

# NONPARAMETRIC QUANTILE REGRESSION FOR SPATIO-TEMPORAL PROCESSES

BY SOUDEEP DEB<sup>1,a</sup>  CLÁUDIA NEVES<sup>2,b</sup>  AND SUBHRAJYOTY ROY<sup>3,c</sup> ,

<sup>1</sup>Decision Sciences, Indian Institute of Management Bangalore, <sup>a</sup>[soudeep@iimb.ac.in](mailto:soudeep@iimb.ac.in)

<sup>2</sup>Department of Mathematics, King's College London, <sup>b</sup>[claudia.neves@kcl.ac.uk](mailto:claudia.neves@kcl.ac.uk)

<sup>3</sup>Applied Statistics Division, Indian Statistical Institute Kolkata, <sup>c</sup>[subhrajyoty\\_r@isical.ac.in](mailto:subhrajyoty_r@isical.ac.in)

We establish convergence rates for this nonparametric infinite-dimensional methodology relating quantile regression, investigating conditions under which it estimation is feasible and of good overall quality. We also address main computational issues faced and discussed how these were averted. Finally, we present a finite sample study the empirical performance of the proposed estimation method as well as testing procedure. Our proposed infinite-dimensional framework for inference builds on a non-trivial combination of moment-type results borne by a functional central limit theorem tailored to martingale differences.

**1. Introduction.** In the context of various economic or environmental research problems, it is often important to estimate the effect of a covariate on the behavior of the quantiles of a continuous response variable, and it is well known that a simple regression describing the conditional expectation of the response variable may be inadequate (see [Reich \(2012\)](#) for relevant discussions). Conditional quantiles, in this regard, can provide a much better description of the spread of the conditional distribution in both tails.

Originally proposed by [Koenker and Bassett Jr \(1978\)](#), quantile regression in the simplest setting serves as an alternative approach to mean regression without requiring the strict assumption about normally distributed residuals in the linear models. In fact, conditional quantiles can provide adequate description of the data about robustness to heavy-tailed error distributions and presence of outliers, cf. [Chaudhuri, Doksum and Samarov \(1997\)](#); [Ezzahrioui and Ould-Saïd \(2008\)](#). Naturally, quantile estimation in time series or space-time data offers a plethora of applications in real life. Our leading example in this article is the analysis of conditional quantiles of electricity load profiles originating from smart-meter households' records. Modelling and predicting the peak-load events can help distribution network providers maintain a stable supply chain relationship at particular times of the day or week so as to ward off potential power outages ([Jacob, Neves and Vukadinović Greetham, 2020](#)). We tackle this problem through a nonparametric approach in the setting of a multivariate time series with minimal assumptions on the data generating process. Before delving deeper into the ideology of our method, a succinct account of extant literature in this context is of the essence here. Following the aforementioned seminal paper of [Koenker and Bassett Jr \(1978\)](#), various parametric, semi-parametric or nonparametric methods have been developed for conditional quantile function estimation. We particularly focus on the appeal of the nonparametric approaches which not only require minimal assumptions about the random structure and distribution of the response variable, but also enable additional flexibility in adjusting for different factors and confounding effects.

There have been primarily two distinct nonparametric approaches towards modelling the conditional quantiles when dealing with a response variable at a time. One such approach

---

*MSC2020 subject classifications:* Primary 00X00, 00X00; secondary 00X00.

*Keywords and phrases:* First keyword, second keyword.

hinges on a pinball type loss function conditional on the value of the covariate and view the resulting estimation as an optimization routine. In order to translate this into a viable and tractable problem, one often needs to restrict attention to continuous functionals of the covariate, or functionals belonging to a reproducible kernel Hilbert space (Takeuchi et al., 2006), or polynomial functions of the covariate (Zhou and Wu, 2009), which would typically result in smooth quantile curves. A caveat is that these methods can suffer from the undesirable quantile crossing effect, thus requiring additional constraints to be factored into the optimization setup. Contrary to that, the second approach relies on nonparametric kernel estimation of the conditional distribution function (cf. Ferraty, Rabhi and Vieu, 2005; Ezzahrioui and Ould-Said, 2008, and references therein), upon which to invert for obtaining the desired conditional quantiles automatically yield conditional quantile curves that satisfy the non-crossing criterion. However, with the inversion step, the smoothness quality of the estimated quantile curve is not guaranteed and can pose a challenge in real life applications.

Turning attention to multivariate time series or spatio-temporal data, we note that despite the vastly accumulating literature on conditional quantile function estimation in the independent setting, progress in the development of feasible methodology in these domains are limited, possibly due to the challenges faced when accounting for the inherent dependencies in a meaningful way that does not break down at the laws of physics underpinning the process generating the data. To the best of our knowledge, only a few methodologies have been proposed thus far. A couple of Bayesian approaches (e.g. Reich, Fuentes and Dunson, 2011; Das and Ghosal, 2017) tend to anchor on basis polynomials with ensuing considerations in terms of what makes a suitable choice for prior distributions to the relevant parameters alongside and which allowances must be brought into practice to make the method feasible. Any misspecification involving the latter is detrimental to the overall quality and feasibility of inference, thus carrying a significant potential for undermining end-users' confidence in the statistical methods being employed. Another approach (e.g. Reich, 2012; Duan et al., 2021) considers each quantile curve of different locations or coordinates separately, with the levels of quantile curves not affecting each other except for the non-crossing condition. It implies that such approaches may fail in the situations where observations at most coordinates follow a general trend (thus remaining close to the median) with only a few locations exhibiting unusual behaviour.

In contrast to the above techniques, in this paper, we introduce a nonparametric estimator of the conditional quantile of a continuous response variable. Our method is shown to deliver smooth estimates whilst demonstrating to effectively mitigate against the above-mentioned quantile crossing effect. The extremity issue is overcome by modelling the multivariate quantiles of a distribution through an appropriate random element of the unit euclidean ball. The asymptotic properties are developed as part of the study for the proposed estimation procedure. We consider the asymptotic framework where both the number of time-points and the number of locations grow to infinity, which demand several modifications to the conventional Taylor series method used for proving consistency and asymptotic normality of estimators. Our approach leverages the sub-Gaussian properties of the data to maintain control over higher-order moments. Moreover, ensuring consistency demanded the convergence to happen in a uniform sense to account for the changing vector spaces with varying dimension, bringing additional complexity. This complexity was managed by employing concentration inequalities and a sequence of compact sets to limit the difference between estimated and true quantiles. Lastly, to address asymptotic normality, the notion needed to be appropriately modified using Portmanteau's theorem by showing that for a rich class of smooth functions with bounded derivatives, the expectation of the functional value of the normalized estimators converges to the expectation of the same functional value but applied on an appropriate slice of Gaussian process.

The methodological details and the key results developed in this article are described in Sections 2, 3 and 4. Following the asymptotic theory, we also develop simultaneous confidence bands for the spatio-temporal quantile curves across all locations and time-points in Section 5, alongside their appertaining statistical tests. In order to maintain the flow of the paper, we defer the proofs of the theoretical results until Section 6. Finally, in Section 7, we present the maiden real-world application of the proposed estimator to electricity load forecasting. We also include appropriate numerical investigations which offer assurances of the good overall performance of our estimator. Indeed, the key large sample properties developed in the paper are mirrored in the finite sample simulation study encompassing Appendix B.

**2. Spatio-temporal framework for quantile regression.** This paper is concerned with the statistical modeling of dependent data endowed with a spatio-temporal structure that builds on as well as expands preceding related works in [De Gooijer, Gannoun and Zerom \(2006\)](#); [Zhou \(2010\)](#); [Deb and Jana \(2021\)](#).

Let  $\{Y(t, s) : t \in \mathcal{T}, s \in \mathcal{S}\}$  be a spatio-temporal process indexed by temporal domain consisting of a (dense) temporal horizon  $\mathcal{T} = [0, 1]$  and by a countable spatial horizon set  $\mathcal{S} \subset \mathbb{R}^2$ . Subsequently, for all  $d \in \mathbb{N}$ , let  $\mathbf{X} : \mathcal{T} \rightarrow \chi \subseteq \mathbb{R}^d$  be a  $d$ -dimensional temporal covariate process such that for each  $t \in \mathcal{T}$ ,  $\mathbf{X}(t) = (X_1(t), \dots, X_d(t))'$  is a predictor variable in  $\chi \subseteq \mathbb{R}^d$ . In the sliding scale of generality, this covariate process can be allowed to be made fairly generic so as to encapsulate external factors that might affect the temporal dependence to the main process  $Y(t, s)$ , to being set out as specialized into a given periodic component aiming to capture inherent seasonality.

Out interest lies in the conditional space-time process  $Y(t, s) | \{\mathbf{X}(t) : t \in \mathcal{T}\}$  with expectation and covariance structure denoted by

$$(1) \quad \mathbb{E}(Y(t, s)) = \mu(\mathbf{X}(t), s) \quad \text{and} \quad \text{Cov}(Y(t, s_1), Y(t, s_2)) = \Sigma(\mathbf{X}(t), s_1, s_2),$$

for every  $s, s_1, s_2 \in \mathcal{S}$  and  $t \in \mathcal{T}$ , bringing to the fore that the temporal dependence in the process  $Y(t, s)$  is essentially driven by that of the nested covariate process  $\mathbf{X}(t)$ . Obviously, it would be beguiling to assume that  $Y(t, s)$  can actually be observed in the way of covering the whole set  $\mathcal{T} \times \mathcal{S}$ , so we must settle with  $Y(t, s)$  indexed by the discrete subset of time points  $\mathcal{T}_n = \{t_1, t_2, \dots, t_n\}$  and locations  $\mathcal{S}_p = \{s_1, s_2, \dots, s_p\}$ , where both  $n$  and  $p$  are assumed to be sufficiently large (i.e. both tending to  $\infty$  as appropriate). This renders a more realistic framework for inference with a practical resonance: for each  $t \in \mathcal{T}_n$ , consider that the data is generated by a nonparametric spatial regression model,

$$(2) \quad \mathbf{Y}(t) = \begin{pmatrix} Y(t, s_1) \\ Y(t, s_2) \\ \vdots \\ Y(t, s_p) \end{pmatrix} = \boldsymbol{\mu}(\mathbf{X}(t)) + \boldsymbol{\Sigma}^{1/2}(\mathbf{X}(t)) \begin{pmatrix} \xi_{t1} \\ \xi_{t2} \\ \vdots \\ \xi_{tp} \end{pmatrix},$$

where  $\boldsymbol{\mu}(\mathbf{X}(t))$  is a random vector in  $\mathbb{R}^p$  of the form

$$\boldsymbol{\mu}(\mathbf{X}(t)) := \begin{pmatrix} \mu(\mathbf{X}(t), s_1) \\ \mu(\mathbf{X}(t), s_2) \\ \vdots \\ \mu(\mathbf{X}(t), s_p) \end{pmatrix},$$

the covariance matrix is given by  $\boldsymbol{\Sigma}(\mathbf{X}(t))$ , whose  $(i, j)$ -th entry is  $\Sigma(\mathbf{X}(t), s_i, s_j)$  for  $1 \leq i, j \leq p$ , and the (unobserved) errors  $(\xi_{tj})_{j=1, \dots, p}$  are independent and identically distributed with mean zero, positive density on an interval containing 0 (albeit not necessarily that of a normal distribution) and unit variance.

The framework (2) lays bare that the key problem of estimating a spatial quantile at any give time  $t \in \mathcal{T}$  fits well into the logic of estimating a  $p$ -multivariate quantile for the distribution of  $\mathbf{Y}(t)$  centered at time-points  $t_1, t_2, \dots, t_n$ . In this sense, the above framework remains quite general, so much so that it will enable existing methods for quantile regression to be employed, including in those situations where the spatio-temporal dependence structure is only implicitly defined as part of  $\Sigma(\mathbf{X}(t))$ . If relevant information about the spatial dependency structure is available *a priori*, then this can be encapsulated in  $\Sigma(\cdot)$  through a set of appropriate restrictions, either spanning the extent of a nonparametric class of models to being made fundamentally reliant on a certain postulated parametric model, in a process that echoes a familiar balancing act: the well-known robustness-efficiency trade-off. Since, in this paper,  $\mathbf{Y}(t)$  essentially depends on the temporal domain through the covariate process  $\mathbf{X}(t)$  (assumed predictable) the focus will be indeed on the nonparametric estimation of conditional multivariate quantiles of  $\mathbf{Y}(t)$ , given  $\mathbf{X}(t) = \mathbf{x}$ .

