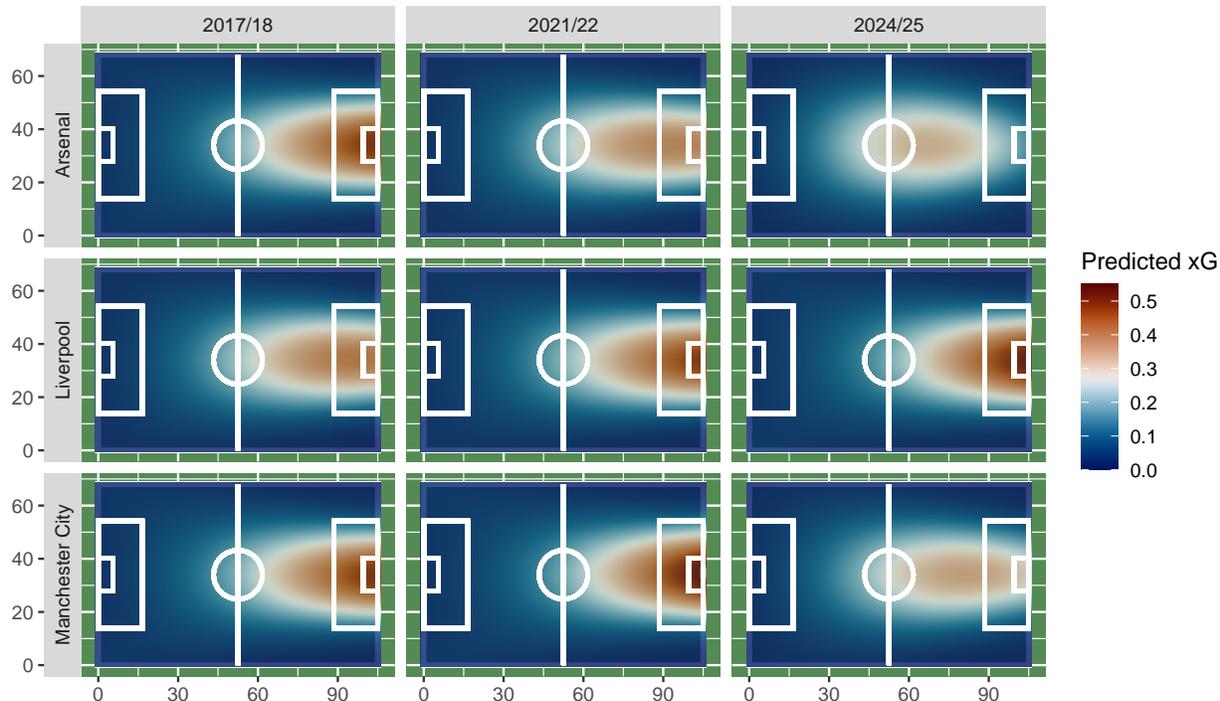
Figure S.1: Estimated Spatial Surface for expected goals (xG) for Arsenal, Liverpool, and Manchester City in different seasons, when playing against an equally strong team.

the extent and spread of moderate-to-high xG regions differ in ways that are consistent with changes in performance across seasons. These patterns provide a compact, interpretable representation of team-specific attacking profiles. With this context in mind, a further advantage of our proposed framework is that these comparisons are not purely descriptive heatmaps: they are conditional spatial surfaces $\hat{\mu}(\boldsymbol{X}_t, s)$, so differences across teams or seasons can be interpreted as systematic effects beyond what is explained by observable team-level covariates. Finally, the results are obtained by fixing equally strong opponent for both the teams. This setting should be read as a controlled scenario in which opponent-related components of $\boldsymbol{X}_t$ are held at comparable levels. Consequently, observed differences in the surfaces are most naturally attributed to team-specific spatial shot-generation profiles rather than imbalances in opponent strength.

S.4 Simulation Studies

To empirically verify the results on the behaviour of the proposed estimator, we consider a simple numerical exercise as follows. Let n be the number of timepoints and p^2 be the number of spatial locations at each timepoint. We pick regularly spaced time and location intervals, namely $\mathcal{T}_n = \{0, 1/(n-1), \dots, 1\}$ and a regularly spaced grid $\mathcal{S}$ on $[0,1]^2$ given by the points $\{(i/(p-1), j/(p-1)) : i, j \in 0, 1, \dots, (p-1)\}$. Since, $\mathcal{S} \subset \mathbb{R}^2$, we pick a set of polynomial basis functions for the Hilbert space given by

$$b_1(s) = 1, \qquad b_2(s) = \sqrt{12}\left(s_1 - \frac{1}{2}\right), \qquad b_3(s) = \sqrt{12}\left(s_2 - \frac{1}{2}\right)$$

$$b_4(s) = 6\sqrt{5}\left(s_1^2 - s_1 + \frac{1}{6}\right), \quad b_5(s) = 6\sqrt{5}\left(s_2^2 - s_2 + \frac{1}{6}\right), \quad b_6(s) = 12\left(s_1 - \frac{1}{2}\right)\left(s_2 - \frac{1}{2}\right),$$

where $s = (s_1, s_2) \in \mathcal{S}$. The above set of basis functions is derived from a shifted version of the Legendre polynomials upto degree 2 on $[0,1]^2$.