To this date, various definitions for a quantile of a multivariate distribution have been put forward, notably those drawing on convex hulls (Eddy, 1985), minimum volume sets (Einmahl and Mason, 1992), geometric generalization of spatial median (Chaudhuri, 1996) and center outward quantile functions (Figalli, 2018). Amongst these, the geometric concept of a quantile in  $\mathbb{R}^d$ ,  $d \geq 2$ , originating from Chaudhuri (1996) and endorsed in Konen and Paidaveine (2022), stands out as the most amenable to our desired aim of extending the quantile regression framework of Koenker and Bassett Jr (1978) to the nonparametric spatio-temporal setting.

Central to this paper is the loss function  $\Phi(\mathbf{u}, \mathbf{v}) = \|\mathbf{v}\| + \langle \mathbf{u}, \mathbf{v} \rangle$ , for all  $\mathbf{u} \in B^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| < 1\}$  and the estimated quantile  $\mathbf{v} \in \mathbb{R}^d$ , in which setting a recognizable form of the multivariate  $\mathbf{u}$ -quantile of a  $d$ -variate random vector  $\mathbf{X}$  beckons:

$$(3) \quad Q(\mathbf{u}) := \arg \min_{\mathbf{q} \in \mathbb{R}^d} \mathbb{E} [\Phi(\mathbf{u}, \mathbf{X} - \mathbf{q})],$$

(cf. Chowdhury and Chaudhuri, 2019, p.397). As desired, a choice of  $\mathbf{u} \in B^{d-1}$  in (3) towards the boundary, i.e.  $\|\mathbf{x}\| \uparrow 1$ , leads to a (high or low) extreme quantile, whereas fixing  $\mathbf{u}$  near the origin yields spatially centered measurable functionals, including expectiles such as the mean vector ascribed to the  $\mathcal{L}_2$ -objective function in (3) or the spatial median in connection with  $\mathcal{L}_1$ -norm. Since the vector  $\mathbf{u}$  depends on the dimension  $p$ , i.e., the number of spatial points at play, which we allow to grow so as to attain asymptotic results, we consider  $\mathbf{u}$  as a spatial process (measurable with respect to the Lebesgue measure on  $\mathbb{R}^p$ ) satisfying

$$\int_{\mathcal{S}} |\mathbf{u}(s)|^2 ds \leq 1,$$

which, for every finite restriction  $\mathcal{S}_p$  of  $\mathcal{S}$ , translates to  $\mathbf{u}|_{\mathcal{S}_p} = (\mathbf{u}(s_1), \dots, \mathbf{u}(s_p)) \in B^p$ . In a similar vein, we now define for every  $t \in \mathcal{T}$ ,  $\mathbf{x} \in \chi$ , the  $\mathbf{u}$ -conditional spatial quantile of  $\mathbf{Y}(t)$  given  $\mathbf{X}(t) = \mathbf{x}$  as

$$(4) \quad Q(\mathbf{u}, \mathbf{x})|_{\mathbb{R}^p} = \arg \min_{\mathbf{q} \in \mathbb{R}^p} \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\| + \langle \mathbf{u}|_{\mathcal{S}_p}, \mathbf{Y}(t) - \mathbf{q} \rangle \mid \mathbf{X}(t) = \mathbf{x}],$$

where  $Q(\mathbf{u}, \mathbf{x})|_{\mathbb{R}^p}$  highlights that the conditional quantile process  $Q(\mathbf{u}, \mathbf{x})$  is indexed by a  $p$ -dimensional spatial domain  $\mathcal{S}_p$ . The set  $\mathcal{S}_p$  has to be asymptotically dense in  $\mathcal{S}$  as  $p \rightarrow \infty$ , i.e., for any  $s \in \mathcal{S}$  and however small  $\epsilon > 0$ , we can find a point  $s^* \in \mathcal{S}_p$  within the ball of radius  $\epsilon$  centered at  $s$  for sufficiently large  $p$ . It is important to stress that the proposed framework prompts a significant generalization to the smoothing spline approach initially proposed in Koenker, Ng and Portnoy (1994); the consideration of a temporal domain adds to a body of contemporary works in a substantive way, which to the extent of our knowledge have thus far been limited to the study of nonparametric spatial quantile regression. Our proposal is distinctive in that the criterion (loss) function presiding the M-estimation is no

longer a fixed function but rather it is allowed to vary with  $p$  whilst preserving many of the fundamental properties for quantile regression investigated in [Chowdhury and Chaudhuri \(2019\)](#) and references therein. This is the key insight to extending the notion of a spatial (multivariate) quantile across a direction, towards the aim of tackling the spatio-temporal regression estimation problem posed in (4). The precise meaning of the latter statement will become apparent in Section 3.

One natural question about the definition of the conditional quantile process given in (4) is that of whether the definition remains identifiable across choices of the set  $\mathcal{S}_p$ . For example, if  $\mathcal{S}^* \subseteq \mathcal{S}$  is a superset of  $\mathcal{S}_p$  for some fixed  $p$ , then the restriction of  $Q(\mathbf{u}, \mathbf{x})|_{\mathcal{S}^*}$  to the set of points  $\mathcal{S}_p$  should be same as  $Q(\mathbf{u}, \mathbf{x})|_{\mathbb{R}^p}$ . To see this, we consider  $\mathcal{S}^* = \mathcal{S}_p \cup (\mathcal{S}^* - \mathcal{S}_p) = \mathcal{S}_1 \cup \mathcal{S}_2$  and correspondingly define  $\mathbf{Y}_1(t), \mathbf{Y}_2(t)$  and  $\mathbf{q}_1, \mathbf{q}_2$  as the restrictions of  $\mathbf{Y}(t)$  and  $\mathbf{q}$  respectively. Finally, we note that

$$Q(\mathbf{u}, \mathbf{x})|_{\mathcal{S}^*} = \arg \min_{\mathbf{q}_1, \mathbf{q}_2} \mathbb{E} \left[ \sqrt{\|\mathbf{Y}_1(t) - \mathbf{q}_1\|^2 + \|\mathbf{Y}_2(t) - \mathbf{q}_2\|^2} \right. \\ \left. + \langle \mathbf{u}|_{\mathcal{S}_1}, \mathbf{Y}_1(t) - \mathbf{q}_1 \rangle + \langle \mathbf{u}|_{\mathcal{S}_2}, \mathbf{Y}_2(t) - \mathbf{q}_2 \rangle | \mathbf{X}(t) = \mathbf{x} \right].$$

It now clearly follows that optimizing with respect to  $\mathbf{q}_1 \in \mathbb{R}^p$  yields a minimizer of (4) for any fixed value of  $\mathbf{q}_2$ , thus ensuring identifiability.

**3. Nonparametric spatio-temporal quantile estimation.** For ease of introduction, we firstly address the case where we look at one location at a time. This corresponds to the case of  $p = 1$  in model (2), whereby the canonical nonparametric quantile estimator arises as the kernel-weighted version of the quantile  $q$  of order  $\tau \in (0, 1)$  of the corresponding empirical probability distribution function of  $Y(t)$  given  $\mathbf{X}(t) = \mathbf{x}$  on  $\chi = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ . Formally, for  $u \equiv u(\tau) = (2\tau - 1)$ ,  $0 < \tau < 1$ , the quantile  $q$  of order  $\tau$  is the minimizer of

$$(5) \quad \mathcal{M}_{u,n}^{(1)}(q) = \sum_{i=1}^n k(\mathbf{x}, \mathbf{X}(t_i)) (|Y(t_i) - q| + u \cdot (Y(t_i) - q)).$$

In this respect, we note that  $\mathcal{M}_{u,n}^{(1)}$  is convex on  $\mathbb{R}$ . With  $\tau = 1/2$  standing for the median as the minimizer of  $\mathcal{M}_{u,n}^{(1)}$ , then  $u$  can be viewed as a gauge for central displacement or tail-relatedness, since the greater the  $|u|$  the more tail-related the quantile. Clearly, finding minimizers of (5) configures an  $M$ -estimation problem, and with the stated sample criterion function incorporating a kernel function of bandwidth  $b_n$ , this is one of a nonparametric nature. By this token, a generalization to the  $p$ -multivariate case can be readily achieved by taking matrix valued kernel functions  $\mathbf{K}(\cdot, \cdot)$ , suitably normalized, into the inner product that constitutes (5) thus giving way to the following definition of a nonparametric spatio-temporal  $\mathbf{u}$ -quantile that mirrors (4) in the associated sequence of statistical experiments: for  $\mathbf{u}|_{\mathcal{S}_p} \in \mathbb{R}^p$  comprised of the set  $\{u(s) : s \in \mathcal{S}_p\}$ , we define the estimator

$$(6) \quad \hat{q}_n(\mathbf{u}, \mathbf{x}) := \arg \min_{\mathbf{q} \in \mathbb{R}^p} \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}),$$

where

$$(7) \quad \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) := \frac{1}{n} \sum_{i=1}^n \left\{ \|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}(t_i)} + \langle \mathbf{u}|_{\mathcal{S}_p}, \mathbf{Y}(t_i) - \mathbf{q} \rangle_{\mathbf{K}(t_i)} \right\},$$

for which we write

$$\mathbf{K}(t_i) := \mathbf{K}(\mathbf{x}, \mathbf{X}(t_i)) \\ \|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)}^2 := \|\mathbf{K}(\mathbf{x}, \mathbf{X}(t))(\mathbf{Y}(t) - \mathbf{q})\|^2, \\ \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)} := \mathbf{u}^\top \mathbf{K}(\mathbf{x}, \mathbf{X}(t))(\mathbf{Y}(t) - \mathbf{q}).$$

If the objective function  $\mathcal{M}_{\mathbf{u},n}^{(p)}$  in (7) has been highlighted to depend on both the parameter dimension  $p$  and sample size  $n$ , in view of (4) its counterpart population version (for criterion function) is laid out as

$$\mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}) := \mathbb{E} \left[ \|\mathbf{Y}(t) - \mathbf{q}\| + \langle \mathbf{u}|_{\mathcal{S}_p}, \mathbf{Y}(t) - \mathbf{q} \rangle \mid \mathbf{X}(t) = \mathbf{x} \right].$$

Later on, in Section 4, this setting will be expanded with  $p$  taken as a sequence of positive integers indexed in  $n$ .

Below, we collect the key assumptions underpinning the asymptotic results established in this paper, as well as noting that each of which will be invoked as deemed appropriate and/or necessary throughout.

- (A1) The covariate process  $\mathbf{X}(t)$  is a predictable process, i.e., there exists a sequence of random variables  $\{\epsilon_t : t = t_1, t_2, \dots\}$ , with its natural filtration  $\mathcal{F}_t$  independent of the random variables  $\xi_{tj}$ , such that  $\mathbf{X}(t_i)$  is  $\mathcal{F}_{t_{(i-1)}}$ -measurable for any  $t_i \in \mathcal{T}_n$ .
- (A2) There exists a sequence of convex, open sets  $\mathcal{U}_1 \subseteq \mathbb{R}, \mathcal{U}_2 \subseteq \mathbb{R}^2, \dots, \mathcal{U}_p \subseteq \mathbb{R}^p$  where the sequence of criterion functions  $\mathcal{M}_{\mathbf{u},n}^{(p)}$ , defined as in (7) takes on the vector-valued  $\mathbf{q} \in \mathcal{U}_p$ , for all  $t_i \in \mathcal{T}$ , and for any fixed  $p \in \mathbb{N}$  not depending on  $n$ . Moreover, we assume that  $\inf_{\mathbf{q} \in \mathcal{U}_p} \inf_{t_i \in \mathcal{T}_p} \|\mathbf{Y}(t_i) - \mathbf{q}\| > \delta$  for some sufficiently small  $\delta > 0$ , independent of  $n$  as well as of  $p$ .
- (A3) The kernel matrix  $\mathbf{K}(\mathbf{x}, \mathbf{X}(t))$  is a positive definite  $(p \times p)$  matrix for all  $t \in \mathcal{T}$  and  $\mathbf{x} \in \chi$ , and the support of the conditional distribution of each  $\mathbf{Y}(t_1), \dots, \mathbf{Y}(t_n)$ , provided  $\mathbf{X}(t) = \mathbf{x}$  is not enclosed in a straight line in  $\mathcal{S}$  (cf. Chowdhury and Chaudhuri, 2019).

Assumption (A1) on the predictable process  $\mathbf{X}(t)$  makes it possible to write  $\mathbf{X}(t_i) = h(\dots, \epsilon_{t_{(i-2)}}, \epsilon_{t_{(i-1)}})$  for some measurable function  $h : \mathbb{R}^\infty \rightarrow \mathbb{R}^d$ . It is a common assumption in related literature. Assumption (A2) is an operational assumption, which in conjunction with (A3), ensures that  $\mathcal{M}_{\mathbf{u},n}^{(p)}$  consists of a sequence of random convex functions on  $\mathcal{U}_p$ , for every  $p \geq 1$ ; hence it forms a sequence of continuous functions. This, in turn, ascertains the existence of M-estimators  $\hat{q}_n$  in the quality of a uniquely defined sequence of minimizers of  $\mathcal{M}_{\mathbf{u},n}^{(p)}$ . Alongside these, we note that the objective function  $\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$  is strictly convex in its argument  $\mathbf{q}$ , and the same conclusion is true for  $\mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q})$  regardless of the square-integrable spatial process  $\mathbf{u}$  such that  $\int_{\mathcal{S}} |u(s)|^2 ds \leq 1$ , and for every fixed  $p \in \mathbb{N}$ . This strict convexity ensures that the true quantile  $\mathbf{q}_0$  is a well-separated minimizer of  $\mathcal{M}_{\mathbf{u}}^{(p)}$ , that is, there exists a  $\mathbf{q}_0 \in \mathcal{U}_p$  for which  $\inf_{\mathbf{q} : \|\mathbf{q} - \mathbf{q}_0\| \geq \delta} \left( \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}_0) \right) > 0$ , for arbitrarily small  $\delta > 0$ , where  $\delta$  may depend on  $p$ .

**3.1. Estimation algorithm.** The idea behind the Iteratively Re-weighted Least Squares (IRLS) type estimation as an effective means to achieve minimization in (6) is to replace the objective function with an equivalent version whose iterative minimization leads to the closest answer to the original problem of obtaining a quantile estimator  $\hat{q}_n(\mathbf{u}, \mathbf{x})$ . Firstly, we expand the sample criterion function  $\mathcal{M}_{\mathbf{u},n}$  as follows:

$$\frac{1}{n} \sum_{i=1}^n \frac{(\mathbf{Y}(t_i) - \mathbf{q})^\top \mathbf{K}^2(t_i) (\mathbf{Y}(t_i) - \mathbf{q})}{\|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}^2(t_i)}} + \langle \mathbf{u}, (\mathbf{Y}(t_i) - \mathbf{q}) \rangle_{\mathbf{K}(t_i)},$$

where we follow the convention that  $\mathbf{y}/\|\mathbf{y}\| = 0$  if  $\mathbf{y} = \mathbf{0}$ . In addition, we highlight that  $\mathbf{q}$  does not change with  $i$ . Henceforth, we shall use  $w_i(\mathbf{q}) := \|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}^{-1}(\mathbf{x}, \mathbf{X}(t_i))^2}^{-1}$  so as to avoid notational burden. Minimization of the above with respect to  $\mathbf{q}$  yields the sequence of



estimators  $\hat{q}_n$ :

$$(8) \quad \hat{q}_n = \left\{ \frac{1}{2} \sum_{i=1}^n \mathbf{K}(t_i) \mathbf{u} + \sum_{i=1}^n w_i(\mathbf{q}) \mathbf{K}^2(t_i) \mathbf{Y}(t_i) \right\} / \sum_{i=1}^n w_i(\mathbf{q}) \mathbf{K}^2(t_i).$$

The above expression suggests that, given an initial estimate  $\hat{q}^{(0)}$ , it is possible to use (8) iteratively in order to improve estimation at the next iteration,  $\mathbf{q}^{(k+1)}$ : by calculating  $w_i(\mathbf{q}^{(k)})$ , for all  $i = 1, 2, \dots, n$ , using the current estimate and then using (8) to go to the next step of the iteration. Despite the matrix  $\sum_{i=1}^n w_i(\mathbf{q}) \mathbf{K}(t_i)$  is guaranteed positive definite with probability tending to one, in practice it may become numerically singular at some stage of the iteration, particularly when the number of locations  $p$  is large. By adding a small positive number  $\rho$  to the diagonal elements of this matrix, such a numerical blunder can be easily averted. The hyperparameter  $\rho$  subsequently becomes the Lagrangian coefficient associated with a new constraint for the objective function (6):  $\|\mathbf{q}\| \leq C$ , for some fixed  $C$  positive number. Indeed, it is of straightforward verification that the IRLS scheme upon (8) will reduce the value of the objective function at the next iteration, i.e.,  $\mathcal{M}_n(\mathbf{q}^{(k+1)}) \leq \mathcal{M}_n(\mathbf{q}^{(k)})$  for all  $k \in \mathbb{N}$ . Combining this fact with the strict convexity of the objective function, it follows that the solution of the algorithm converges to the global minimizer of  $\mathcal{M}_n(\mathbf{q})$  since it is inherently assumed that the set of global minimizers is non-empty.

**3.2. Choice of the kernel and associated bandwidth.** The matrix-valued kernel function  $\mathbf{K}(\mathbf{x}, \mathbf{X}(t))$  involving the bandwidth  $b_n > 0$  requires suitable normalizing constants  $c_n > 0$ . Specifically, we consider the matrix

$$\mathbf{K}(\mathbf{x}, \mathbf{X}(t)) = c_n^{-1} \mathcal{K} \left( \frac{\mathbf{x} - \mathbf{X}(t)}{b_n} \right),$$

given in terms of a symmetric and continuous kernel  $\mathcal{K} : \mathbb{R}^d \rightarrow \mathbb{R}^{p \times p}$ , with compact support on  $B^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq 1\}$ , in keeping with the formulation we typically encounter for such nonparametric adjacency matrices of weights (see e.g. Deb and Jana, 2021; Zhao and Wu, 2008). Specific choices for constants  $b_n, c_n > 0$  must be based not only on theoretical considerations, but also be informed by empirical evidence on the behavior of competing estimators in situations of repeated sampling with varying size, all considered against the backdrop of assumptions threaded in the relevant subject literature. For example, univariate nonparametric kernel estimators for a trend in time series, which have been proposed in Zhao and Wu (2008), appear suitably normalised via  $c_n = \sqrt{n b_n}$ . We also assume that  $\mathcal{K}(0) = c_n \mathbf{I}_p$ , similar to the univariate case where the kernel function at origin is taken to be equal to 1.

**3.3. Key properties of the nonparametric quantile estimator.** The results encompassing this section, summarized in Lemma 3.1 and ensuing considerations, are viewed as fundamental steps to establishing performance guarantees of proposed nonparametric methodology for quantile estimation. Indeed, a desirable property of a quantile estimator is affine equivariance: if the samples  $\mathbf{Y}(t)$  are changed to  $\mathbf{A}\mathbf{Y}(t) + \mathbf{b}$  for some non-singular matrix  $\mathbf{A}$ , the estimator should also exhibit the same affine transformation. We note that the proposed quantile estimator  $\hat{q}(\mathbf{u}, \mathbf{x})$  resulting from (6) is equivariant under location transformation, i.e., employing a shift to observations through  $\mathbf{Y}(t_i) + \mathbf{a}$  changes the sequence of estimators accordingly to  $\hat{q}_n(\mathbf{u}, \mathbf{x}) + \mathbf{a}$ . Although the equivariance also holds in general with respect to a scale transformation  $c\mathbf{Y}(t_i)$ , for some scalar  $c \in (0, \infty)$ , verification of the orthogonal equivariance property (or more general affine equivariance property) is dependent on the incorporating kernel matrix  $\mathbf{K}(t_i)$  (see Chaudhuri, 1996; Chen, 1996; Konen and Paidaveine, 2022).

As highlighted in Section 1, quantile estimators should generally exhibit two key properties: non-crossing of the quantiles and smoothness of the estimator as a function of the data. In

the following discussion, we demonstrate how upon a very minimal set of assumptions, both these properties follow in relation to  $\hat{q}_n(\mathbf{u}, \mathbf{x})$ .

For each  $n \geq 1$ , the quantile estimator  $\hat{q}_n(\mathbf{u}, \mathbf{x})$  is continuous with respect to observables  $\mathbf{Y} = (\mathbf{Y}(t_1), \dots, \mathbf{Y}(t_n))$ . In view of assumption (A2), we can without loss of generality restrict the optimization to a compact set  $\mathcal{Q} \subseteq \mathbb{R}^p$  containing the true targeted quantile. Consequently,  $\hat{q}_n(\mathbf{u}, \mathbf{x}) = \mathbf{q}(\mathbf{Y}) = \arg \min_{\mathbf{q} \in \mathcal{Q}} \mathcal{M}_{\mathbf{u}, n}(\mathbf{q}, \mathbf{Y})$ . Now, let  $\mathbf{Y}_r \rightarrow \mathbf{Y}_0$  be a converging sequence with  $r = 1, 2, \dots$ , and let  $\mathbf{q}_r = \mathbf{q}(\mathbf{Y}_r)$  for every  $r \in \mathbb{N}$ . We proceed with a subsequencing argument: for any  $\{\mathbf{q}_{r_k}\}_{k \in \mathbb{N}}$ , the compactness of  $\mathcal{Q}$  ensures that  $\{\mathbf{q}_{r_k}\}$  has an accumulation point  $\mathbf{q}'$ , say. Hence, we have that  $\mathcal{M}_{\mathbf{u}, n}(\mathbf{q}_{r_k}, \mathbf{Y}_{r_k}) \leq \mathcal{M}_{\mathbf{u}, n}(\mathbf{q}_0, \mathbf{Y}_{r_k})$ ,  $r_k \in \mathbb{N}$ . Since  $\mathcal{M}_{\mathbf{u}, n}$  is a continuous function of  $\mathbf{q}$  and  $r_k \rightarrow \infty$ , then  $\mathcal{M}_{\mathbf{u}, n}(\mathbf{q}', \mathbf{Y}_0) \leq \mathcal{M}_{\mathbf{u}, n}(\mathbf{q}_0, \mathbf{Y}_0)$ , thus entailing  $\mathbf{q}' = \mathbf{q}_0$  for any subsequence, from which the continuity of the estimator ensues.

The next lemma outlines the desirable non-crossing property in any quantile estimation methodology (e.g. in line with Theorem 2.1.2 of Chaudhuri (1996)). It is mainly a technical lemma aiming to justify that the M-estimator  $\hat{q}_n(\mathbf{u}, \mathbf{x})$ , viewed as the approximate (asymptotic) solution of the convex optimization problem (6), gains the configuration of a Z-estimation problem for every fixed  $(\mathbf{u}, \mathbf{x})$ . This is envisaged to simplify substantially the later proofs for asymptotic normality. However, because we will need to take into account the spatial dependence structure on the basis of an infinitely growing number of locations, the rate of convergence will differ from the familiar  $\sqrt{n}$  for tightness mustered by predominant asymptotic theory for Z-estimators (see van der Vaart (1998); Kosorok (2008)).