First, we generate the covariate X_t based on an autoregressive model

$$X_t = 1 + 0.5 X_{t-1} + e_t, \ e_t \sim N(0, 0.1^2).$$

Then, we generate the response variable as

$$Y_t(s) = X_t(s_1 + s_2) + \epsilon_t(s), (\epsilon_t(s_1), \dots, \epsilon_t(s_{p^2})) \sim N\left(0, 0.1^2((e^{-\|s_i - s_j\|^2}))_{i,j=1}^{p^2}\right),$$

so that $\mu(X_t, s) = X_t(s_1 + s_2)$ and $\sigma(X_t, s) = 1$. It is clear that, the provided truncated basis functions can exactly express the mean and the covariance terms as a linear combination without requiring the higher order terms. Figure S.2 showcases one such sample realization.
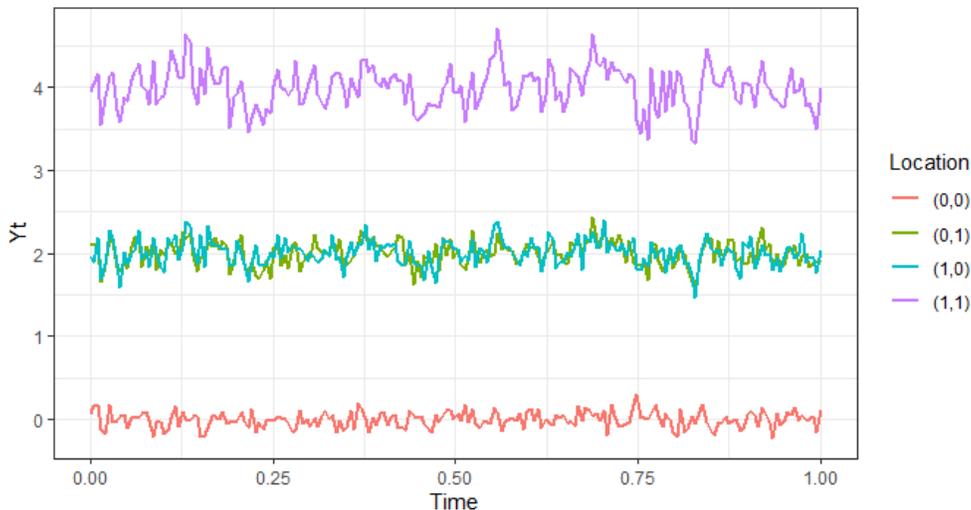

Figure S.2: A sample realization of the simulated data for $n = 200$ and $p = 2$ (i.e., 4 locations)

After we simulate the data from the above model with $n = 100$ and $p = 15$, we use the proposed methodologies to predict the mean function $\hat{\mu}$, and as a result, obtain prediction

21

for $\widehat{Y}_t(s)$ for the last 10 timepoints (among those $n = 100$) for all locations using the first 90 timepoints as observations. About the choice of the covariate, we make three different situations as follows:

(S1) At timepoint t, the original covariate X_t is known, i.e., we can estimate $\widehat{Y}_t(s)$ by $\widehat{\mu}(X_t, s)$.

(S2) At timepoint t, the original covariate X_t is not known, but only its lagged value X_{t-1} is known. Since, $X_t = 1 + 0.5X_{t-1} + e_t$, it follows that

$$Y_t(s) = (s_1 + s_2) + 0.5X_{t-1}(s_1 + s_2) + \epsilon_t(s) + (s_1 + s_2)e_t,$$

which is also covered by the proposed model in (1).

(S3) At timepoint t, only the lagged value of response Y_{t-1} is known. Note that, the average of $Y_t(s)$ over the spatial grid is equal to $\sum_{s \in \mathcal{S}_p} Y_t(s) = X_t \times C + \sum_{s \in \mathcal{S}_p} \epsilon_t(s)$, where C is a fixed deterministic constant. Therefore, the knowledge of Y_{t-1} can be incorporated into model (1) by taking a proxy covariate X_t' to be the average of lagged responses across all locations, which reduces it to the former case (S2).

We consider $B = 100$ replications of this dataset, and for each replication we compute the estimates. For demonstration purposes, we tabulate the metrics such as Bias, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) for each timepoints across all replications and all locations in the Table S.1.