**LEMMA 3.1.** *Under conditions (A1)–(A3), for every  $n, p \in \mathbb{N}$ ,  $\mathbf{x} \in \mathbb{R}^d$  and  $\mathbf{u} \in B^{p-1}$ , the sequence of minimizers  $\hat{q}_n(\mathbf{u}, \mathbf{x})$  of the objective function  $\mathcal{M}_{\mathbf{u}, n}^{(p)}$  (as in (7)) satisfies*

$$(9) \quad \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left[ \frac{\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \hat{q}_n(\mathbf{u}, \mathbf{x}))}{\|\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \hat{q}_n(\mathbf{u}, \mathbf{x}))\|} + \mathbf{u} \right] = \mathbf{0},$$

for  $\hat{q}_n(\mathbf{u}, \mathbf{x}) \notin \{\mathbf{Y}(t_1), \dots, \mathbf{Y}(t_n)\}$ . The true quantile  $Q(\mathbf{u}, \mathbf{x})$  in the quality of minimizer of  $\mathcal{M}_{\mathbf{u}}^{(p)}$  verifies the population version of (9), specifically

$$(10) \quad \mathbb{E}_{\mathbf{Y}(t)|\mathbf{X}(t)=\mathbf{x}} \left[ \frac{\mathbf{Y}(t) - Q(\mathbf{u}, \mathbf{x})}{\|\mathbf{Y}(t) - Q(\mathbf{u}, \mathbf{x})\|} + \mathbf{u} \right] = \mathbf{0}.$$

The actionable idea stemming from Lemma 3.1 – that we may regard (9) as an estimating equation for obtaining  $\hat{q}_n(\mathbf{u}, \mathbf{x})$  from a finite sample data with  $n$  time-points and  $p$  spatial points – gives rise to an alternative Newton-Raphson type algorithm for obtaining the estimator. Notice that, for  $\mathbf{u} = \mathbf{0}$ , the estimating equation simply corresponds to a weighted spatial median estimate of  $\mathbf{Y}(t_i)$  values, weighted by the covariate process  $\{\mathbf{X}(t)\}$  through the kernel matrix  $\mathbf{K}(t) = \mathbf{K}(\mathbf{x}, \mathbf{X}(t))$ . However, for  $\mathbf{u} \neq \mathbf{0}$ , there is an additional term creeping up in the way of an approximation bias, thereby increasing or decreasing the coordinates of  $\hat{q}_n(\mathbf{x}, \mathbf{u})$  as appropriate in a data-driven manner. This is found to play a key role to ensuring that the estimated quantile curves yielded by our methodology enjoy the non-crossing property. Further details are provided in the next section.

**4. Results.** The nonparametric quantile estimator introduced in Section 3 enjoys a number of desirable large-sample properties. This section is aimed at addressing these by allowing the number of spatial locations  $p$  tend to infinity, albeit in a controlled manner relatively to  $n \rightarrow \infty$ . The obtained asymptotic properties provide the theoretical justification for the distinctive features of the quantile estimation and ensuing testing methodology proposed in this paper.



We now fix some more notation in order to progress. A bounded sequence of numbers  $\{x_n\}$  is throughout written as  $x_n = \mathcal{O}(1)$ , springing out to the notation  $x_n = \mathcal{O}(y_n)$  if  $x_n/y_n = \mathcal{O}(1)$ . In general, if the random sequence  $X_n$  is tight, we write  $\|X\| = \mathcal{O}_p(1)$ , and if  $\|X\|/\|Y\| \xrightarrow{P} 0$  we denote this fact by  $X_n = o_p(Y_n)$ , as  $n \rightarrow \infty$ .

Let  $\mathcal{T}_n = t_1, \dots, t_n$  denote the set of observable time-points that satisfy the smoothing condition  $\max_{2 \leq i \leq n} |t_i - t_{i-1}| \rightarrow 0$  as  $n \rightarrow \infty$ . We also require that  $\mathcal{T}_n$  is asymptotically dense in the temporal horizon  $\mathcal{T}$ . In other words, for any  $\tilde{t} \in \mathcal{T}$ , there exists a sequence of points  $\tilde{t}_n \in \mathcal{T}_n$  such that  $|\tilde{t} - \tilde{t}_n| \rightarrow 0$ .

**THEOREM 4.1.** *Assume conditions (A1)-(A3) hold. Suppose that the number of spatial locations  $p$  is such that  $p = p_n = o(n^{1/2})$ , as  $n \rightarrow \infty$ . Then, for all  $\mathbf{x}$  lying in the compact set  $\chi$ , the sequence of estimators  $\hat{q}_n(\mathbf{u}, \mathbf{x})$  of the population (theoretical) quantile  $Q(\mathbf{u}, \mathbf{x})$  satisfies*

$$(11) \quad \|\hat{q}_n(\mathbf{u}, \mathbf{x}) - Q(\mathbf{u}, \mathbf{x})\| = \mathcal{O}_p(p^2/n),$$

*provided a square-integrable spatial process  $\mathbf{u}$ , i.e.  $\int_{\mathcal{S}} |u(s)|^2 ds < 1$ .*

As part of Theorem 4.1, the number of spatial locations  $p$  is allowed to tend to infinity but at a vanishing rate relatively to  $n^{1/2}$ . The requirement that  $p_n = o(n^{1/2})$ , as  $n \rightarrow \infty$ , means in particular that the case of positive constant  $p$  stays within grasp. It turns out that such a mild restriction on the growth rate of  $p_n$  is indeed what it takes to ensuring the consistency of the quantile estimator established in the theorem. For ascertaining the asymptotic normality of the proposed quantile estimator, however, the number of spatial locations must align with the initial requirement that  $\mathcal{S}_p$  becomes asymptotically dense in  $\mathcal{S}$ .

Analogously to the temporal domain, we fix the setting where the observable spatial points  $\mathcal{S}_p = s_1, s_2, \dots, s_p$  satisfy the condition  $\max_{1 \leq j \neq k \leq n} d(s_j, s_k) \rightarrow 0$  as  $p \rightarrow \infty$ . Here,  $d(s_j, s_k)$  represents the geodesic distance between the spatial coordinates  $s_j$  and  $s_k$ . This ensures that  $\mathcal{S}_p$  becomes dense in  $\mathcal{S}$  with a growing number of spatial locations  $p_n$  based on the geodesic metric. Also, the bandwidth  $b_n$  of the kernel matrix, architected and discussed in Section 3.2 as  $b_n \rightarrow 0$ , must be such that and  $nb_n \rightarrow \infty$  as  $n \rightarrow \infty$  so that the maximum eigenvalue of  $\mathbf{K}(\mathbf{x}, \mathbf{X}(t))$  is uniformly tight.

**4.1. Assumptions.** While assumptions (A1)-(A3) in Section 3 are aimed at laying the groundwork for this paper, the next set of assumptions (B1)-(B5) are required for establishing our main results appertaining the limiting distribution for the proposed quantile estimator.

Let  $f_t$  denote the joint probability density function of  $\mathbf{X}(t)$  and  $\mathcal{G}_k$  be the  $\sigma$ -algebra generated by  $\{\mathbf{Y}(t, s), \mathbf{X}(t) : t \leq t_k, s \in \mathcal{S}\}$ . Additionally, let  $\mathcal{P}_k$  be the projection operator defined as  $\mathcal{P}_k(Z) := \mathbb{E}(Z | \mathcal{G}_k) - \mathbb{E}(Z | \mathcal{G}_{k-1})$ , for all  $k \in \mathbb{Z}$ . Akin to Zhao and Wu (2008), we define the cumulative effect

$$\Xi_n := n\Theta_{2n}^2 + \alpha_n$$

of the first error  $\xi_0$  in predicting the covariate process  $\mathbf{X}(n)$ , where

$$\begin{aligned} \Theta_n &= \sum_{i=1}^n \sup_{\mathbf{x} \in \chi} \|\mathcal{P}_0(f_t(\mathbf{x} | \mathcal{G}_i))\| + \sup_{\mathbf{x} \in \chi} \|\mathcal{P}_0(f'_t(\mathbf{x} | \mathcal{G}_i))\|, \\ \alpha_n &= \lim_{m \rightarrow \infty} \sum_{k=n}^m (\Theta_{n+k} - \Theta_k)^2. \end{aligned}$$

(B1) Suppose that the joint probability density function  $f_t$  is differentiable with respect to each coordinate of  $\mathbf{X}(t)$ , for each  $t \in \mathcal{T}$ . This takes care of finite-dimensional distributions of the observable process.

(B2) The bandwidth  $b_n > 0$  satisfies the following boundary condition, as  $n \rightarrow \infty$ ,

$$\max \left( b_n^{4/3} \log n, \frac{(\log n)^3}{n b_n^3}, \frac{(\log n)^2 \Xi_n}{b_n^{4/3} n^2} \right) = o(1).$$

(B3) There exists a constant  $c > 1$  such that

$$\frac{1}{c} < \sup_{t \in \mathcal{T}} \frac{\lambda_1(\mathbf{K}(t))}{\lambda_p(\mathbf{K}(t))} < c,$$

where  $\mathbf{K}(t) = \mathbf{K}(\mathbf{x}, \mathbf{X}(t))$  is the  $p \times p$  the kernel matrix as before and  $\lambda_1(\mathbf{A})$  and  $\lambda_p(\mathbf{A})$  denote the largest and the smallest eigenvalues of the matrix  $\mathbf{A}$ , respectively.

(B4) The number of spatial points  $p_n$  is such that  $p_n b_n \rightarrow 0$  and  $p_n^{9/8} b_n \rightarrow \infty$ , as  $n \rightarrow \infty$ .

Assumption (B1) is a natural and indeed a customary assumption that ensures smoothness of the probability density functions, allowing one to employ the standard techniques in asymptotic analysis. Assumption (B2) is a further restriction to the growth of the bandwidth  $b_n$  of the kernel matrix, which exerts influence on the strength of dependence of the process. From (B2), it results that  $b_n^{-1}(n^{-1/3} \log n) = o(1)$ , thus imposing a lower bound on  $b_n$ . Analogously, (B2) implies that  $b_n(\log n)^{3/4} = o(1)$ , which provides an upper bound for  $b_n$ . A popular choice for  $b_n > 0$  is that of the type  $c_0 n^{-\beta}$  for some (fixed) constant  $c_0 > 0$ ; if used in tandem with the kernel proposed in Section 3.2, then  $b_n = n^{-1/5}$  is singled out in Zhao and Wu (2008) for producing satisfactory results. Next, we note that the third term embedded in (B2),

$$\left( \frac{\log n}{n^{1/3} b_n} \right)^2 (n b_n)^{2/3} \frac{\Xi_n}{n^2}.$$

implies that  $\Xi_n/n^2 = o(1)$  due to the earlier condition  $b_n^{-1}(n^{-1/3} \log n) = o(1)$  in conjunction with  $n b_n \rightarrow \infty$ . Because each summand of the partial  $\Theta_n$  gives the contribution of  $\epsilon_0$  (as in (A1)) to the associated predictor  $\mathbf{X}(t_{i+1})$ , this term plays a relevant role in moderating the dependence of  $\mathbf{X}(t)$ . In particular, the random vectors  $\mathbf{X}(t_i)$  must satisfy a summability property designed to maintain the integrity of the predictor process  $\mathbf{X}(t)$  by way of ensuring that the covariance structure of the spatio-temporal process is well-defined and can always be estimated. On the whole, this assumption (A1) has a number of parallels with published works on inference for spatio-temporal processes (see e.g. Zhao and Wu (2008); Zhou and Wu (2009); Deb and Jana (2021)). According to Zhao and Wu (2008), a condition such as this corrals virtually the full spectrum for the strength of dependence, from the boundary case of the predictor process  $\mathbf{X}(t)$  exhibiting only short-range dependence (SRD) (determined by  $\lim_{n \rightarrow \infty} \Xi_n < \infty$ ) to a weakening long-range dependence (LRD). Assumption (B3) is required in order to control random oscillations of the kernel matrix. In keeping with the random matrix theory, Wigner's law asserts that the eigenvalues of the  $p \times p$  covariance matrix  $\Sigma(\mathbf{X}(t))$  are of order  $\sqrt{p}$  (Vershynin, 2010) for sufficiently large  $p$ , hence monotonically increasing with  $n$  and  $p$ . This means that for any fixed  $t \in \mathcal{T}$ , the ratio of the largest and the smallest eigenvalue of  $\mathbf{K}(t)$  remains bounded for sufficiently large  $p$ . For the first finitely many terms, boundedness of this ratio is ensured by tightness and positive definiteness of the kernel matrix  $\mathbf{K}(t)$ . Assumption (B3) is a stronger version of this statement that requires uniform boundedness over all choices of  $t \in \mathcal{T}$ . This assumption guarantees that there exists a suitable kernel matrix such that eigenvalues of the covariance matrices of the kernel-weighted observables  $\mathbf{Y}(t)$  remain bounded. Finally, assumption (B4) in conjunction with the bounds on bandwidth  $b_n$  that ensue from (B2) implies that  $(\log n)^{-2/3} \ll p_n \ll (n/\log n)^{1/3}$ .

4.2. *Asymptotic distribution.* In this section, we establish the gaussian limiting distribution attained by the proposed quantile estimator. The consistency of the proposed quantile estimator expounded in Theorem 4.1 and accompanying text herald the rate at which the number of locations  $p$  can grow in relation to the number of time-points  $n$  so as to make this nonparametric quantile estimation feasible.

Continuing along the lines of (7) and adding to the notation introduced as part of Lemma 3.1, we define

$$\nabla \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) := -\frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left[ \frac{\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q})}{\|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}(t_i)}} + \mathbf{u} \right].$$

Let  $\nabla^2 \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}_0)$  be the matrix whose  $i$ -th row is comprised of the gradient of  $\nabla \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$  with respect to the  $i$ -th coordinate of  $\mathbf{q}$ . Also, let us denote by  $\nabla^3 \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$  the third order tensor whose  $(i, j, k)$ -th entry is the third order derivative of  $\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$  with respect to the  $i, j$  and  $k$ -th coordinate of  $\mathbf{q}$ . We have the following proposition.

PROPOSITION 1. Under assumptions (A1)-(A3) and (B1)-(B4), the following hold:

(P1) There exists a  $p$ -variate Gaussian random variable  $\mathbf{Z}(\mathbf{q}_0)$  with mean parameter  $\eta_t(\mathbf{q}_0)$  and dispersion matrix  $\Omega_t(\mathbf{q}_0)$ , such that

$$\sup_{\|\mathbf{u}\| \leq 1} \sup_{f \in \mathcal{F}} \left\| \mathbb{E} \left[ f \left( (pb_n)^{-1} \nabla \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}_0) \right) \right] - \mathbb{E} [f(\mathbf{Z}(\mathbf{q}_0))] \right\| = \mathcal{O}(p^2/n)$$

with probability tending to 1, where  $\mathcal{F}$  is the set of bounded thrice differentiable functions from  $\mathbb{R}^p$  to  $\mathbb{R}$  with a bounded derivative up to the third order.

(P2) There exists a collection of positive definite matrices  $\{\Psi_t(\mathbf{q}_0)\}$  for all possible  $\mathbf{q}_0 \in \mathcal{Q}_p$  as a minimizer of (4) with respect to the spatial points  $\mathcal{S}_p$ , such that

$$\sup_{\mathbf{q}_0 \in \mathcal{Q}_p} \left\| \nabla^2 \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}_0) - \Psi_t(\mathbf{q}_0) \right\| \xrightarrow{P} 0.$$

(P3) There exists  $M > 0$  (not depending on  $n$ ) such that

$$\sup_{\mathbf{q} \in \mathcal{Q}_p} \left\| \nabla^3 (\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})) \right\| \leq M.$$

The three parts (P1)-(P3) embodying Proposition 1, whose proof is deferred to Appendix A, mirror each of the usual necessary steps to demonstrating the asymptotic normality of  $M$ -estimators (cf. Huber, 2011). The asymptotic development for the proposed estimator of the conditional quantile of  $\mathbf{Y}$ , given that  $\mathbf{X}(t) = \mathbf{x}$  for any fixed  $t \in \mathcal{T}$ , stays within this remit. Although general in their principle, the innovation in Proposition (P1) is that it amplifies complementary works on quantile regression by enabling the increasing dimension  $p = p_n \rightarrow \infty$ , as  $n \rightarrow \infty$ . The simpler i.i.d. version analogous to this proposition can be traced back to Portnoy (1986), and contemporary related results accommodating a changing  $p_n$  can be found in Das and Lahiri (2020). Our infinite-dimensional framework shall build on a combination of (pre-asymptotic) moment-type results approached by a functional central limit theorem for martingale differences (cf. Belloni and Oliveira, 2018).

In order to lessen the notational burden throughout the remainder of this paper, we employ the substitutions outlined in the following:

$$(12) \quad \boldsymbol{\eta}(t_i) = \mathbb{E} [\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q}_0) \mid \mathcal{G}_{t_i-}] = \mathbf{K}(t_i) (\boldsymbol{\mu}(t_i) - \mathbf{q}_0),$$

$$(13) \quad \mathbf{V}_1(t_i) = \text{Var} [\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q}_0) \mid \mathcal{G}_{t_i-}] = \mathbf{K}(t_i) \boldsymbol{\Sigma}(t_i) \mathbf{K}(t_i),$$

$$(14) \quad \mathbf{V}_2(t_i) = \left( \frac{1}{\|\boldsymbol{\eta}(t_i)\|} \mathbf{I}_p - \frac{\boldsymbol{\eta}(t_i) \boldsymbol{\eta}(t_i)^\top}{\|\boldsymbol{\eta}(t_i)\|^3} \right),$$

where  $\mathbf{I}_p$  denotes the identity matrix of order  $p$ .

With the key quantities  $\boldsymbol{\eta}_t(\mathbf{q})$ ,  $\boldsymbol{\Omega}_t(\mathbf{q})$  and  $\boldsymbol{\Psi}_t(\mathbf{q})$ , encapsulating the mean and covariance that characterizes the dependence structure of the finite-dimensional distributions associated with the limiting gaussian process for the proposed quantile estimation, namely

$$(15) \quad \boldsymbol{\eta}_t(\mathbf{q}_0) := \frac{1}{npb_n} \sum_{i=1}^n \mathbf{K}(t_i) \left[ \frac{\boldsymbol{\mu}(t) - \mathbf{q}_0}{\|\boldsymbol{\mu}(t) - \mathbf{q}_0\|} - \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} \right]$$

$$(16) \quad \boldsymbol{\Omega}_t(\mathbf{q}_0) := \frac{1}{n^2 p^2 b_n^2} \sum_{i=1}^n \mathbf{K}(t_i) \mathbf{V}_2(t_i) \mathbf{V}_1(t_i) \mathbf{V}_2(t_i)^\top \mathbf{K}(t_i)$$

$$(17) \quad \boldsymbol{\Psi}_t(\mathbf{q}_0) := \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \mathbf{V}_2(t_i) \mathbf{K}(t_i),$$

we arrive at the main result in this paper, stated in the following theorem.

**THEOREM 4.2.** *Define  $\mathbf{q}_0 := Q(\mathbf{u}, \mathbf{x})$ . Under the assumptions (A1)-(A3) and (B1)-(B4), with  $\mathcal{F}$  from (P1), it holds that*

$$\sup_{\|\mathbf{u}\| < 1} \sup_{f \in \mathcal{F}} \left\| \mathbb{E} \left( f \left( (pb_n)^{-1} (\hat{q}_n(\mathbf{u}, \mathbf{x}) - \mathbf{q}_0) \right) \right) - \mathbb{E} \left( f \left( -\boldsymbol{\Psi}_t(\mathbf{q}_0)^{-1} \mathbf{Z}(\mathbf{q}_0) \right) \right) \right\| \xrightarrow{P} 0$$

as  $n \rightarrow \infty$ , where  $\mathbf{Z}(\mathbf{q}_0)$  is a  $p$ -variate Gaussian random vector with mean vector  $\boldsymbol{\eta}_t(\mathbf{q}_0)$  and dispersion matrix  $\boldsymbol{\Omega}_t(\mathbf{q}_0)$ .

At this point, we tactically turn focus to sketching a proof for Proposition 1 since this forms the basis for the proof of Theorem 4.2. For the first part in reference, we start by noting that the challenge in obtaining the asymptotic distribution of  $(pb_n)^{-1} \nabla \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}_0)$  stems from the non-linearity of the normalized term  $(\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q}_0)) / \|\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q}_0)\|$  with respect to the random variables  $\mathbf{Y}(t_i)$ . A standard trick to thwart this issue is to employ a suitable Taylor's expansion for obtaining a sufficiently sharp approximation to a linear functional of  $\mathbf{Y}(t_i)$  or  $\boldsymbol{\xi}(t_i)$ . Indeed, both expressions (12)-(14) have already been carved out in such a way as to become amenable to such an expansion: writing  $\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q}_0) = \boldsymbol{\eta}(t_i) + \mathbf{V}_1^{1/2}(t_i)\boldsymbol{\xi}(t_i)$ , application of Taylor's expansion on the errors  $\xi_{tj}$  thus precipitates the development

$$\frac{\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q}_0)}{\|\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q}_0)\|} = \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} + \mathbf{V}_2(t_i) \mathbf{V}_1^{1/2}(t_i) \boldsymbol{\xi}(t_i) + O \left( \frac{\|\mathbf{V}_1^{1/2}(t_i)\|}{\|\boldsymbol{\eta}(t_i)\|^3} \|\boldsymbol{\xi}(t_i)\|^2 \right),$$

i.e.,

$$\begin{aligned} \nabla \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}_0) &= \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} + \mathbf{u} \right) + \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \mathbf{V}_2(t_i) \mathbf{V}_1^{1/2}(t_i) \boldsymbol{\xi}(t_i) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \mathcal{O}(\|\mathbf{K}(t_i)\| \|\boldsymbol{\eta}(t_i)\|^{-3} \|\mathbf{V}_1^{1/2}(t_i)\| \|\boldsymbol{\xi}(t_i)\|^2). \end{aligned}$$

The proof now follows in three steps. Owing to Lemma 3.1, the first term can be shown to converge to a constant which constitutes the bias. Next, application of a bespoke central limit theorem for the multidimensional martingale provided by Belloni and Oliveira (2018), ensures that the asymptotic weak limit to the second term is a gaussian random variable with mean zero. Finally, by assumption (B4), the third term vanishes to zero at the rate

$\mathcal{O}(\max\{p^{-9/4}b_n^{-2}, p^{-3/2}b_n^{-1}\})$ . Further details underpinning this proof for Proposition (P1) can be found in Appendix A.1.

Proposition (P2), whose detailed proof is deferred to Appendix A.2, is outline as follows: by computing the second order derivative of the objective function  $\mathcal{M}_{u,n}^{(p)}(\mathbf{q})$ , we find that

$$\begin{aligned} \nabla^2 \mathcal{M}_{u,n}^{(p)}(\mathbf{q}) = \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) & \left[ \frac{I_p}{\|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}(t_i)}} \right. \\ & \left. - \frac{\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q})(\mathbf{Y}(t_i) - \mathbf{q})^\top \mathbf{K}(t_i)}{\|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}(t_i)}^3} \right] \mathbf{K}(t_i). \end{aligned}$$

Upon splitting the above expression into a sum of martingale difference terms and a remainder term, the former (i.e. sum) normalized appropriately, can be shown to converge to zero via application of a martingale central limit theorem (see Belloni and Oliveira, 2018). The dominant component in the remainder term boils down to  $\Psi_t(\mathbf{q}_0)$  with everything else being asymptotically negligible. This is achieved through a suitable Taylor's expansion.

Finally, regarding Proposition (P3), we find it convenient and useful to prove the stronger statement that  $\nabla^2 \mathcal{M}_{u,n}^{(p)}(\mathbf{q})$  is Lipschitz with a Lipschitz constant  $M$  independent of  $n$  and  $p$ . Starting with assumption (A2), we note that since  $\mathcal{Q}_p$  is a compact subset of  $\mathcal{U}_p$ ,  $\|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}(t_i)} \geq \lambda_p(\mathbf{K}(t_i))\delta$  for all  $\mathbf{q} \in \mathcal{Q}_p$ , where  $\lambda_p(\mathbf{A})$  is the smallest eigenvalue of the matrix  $\mathbf{A}$ . Since the kernel matrix is positive definite and has bounded entries, it readily follows that  $\nabla^2 \mathcal{M}_{u,n}^{(p)}(\mathbf{q})$  is a Lipschitz function of  $\mathbf{q}$  with a Lipschitz constant  $(\delta^{-1} \sup_{t_i \in \mathcal{T}_p} \lambda_1(\mathbf{K}(t_i)) / \lambda_p(\mathbf{K}(t_i)))^{-3}$  where  $\lambda_1(\mathbf{A})$  is the largest eigenvalue of the matrix  $\mathbf{A}$ . By assumption (B3), this quantity remains bounded.