Scenario	Metric	Time Index (t)
		0.909	0.919	0.929	0.939	0.949	0.960	0.970	0.980	0.990	1.000
(S1)	Bias	-0.100	-0.106	-0.104	-0.103	-0.0980	-0.0950	-0.107	-0.0981	-0.108	-0.102
	MAE	0.439	0.439	0.440	0.440	0.439	0.438	0.440	0.437	0.441	0.439
	RMSE	0.543	0.543	0.544	0.544	0.542	0.540	0.544	0.540	0.545	0.543
	MAPE (%)	25.5	25.2	25.4	25.3	25.5	25.5	25.3	25.4	25.2	25.4
(S2)	Bias	-0.0935	-0.0999	-0.0923	-0.0971	-0.0919	-0.0952	-0.0993	-0.0985	-0.0920	-0.0961
	MAE	0.444	0.445	0.446	0.444	0.445	0.445	0.447	0.443	0.449	0.447
	RMSE	0.550	0.551	0.553	0.550	0.550	0.551	0.554	0.548	0.558	0.555
	MAPE (%)	25.9	25.7	26.0	25.7	25.9	25.8	25.8	25.6	26.1	25.9
(S3)	Bias	-0.111	-0.154	-0.148	-0.152	-0.153	-0.167	-0.157	-0.170	-0.173	-0.176
	MAE	0.449	0.566	0.567	0.570	0.588	0.595	0.596	0.606	0.606	0.621
	RMSE	0.558	0.704	0.704	0.708	0.732	0.743	0.744	0.759	0.759	0.784
	MAPE (%)	25.6	32.8	33.1	33.1	34.7	34.9	35.6	36.2	35.5	37.6

Table S.1: Simulation results for scenarios (S1)-(S3) with $n = 100, p = 15$.

It is also interesting to see how the prediction error would decrease as a function of the p, the number of gridpoints on the spatial horizon. Figure S.3 demonstrates this behaviour which shows that the MAPE decreases as the number of spatial locations increases, as assured by the consistency result in Theorem 2.

References

Andrews, D. W. K. (1988), 'Laws of large numbers for dependent non-identically distributed random variables', *Econometric Theory* **4**(3), 458–467.
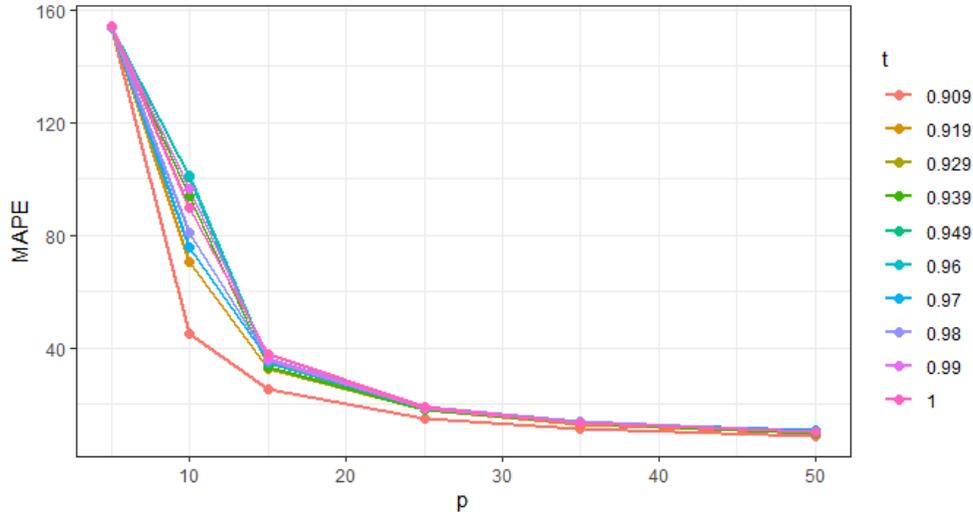
Figure S.3: Resulting MAPE in predicting different timepoints as a function of p, the number of grid points

Anzer, G. & Bauer, P. (2021), 'A goal scoring probability model for shots based on synchronized positional and event data in football (soccer)', *Frontiers in sports and active living* **3**, 624475.

Hong, S. Y. & Linton, O. (2020), 'Nonparametric estimation of infinite order regression and its application to the risk-return tradeoff', *Journal of Econometrics* **219**(2), 389–424.

Masry, E. (2005), 'Nonparametric regression estimation for dependent functional data: asymptotic normality', *Stochastic Processes and their Applications* **115**(1), 155–177.

Zhao, Z. & Wu, W. B. (2008), 'Confidence bands in nonparametric time series regression', *The Annals of Statistics* **36**(4), 1854–1878.

Zivkovic, J. (2022), *worldfootballR: Extract and Clean World Football (Soccer) Data.* R package version 0.6.2.