**5. Statistical inference: interval estimation and hypotheses testing.** While the estimation of quantile curves yielding from (6) possesses several desirable qualities, including the consistency and asymptotic normality, the underpinning convergence is pointwise in the temporal domain, that is, it is obtained for each value of the covariate process  $\{\mathbf{X}(t) = \mathbf{x}\}$ . From the practical stance, an interesting approach resides in obtaining simultaneous confidence band. Inevitably, this will require extending the asymptotic properties of the estimator to the entire temporal domain, uniformly on the index range of  $\mathbf{X}(\cdot)$ . The means to achieve this has a number of parallels with the techniques exploited in Zhao and Wu (2008).

To this end, we shall consider a set of values for the covariates, sufficiently apart in order that at every time-point  $t$ , the kernel estimator is affected by at most one of the values in that set. This is to ensure that no correlation between estimates are at play for sufficiently distant time-points. Upon the already obtained asymptotic normality results, this consideration opens way to making uniform statistical inference possible: asymptotic behavior of the proposed nonparametric estimator's sequence across multiple time-points, uniformly, is now within reach. The ultimate aim is to devise meaningful asymptotic results pertaining to partial maxima of this sequence of estimators by drawing on extreme value theory.

We define  $\chi_n = \{\mathbf{X}(t) : t \in \mathcal{T}_n\} \subseteq \chi$  as the range of the covariate process  $\{\mathbf{X}(t)\}$ . We assume that, for any  $\mathbf{x}, \mathbf{x}' \in \chi_n$ ,  $\|\mathbf{x} - \mathbf{x}'\| > 2b_n \rightarrow 0$  thus implying that  $\chi_n$  becomes dense within  $\chi$  as  $n$  tends to infinity. The cardinality of  $\chi_n$  is encapsulated in  $m_n$ .

**5.1. Simultaneous confidence bounds.** For the sake of simplicity, the results comprising in this section refrain to  $u$ -quantiles with  $u = (2\tau - 1)\mathbf{1}/\sqrt{p}$  for some  $\tau \in (0, 1)$ . This framework essentially renders quantiles that spread evenly among the considered spatial locations. We shall denote the corresponding quantile estimator by  $\hat{q}_n(\tau, \mathbf{x})$  (or sometimes simply by  $\hat{q}(\tau, \mathbf{x})$ )

whereas the true targeted quantile is denoted by  $Q(\tau, \mathbf{x})$ . In spite of this slight restriction to the framework for inference, all results derived in this section can be adapted to incorporate the more general multivariate case as before, in a relatively straightforward manner.

By construction, it follows from the definition of  $\chi_n$  that, for any two points  $\mathbf{x}_{i_1}, \mathbf{x}_{i_2} \in \chi_n$ ,  $i_1 \neq i_2$ , and every  $\delta > 0$ , there exists  $n_0 \in \mathbb{N}$  such that  $\|\mathbf{x}_{i_1} - \mathbf{x}_{i_2}\| < 2\delta$ ,  $i_1, i_2 > n_0$ . Hence, for every  $t \in \mathcal{T}_n$ ,  $\|\mathbf{X}(t) - \mathbf{x}_j\| < \delta$ ,  $j = i_1, i_2$ , which results from the triangle inequality. Since  $\mathcal{K}$  has bounded support in the unit hypersphere  $B^d$ , it thus follows that the cross products of the weights intervening in the objective function resolve to zero, i.e.

$$(18) \quad \mathbf{K}(\mathbf{X}(t), \mathbf{x}_{i_1}) \mathbf{K}(\mathbf{X}(t), \mathbf{x}_{i_2}) = c_n^{-2} \mathcal{K} \left( \frac{\mathbf{X}(t) - \mathbf{x}_{i_1}}{b_n} \right) \mathcal{K} \left( \frac{\mathbf{X}(t) - \mathbf{x}_{i_2}}{b_n} \right) = 0$$

(the reader may wish to recall the discussion on  $c_n$  as part of Section 3.2). Equation (18) lays bare that fixing two quantile levels  $\tau$  and  $\tau'$  ( $0 < \tau, \tau' < 1$ ), so that objective function if optimised for  $\hat{q}(\tau, \mathbf{x}_{i_1})$  contains the terms with weights  $\mathbf{K}(\mathbf{X}(t), \mathbf{x}_{i_1})$  for some  $t \in \mathcal{T}_n$ , whereas for the sequence of estimators  $\hat{q}(\tau', \mathbf{x}_{i_2})$  these weights amount to  $\mathbf{K}(\mathbf{X}(t), \mathbf{x}_{i_2})$ , results in the two objective functions (both for  $\hat{q}(\tau, \mathbf{x}_{i_1})$  and  $\hat{q}(\tau', \mathbf{x}_{i_2})$ ) being uncorrelated, thus implying that the estimators  $\hat{q}(\tau, \mathbf{x}_{i_1})$  and  $\hat{q}(\tau', \mathbf{x}_{i_2})$  are also uncorrelated.

[CN: Please check whether the above makes sense to you. Thanks! ]

Therefore, by taking a multitude of points on the set  $\chi_n = \{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_{m_n}}\}$ , we obtain from Theorem 4.2 that the estimator  $\hat{q}_n(\mathbf{x}, \tau)(s)$  becomes asymptotically a Gaussian process on both temporal and spatial domains, for every  $\mathbf{x} \in \chi_n$  and  $s \in \mathcal{S}$ . More concretely, on the finite set of spatial points  $S_d = \{s_1, s_2, \dots, s_d\}$  and with unit vector  $\mathbf{l} \in \mathbb{R}^d$ , we have for every  $\mathbf{x} \in \chi_n$ , as  $n \rightarrow \infty$ , that

$$(19) \quad R_n(\mathbf{x}, \mathbf{l}, S_d) = (pb_n)^{-1} \frac{\mathbf{l}^\top (\hat{q}_n(\tau, \mathbf{x})|_{S_d} - Q(\tau, \mathbf{x})|_{S_d} - \boldsymbol{\mu}_q(\mathbf{x})|_{S_d})}{\sqrt{\mathbf{l}^\top \boldsymbol{\Sigma}_q(\mathbf{x})|_{S_d} \mathbf{l}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

We highlight that  $\hat{q}_n(\mathbf{x}, \tau)|_{S_d}$  and  $Q(\mathbf{x}, \tau)|_{S_d}$  respectively denote  $\hat{q}_n(\mathbf{x}, \tau)$  and  $Q(\mathbf{x}, \tau)$  restricted to the spatial locations in  $S_d$ . Letting  $q_0 = Q(\tau, \mathbf{x})$ , then  $\boldsymbol{\mu}_q(\mathbf{x})|_{S_d}$  and  $\boldsymbol{\Sigma}_q(\mathbf{x})|_{S_d}$  respectively denote mean and covariance structure, evaluated at the spatial locations in the set  $S_d$ , of the Gaussian process  $\{G(\mathbf{x}, s) : \mathbf{x} \in \chi, s \in \mathcal{S}\}$  characterized by

$$(20) \quad \begin{aligned} \mathbb{E}[G(\mathbf{x}, s)] &= \boldsymbol{\Psi}_{\mathbf{X}(t)=\mathbf{x}}^{-1}(\mathbf{q}_0) \boldsymbol{\eta}_{\mathbf{X}(t)=\mathbf{x}}(\mathbf{q}_0) \\ \text{Cov}(G(\mathbf{x}, s), G(\mathbf{x}', s)) &= \begin{cases} \boldsymbol{\Psi}_{\mathbf{X}(t)=\mathbf{x}}^{-1}(\mathbf{q}_0) \boldsymbol{\Omega}_{\mathbf{X}(t)=\mathbf{x}}(\mathbf{q}_0) \boldsymbol{\Psi}_{\mathbf{X}(t)=\mathbf{x}}^{-1}(\mathbf{q}_0), & \mathbf{x} = \mathbf{x}'; \\ 0, & \mathbf{x} \neq \mathbf{x}'. \end{cases} \end{aligned}$$

where the quantities  $\boldsymbol{\eta}(\mathbf{q}_0)$ ,  $\boldsymbol{\Omega}(\mathbf{q}_0)$  and  $\boldsymbol{\Psi}(\mathbf{q}_0)$  are as defined in (15)-(17).

We now progress to obtaining the asymptotic behaviour of the supremum of  $G(\mathbf{x}, s)$  over the temporal index through  $\sup_{\mathbf{x} \in \chi_n} \hat{q}_n(\tau, \mathbf{x})(s)$  for any  $s \in \mathcal{S}$ . From the case in point that  $\hat{q}_n(\tau, \mathbf{x}_{i_1})$  and  $\hat{q}_n(\tau', \mathbf{x}_{i_2})$  are asymptotically uncorrelated, the next step is to apply the extreme value theorem upon the univariate martingale difference sequence  $R_n(\mathbf{x}_{i_k}, \mathbf{l}, S_d)$ ,  $k = 0, 1, \dots, m_n - 1$ . Owing to the rich theory of extreme values, there exist normalizing functions  $C_t$  specific to the folded normal distribution, i.e.

$$(21) \quad C_t(z) := \frac{1}{\sqrt{2 \log t}} \left( z - \frac{\log \log t + \log(4\pi)}{2} \right)$$

(see Resnick, 1987, p.72), such that the weak convergence to the Gumbel distribution function is attained in the following sense: for all  $z \in \mathbb{R}$ ,

$$(22) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{\mathbf{x} \in \chi_{m_n}} |R_n(\mathbf{x}, \mathbf{l}, S_d)| \leq C_{m_n}(z + 2 \log m_n) \right) = \exp\{-e^{-z}\}.$$



Next, we consider the maximum eigenvalue of the matrix  $M$ , denoted as  $\lambda_1(M)$ , for establishing the identification

$$\sup_{\|l\|=1} \frac{l^\top A l}{l^\top B l} = \lambda_1 \left( B^{-1/2} A B^{-1/2} \right),$$

where  $A$  and  $B$  are positive definite matrices. In conjunction with the limiting distribution statement in (22), we subsequently find that

$$(23) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{x \in \chi_{m_n}} \frac{1}{pb_n} \lambda_1 \left[ \Sigma_q(x)^{-1/2} |_{S_d} e(x) |_{S_d} e(x) |_{S_d}^\top \Sigma_q(x)^{-1/2} |_{S_d} \right] \right. \\ \left. < C_{m_n}(z) + \sqrt{2 \log m_n} \right) = \exp\{-e^{-z}\},$$

by taking supremum over all possible unit vectors  $l$  and whereby  $e(x)|_{S_d} = (\hat{q}_n(\tau, x)|_{S_d} - Q(\tau, x)|_{S_d} - \mu_q(x)|_{S_d})$ . Hence, under the assumptions (A1)-(A3) and (B1)-(B4), with a judicious choice of  $\chi_{m(n)}$ , i.e., so long as  $\log m_n = o((pb_n)^{-2})$  is satisfied, the sequence of estimators  $\hat{q}_n(\tau, x)$  can be guaranteed to fulfill (22) and (23) on a finite set of spatial locations  $S_d = \{s_1, s_2, \dots, s_d\}$  from the “larger” set  $\mathcal{S}$ . This result, which is along the lines of that in Theorem 2 of Zhao and Wu (2008), serves as basis to introducing simultaneous confidence bands for the  $\tau$ -quantile curves  $Q(\tau, x)$  ascribed to multiple values  $x \in \chi_n$ . Because  $\chi_n$  becomes increasingly dense in  $\chi$  as  $n$  tends to  $\infty$ , an asymptotic confidence band for the quantile curve  $\{Q(\tau, x) : x \in \chi\}$  naturally arises from (22).

Indeed, a contrast-like linear combination of the quantile curves across different locations stems from (22) as follows: with a proper specification of the set  $S_d$  and vector  $l \in \mathbb{R}^d$  so that  $l^\top Q(x, \tau)$  becomes the desired contrast-like quantile curve, we find that the asymptotic simultaneous  $100(1 - \alpha^*)\%$ -confidence bounds for  $l^\top Q(x, \tau)|_{S_d}$  consist of

$$l^\top \hat{q}_n(x, \tau)|_{S_d} - l^\top \hat{\mu}_q(x)|_{S_d} \\ \pm \left[ C_{m_n}(-\log(-\log(1 - \alpha^*/2))) + \sqrt{2 \log m_n} \right] (pb_n) \left( l^\top \hat{\Sigma}_q(x)|_{S_d} l \right)^{1/2},$$

with embedded consistent estimators  $\hat{\mu}_q(x)$  and  $\hat{\Sigma}_q(x)$  for the bias term  $\mu_q(x)$  and covariance  $\Sigma_q(x)$ , respectively. We note that the symmetry in the confidence interval is justified by the same limiting extreme value distribution of  $\inf_x R_n$  and  $-\sup_x |R_n|$ , a remark that will prove useful to constructing one-sided hypothesis testing procedures.

**5.2. Two tests.** The statement in (23) can be adapted to the purpose of hypothesis testing, namely in order to validate structural assumptions on the quantile curves across all locations simultaneously. For instance, given a null hypothesis about the structural restriction of the quantile curve  $Q(\tau, x)$  at some specific quantile  $\tau \in [0, 1]$ , we can calculate the observed value of the test statistic

$$(24) \quad e_0(x)|_{S_p} = (\hat{q}_n(\tau, x)|_{S_p} - \hat{q}_0(\tau, x)|_{S_p} - \mu_q(x)|_{S_p})$$

where  $\hat{q}_0(\tau, x)$  stands the estimate of  $Q(\tau, x)$  under the postulated null hypothesis  $H_0$  regarding a stable regime. Then, the null hypothesis is rejected if (24) exhibits large oscillation, i.e., at the nominal level of the test  $\alpha^*$ ,  $H_0$  is rejected if

$$\frac{1}{pb_n} \sup_{x \in \chi_n} \lambda_1 \left( \hat{\Sigma}_q(x)^{-1/2} |_{S_p} e_0(x) |_{S_p} e_0(x) |_{S_p}^\top \hat{\Sigma}_q(x)^{-1/2} |_{S_p} \right)$$

exceeds the critical value  $c_{\alpha^*} = C_{m_n}(-\log(-\log(1 - \alpha^*))) + \sqrt{2\log m_n}$ . Henceforth, we define  $\mathbf{u}_\tau := (2\tau - 1)\mathbf{1}_p/\sqrt{p}$ .

A structural assumption that is often put to testing is that of homogeneity of the covariate effect across the spatial domain, which if true, automatically narrows down the spatio-temporal structure of the data to a temporal one. This is statistically parlayed in the null hypothesis  $H_0^{(1)} : Q(\tau, \mathbf{x}) = \beta(\mathbf{x})\mathbf{u}_\tau$ , for all  $\tau \in [0, 1]$ . Under  $H_0^{(1)}$ , the estimator  $\hat{\beta}(\mathbf{x})$  arises from (9) as solution of:

$$\sum_{i=1}^n \mathbf{K}(\mathbf{x}, \mathbf{X}(t_i)) \left[ \frac{\mathbf{K}(\mathbf{x}, \mathbf{X}(t_i))(\mathbf{Y}(t_i) - \hat{\beta}(\mathbf{x})\mathbf{u}_\tau)}{\|\mathbf{K}(\mathbf{x}, \mathbf{X}(t_i))(\mathbf{Y}(t_i) - \hat{\beta}(\mathbf{x})\mathbf{u}_\tau)\|} + \mathbf{u}_\tau \right] = 0,$$

and the null hypothesis is rejected for large values of the test statistic, i.e.,

$$\frac{1}{pb_n} \sup_{\mathbf{x} \in \chi_n} \lambda_1 \left( \hat{\Sigma}_q(\mathbf{x})^{-1/2} |_{\mathcal{S}_p} \mathbf{e}_1(\mathbf{x}) |_{\mathcal{S}_p} \mathbf{e}_1(\mathbf{x}) |_{\mathcal{S}_p}^\top \hat{\Sigma}_q(\mathbf{x})^{-1/2} |_{\mathcal{S}_p} \right) > c_{\alpha^*},$$

on the relevant set  $\mathcal{S}_p$  of spatial locations, where  $\mathbf{e}_1(\mathbf{x})|_{\mathcal{S}_p} = (\hat{q}_n(\mathbf{u}, \mathbf{x}_\tau)|_{\mathcal{S}_p} - \hat{\beta}(\mathbf{x})\mathbf{u}_\tau - \hat{\mu}_q(\mathbf{x})|_{\mathcal{S}_p})$  can be defined.

The second testing procedure with practical importance under the realm of (2) ascribes time to a covariate, i.e.  $\mathbf{X}(t) = t$ . The hypothesis of homogeneity over time then translates into  $H_0^{(2)} : Q(\mathbf{u}_\tau, t)/Q(\mathbf{u}_\tau, 1) = t$ , for all  $t \in (0, 1]$  and  $\tau \in [0, 1]$ . At the significance level  $\alpha^*$ ,  $H_0^{(2)}$  is rejected if

$$\frac{1}{pb_n} \sup_{t \in (0, 1]} \lambda_1 \left( \hat{\Sigma}_q(t)^{-1/2} |_{\mathcal{S}_p} \mathbf{e}_2(t) |_{\mathcal{S}_p} \mathbf{e}_2(t) |_{\mathcal{S}_p}^\top \hat{\Sigma}_q(t)^{-1/2} |_{\mathcal{S}_p} \right) > c_{\alpha^*},$$

with associated  $\mathbf{e}_2(t)|_{\mathcal{S}_p} = (\hat{q}_n(\mathbf{u}_\tau, t)|_{\mathcal{S}_p} - t\hat{q}_n(\mathbf{u}_\tau, 1)|_{\mathcal{S}_p} - \hat{\mu}_q(\mathbf{x})|_{\mathcal{S}_p})$ .

**6. Proofs.** This section contains the proofs of the key theorems in this paper as well as of their ancillary results. We begin by addressing Lemma 3.1, which consolidates the estimating equation for the proposed nonparametric quantile estimator.

**Proof of Lemma 3.1:** Under assumption (A3),  $\mathbf{K}(t) = \mathbf{K}(\mathbf{x}, \mathbf{X}(t))$  is a positive definite matrix for any  $t \in \mathcal{T}$ . Hence,  $\Phi_K(\mathbf{u}, \mathbf{Y}(t) - \mathbf{q}) = \|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)}$  is a strictly convex function of  $\mathbf{q}$ . Since  $\hat{\mathbf{q}} = \hat{q}_n(\mathbf{u}, \mathbf{x})$  is the minimizer of  $\mathcal{M}_{\mathbf{u}, n}^{(p)}(\mathbf{q}) = \frac{1}{n} \sum_{i=1}^n \Phi_K(\mathbf{u}, \mathbf{Y}(t_i) - \mathbf{q})$ , then for every  $\mathbf{h} \in \mathbb{R}^d$ ,

$$\begin{aligned} & \lim_{\epsilon \downarrow 0} \frac{1}{n\epsilon} \sum_{i=1}^n [\Phi_K(\mathbf{u}, \mathbf{Y}(t_i) - \hat{\mathbf{q}} + \epsilon \mathbf{h}) - \Phi_K(\mathbf{u}, \mathbf{Y}(t_i) - \hat{\mathbf{q}})] \geq 0 \\ \implies & \frac{1}{n} \sum_{i=1}^n \lim_{\epsilon \downarrow 0} \frac{\|\mathbf{Y}(t_i) - \hat{\mathbf{q}} + \epsilon \mathbf{h}\|_{\mathbf{K}(t_i)} - \|\mathbf{Y}(t_i) - \hat{\mathbf{q}}\|_{\mathbf{K}(t_i)}}{\epsilon} + \lim_{\epsilon \downarrow 0} \frac{\langle \mathbf{u}, \epsilon \mathbf{h} \rangle_{\mathbf{K}(t_i)}}{\epsilon} \geq 0 \\ \implies & \frac{1}{n} \sum_{i=1}^n \left[ \frac{\langle \mathbf{Y}(t_i) - \hat{\mathbf{q}}, \mathbf{h} \rangle_{\mathbf{K}(t_i)}}{\|\mathbf{Y}(t_i) - \hat{\mathbf{q}}\|_{\mathbf{K}(t_i)}} + \langle \mathbf{u}, \mathbf{h} \rangle_{\mathbf{K}(t_i)} \right] \geq 0. \end{aligned}$$

Using the same steps with  $(-\mathbf{h})$  in place of  $\mathbf{h}$  we arrive (9). Similarly, (10) stems from (4).  $\square$

**Proof of Theorem 4.1:** Defining  $\hat{\mathbf{q}} = \hat{\mathbf{q}}_n(\mathbf{u}, \mathbf{x})$  and  $\mathbf{q}_0 = Q(\mathbf{u}, \mathbf{x})$ , we write the chain of inequalities:

$$\begin{aligned} \mathcal{M}_{\mathbf{u}}^{(p)}(\hat{\mathbf{q}}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}_0) &\leq \mathcal{M}_{\mathbf{u}}^{(p)}(\hat{\mathbf{q}}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}_0) - \mathcal{M}_{\mathbf{u},n}^{(p)}(\hat{\mathbf{q}}) + \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}_0) \\ &\leq 2 \sup_{\mathbf{q} \in \mathcal{Q}_p} |\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q})|, \end{aligned}$$

where  $\mathcal{Q}_p \subseteq \mathbb{R}^p$  is a set containing both  $\mathbf{q}_0$  and the sequence of estimators  $\hat{\mathbf{q}}_n(\mathbf{u}, \mathbf{x})$ , thus enabling

$$\begin{aligned} &\mathbb{P}(\|\hat{\mathbf{q}} - \mathbf{q}_0\| > Ap^2/n) \\ &\leq \mathbb{P}\left(\|\hat{\mathbf{q}} - \mathbf{q}_0\| > Ap^2/n, \mathcal{M}_{\mathbf{u}}^{(p)}(\hat{\mathbf{q}}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}_0) < Bp^2/n\right) \\ &\quad + \mathbb{P}\left(\mathcal{M}_{\mathbf{u}}^{(p)}(\hat{\mathbf{q}}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}_0) \geq Bp^2/n\right) \\ (25) \quad &\leq \mathbb{P}\left(\|\hat{\mathbf{q}} - \mathbf{q}_0\| > Ap^2/n, \mathcal{M}_{\mathbf{u}}^{(p)}(\hat{\mathbf{q}}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}_0) < Bp^2/n\right) \\ (26) \quad &\quad + \mathbb{P}\left(\sup_{\mathbf{q} \in \mathcal{Q}_p} |\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q})| \geq Bp^2/2n\right). \end{aligned}$$

Owing to the convexity lemma by Pollard (1991), the objective function  $\mathcal{M}_{\mathbf{u}}$  as part of (4) is convex on  $\mathcal{Q}_p$ . Consequently, the sequence of conditional regression-quantile estimators  $\hat{\mathbf{q}}_n$  along  $\mathbf{u}$  is well-defined and the true quantile  $\mathbf{q}_0$  is a well-separated minimizer of  $\mathcal{M}_{\mathbf{u}}$ , that is, there exists a  $\mathbf{q}_0 \in \mathcal{U}_p$  for which  $\inf_{\mathbf{q}: \|\mathbf{q} - \mathbf{q}_0\| \geq \delta} (\mathcal{M}_{\mathbf{u}}(\mathbf{q}) - \mathcal{M}_{\mathbf{u}}(\mathbf{q}_0)) > 0$ , for arbitrarily small  $\delta > 0$ . This ensures the boundedness of the first term in the above sum. Through application of the triangle inequality in conjunction with the Cauchy-Schwartz inequality, we obtain:

$$\mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}) \geq \|\hat{\mathbf{q}} - \mathbf{q}_0\|(1 - \|\mathbf{u}\|_{\mathcal{S}_p}),$$

which implies in turn that

$$\mathbb{P}\left(\|\hat{\mathbf{q}} - \mathbf{q}_0\| > Ap^2/n, \mathcal{M}_{\mathbf{u}}^{(p)}(\hat{\mathbf{q}}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}_0) < Bp^2/n\right) \leq \mathbb{P}(\|\mathbf{u}\|_{\mathcal{S}_p} > 1 - B/A).$$

If  $\|\mathbf{u}\| < 1$ , i.e., if not addressing an extreme quantile or the case of bounded support, it is possible to fix  $B = A\epsilon$  with arbitrarily small  $\epsilon > 0$ , so as to ensure that  $\|\mathbf{u}\|_{\mathcal{S}_p} \leq (1 - \epsilon)$ , in which case the above upper bound becomes 0. This takes care of (25). Consequently, the gist of the proof is in verifying the uniform convergence of uniform convergence embedded in (26). This is done as follows. From the subgaussian property of  $\mathbf{Y}(t)$  together with the law of large numbers, it results that

$$(27) \quad \left| \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) - \mathbb{E}[\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)}] \right| = \mathcal{O}_p(\sqrt{p/n})$$

In order to connect the weighted expectation affecting the kernel  $\mathbf{K}(t)$  with  $\mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q})$ , we invoke the law of iterated expectations upon the conditioning random variable  $R = \|\mathbf{X}(t) - \mathbf{x}\|$ :

$$\mathbb{E}[\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)}] = \mathbb{E}[\mathbb{E}[\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)} | R]].$$

There is a one-to-one mapping between  $R$  and the continuous measurable functional that is the kernel  $\mathbf{K}(t)$ . In particular, the event  $R = 0$ , amounts to  $\mathbf{K}(t) = \mathbf{I}_p$  due to  $\mathcal{K}(0) = c_n \mathbf{I}_p$ . Whence, there exists  $\delta_R > 0$  such that for every  $R < \delta_R$ , all eigenvalues of the matrix

$(\mathbf{K}(t) - \mathbf{I}_p)$  are bounded by  $\sqrt{p/n}$ , rendering valid the chain of inequalities:

$$\begin{aligned}
& \left| \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)}] - \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\| + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle \mid \mathbf{X}(t) = \mathbf{x}] \right| \\
&= \left| \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)} \mid R] - \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\| + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle \mid R=0] \right| \\
&= \left| \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)} \mid R \leq \delta_R] \mathbb{P}(R \leq \delta_R) \right. \\
&\quad \left. + \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)} \mid R > \delta_R] \mathbb{P}(R > \delta_R) \right. \\
&\quad \left. - \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\| + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle \mid R=0] \right| \\
&\leq \left| \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} - \|\mathbf{Y}(t) - \mathbf{q}\| \mid R \leq \delta_R] \right| + \left| \mathbb{E} [\langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)} - \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle \mid R \leq \delta_R] \right| \\
&\quad \left| \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)} \mid R > \delta_R] \right|
\end{aligned}$$

The first term can be bounded by the product of the absolute value of the maximum eigenvalue of  $(\mathbf{K}(t) - \mathbf{I}_p)$  and the norm of  $\|\mathbf{Y}(t) - \mathbf{q}\|$ . When  $R \leq \delta_R$ , combining with the subgaussian behaviour of  $\mathbf{Y}(t)$ , the first term can be bounded by  $\mathcal{O}(p/\sqrt{n})$ . For the second term, an application of the Cauchy-Schwartz inequality combined with the fact that  $\|\mathbf{u}\| \leq 1$  yields that it is also bounded by  $\mathcal{O}(p/\sqrt{n})$ . For the third term, note that as the bandwidth  $b_n \rightarrow 0$ , for sufficiently large  $n$ ,  $\delta_R > b_n$ . Since the kernel  $\mathcal{K}$  has bounded support in the unit hypersphere  $B^d$ , it follows that  $\mathbf{K}(t)$  can be made equal to  $\mathbf{0}$  in the third term. Therefore,

$$\left| \mathbb{E} [\|\mathbf{Y}(t) - \mathbf{q}\|_{\mathbf{K}(t)} + \langle \mathbf{u}, \mathbf{Y}(t) - \mathbf{q} \rangle_{\mathbf{K}(t)}] - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q}) \right| = o(p/\sqrt{n}) = o(p^2/n)$$

**[CN: Please check (and then indicate) which of the above  $\mathcal{O}$ -terms is a small  $o$  ]** **[SR: Modified accordingly. ]** Combining this with (27), we obtain that for each fixed  $\mathbf{q} \in \mathbb{R}^p$ ,

$$\|\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) - \mathcal{M}_{\mathbf{u}}^{(p)}(\mathbf{q})\| = \mathcal{O}_p(p^2/n)$$

The uniform convergence now follows if we can show that  $\mathcal{Q}_p$  is a compact set. Thus, it is enough to show that the sequence of estimators  $\mathbf{q}_n$  minimizing the sequence of objective functions  $\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$  are uniformly bounded. Without the loss of generality, consider that  $\mathbf{0}$  belongs to the convex hull of the datapoints  $\{\mathbf{Y}(t_i) : i = 1, 2, \dots, n\}$ , otherwise we can translate all datapoints suitably. By applying the sub-gaussian property of  $\mathbf{Y}(t_i)$  and union bound, it follows that there exists some  $c > 0$  such that

$$\mathbb{P} \left( \sup_{1 \leq i \leq n} \|\mathbf{Y}(t_i)\| > M \right) \leq ne^{-cp}.$$

By choosing sufficiently large  $p = O(\log n)$ , we can now make  $\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{0}) < M$  with probability arbitrarily close to 1, by bounding each  $\mathbf{Y}(t_i)$ . Also, since  $\|\mathbf{u}\| < 1$ , there exists a  $\delta \in (0, 1)$  such that  $\|\mathbf{u}\| < (1 - \delta)$ , and a vector  $\mathbf{v}$  such that  $\|\mathbf{v}\| > 2M\delta^{-1}$ . Consider the objective function at a point  $\mathbf{Y}(t_j) + \mathbf{v}$  for any  $j = 1, 2, \dots, n$ ;

$$\begin{aligned}
\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{Y}(t_j) + \mathbf{v}) &= \frac{1}{n} \sum_{i=1}^n [\|\mathbf{Y}(t_i) - \mathbf{Y}(t_j) - \mathbf{v}\|_{\mathcal{K}(t_i)} + \langle \mathbf{u}, \mathbf{Y}(t_i) - \mathbf{Y}(t_j) - \mathbf{v} \rangle_{\mathcal{K}(t_i)}] \\
&\geq \|\mathbf{v}\|/2 - (1 - \delta)\|\mathbf{v}\|/2, \text{ by an application of Cauchy-Schwartz inequality} \\
&\geq M \geq \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{0})
\end{aligned}$$

By general positioning of these vectors,  $\{\mathbf{Y}(t_j) + \mathbf{v}, \mathbf{Y}(t_j) - \mathbf{v} : j = 1, \dots, n\}$  do not lie in a subspace with a dimension strictly less than  $p$ . Combining this with the convex nature of the objective function  $\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$ , it follows that along the boundary of the convex hull of

these points also, the value of the objective function is more than  $M$ . Using continuity of the objective function, we find that there is a  $(p-1)$ -dimensional hypersphere, independent of the choice of  $n$ , containing the origin  $0$  such that the  $\mathcal{M}_{u,n}^{(p)}(\mathbf{q}) > \mathcal{M}_{u,n}^{(p)}(\mathbf{0})$  for any  $\mathbf{q}$  lying on the boundary of that hypersphere for any  $n$ . Now it follows from the strict convexity of each  $\mathcal{M}_{u,n}^{(p)}(\mathbf{q})$  that they have a local minimum within that hypersphere. Choosing the  $\hat{\mathbf{q}}_n$  as those local minima establishes the boundedness of them.  $\square$

**Proof of Theorem 4.2:** We start by assuming that the Propostions (P1)-(P3) hold.

Consider a Taylor series expansion of  $\nabla \mathcal{M}_{u,n}^{(p)}(\mathbf{q})$  about the true parameter  $\mathbf{q}_0 := Q(\mathbf{u}, \mathbf{x})$  and evaluate at  $\hat{\mathbf{q}} := \hat{\mathbf{q}}_n(\mathbf{u}, \mathbf{x})$ ,

$$\begin{aligned} 0 &= \nabla \mathcal{M}_{u,n}^{(p)}(\hat{\mathbf{q}}) \\ &= \nabla \mathcal{M}_{u,n}^{(p)}(\mathbf{q}_0) + \nabla^2 \mathcal{M}_{u,n}^{(p)}(\mathbf{q}_0)(\hat{\mathbf{q}} - \mathbf{q}_0) + \mathcal{O}(\|\hat{\mathbf{q}} - \mathbf{q}_0\|^2 \|\nabla^3 \mathcal{M}_{u,n}^{(p)}(\mathbf{q})\|) \end{aligned}$$

Denoting  $\nabla^i \mathcal{M}_{u,n}^{(p)}(\mathbf{q}_0) = \Delta_i$ , we obtain that

$$a_n(\hat{\mathbf{q}} - \mathbf{q}_0) = -a_n \Delta_2^{-1} \Delta_1 + \mathcal{O}(\|\Delta_2^{-1}\| p^4/n)$$

where we use the consistency result of  $\hat{\mathbf{q}}$  along with proposition (P3) to bound  $\|\Delta_3\|$ . By proposition (P2), it follows that for sufficiently large  $n$ ,

$$\|\Delta_2^{-1} - \Psi(\mathbf{q}_0)^{-1}\| \leq \frac{2}{\|\Psi(\mathbf{q}_0)\|^2} \|\Psi(\mathbf{q}_0) - \Delta_2\| \leq \frac{Cp^2}{n\|\Psi(\mathbf{q}_0)\|^2}$$

Therefore,

$$\begin{aligned} & |\mathbb{E}(f(a_n(\hat{\mathbf{q}} - \mathbf{q}_0))) - \mathbb{E}(f(\Psi^{-1}(\mathbf{q}_0)Z(\mathbf{q}_0)))| \\ & \leq |\mathbb{E}(f(-a_n \Delta_2^{-1} \Delta_1 + \mathcal{O}(\|\Delta_2^{-1}\| p^4/n))) - \mathbb{E}(f(\Psi^{-1}(\mathbf{q}_0)Z(\mathbf{q}_0)))| \\ & \leq |\mathbb{E}(f(-a_n \Delta_2^{-1} \Delta_1 + \mathcal{O}(\|\Delta_2^{-1}\| p^4/n))) - \mathbb{E}(f(-a_n \Psi^{-1}(\mathbf{q}_0) \Delta_1))| + \\ & \quad |\mathbb{E}(f(-a_n \Psi^{-1}(\mathbf{q}_0) \Delta_1)) - \mathbb{E}(f(\Psi^{-1}(\mathbf{q}_0)Z(\mathbf{q}_0)))| \\ & \leq \mathbb{E}(\|f'(\tilde{\mathbf{q}})\| (\mathcal{O}(\|\Delta_2^{-1}\| p^4/n) + |a_n| \|\Delta_1\| \|\Delta_2^{-1} - \Psi(\mathbf{q}_0)^{-1}\|)) + \mathcal{O}(p^2/n) \end{aligned}$$

where  $\tilde{\mathbf{q}}$  is some vector lying between  $-a_n \Psi^{-1}(\mathbf{q}_0) \Delta_1$  and  $-a_n \Delta_2^{-1} \Delta_1$ , which is obtained through an application of mean value theorem. Here, we also use the fact that the function  $g(x) := \Psi^{-1}(\mathbf{q}_0)f(x)$  is also an element of  $\mathcal{F}$ , hence  $|\mathbb{E}(g(-a_n \Delta_1)) - \mathbb{E}(g(Z(\mathbf{q}_0)))|$  is  $\mathcal{O}(p^2/n)$  by proposition (P1). Along with that, we also note that  $\|\Psi(\mathbf{q}_0)\|$  is positive definite and has bounded entries, hence its eigenvalues are of order  $\mathcal{O}(\sqrt{p})$ . Therefore, we obtain

$$|\mathbb{E}(f(a_n(\hat{\mathbf{q}} - \mathbf{q}_0))) - \mathbb{E}(f(\Psi^{-1}(\mathbf{q}_0)Z(\mathbf{q}_0)))| = \mathcal{O}(p^2/n)$$

Now, note that the bound obtained is independent of the choice of  $f \in \mathcal{F}$ , therefore the above conclusion holds if we replace the left-hand side with the supremum over  $f \in \mathcal{F}$ . Finally, the choice of  $\mathbf{u}$  is restricted within the compact set  $\{\mathbf{u} : \|\mathbf{b}\|^2 \leq 1\}$ , hence taking a further supremum over a finite cover of this set, the result follows.  $\square$

**7. Forecasting of energy demand.** To experimentally verify the correctness of the proposed quantile estimator, we perform simulation studies corresponding to two different setups, one without any explicit spatial information and another setup with proper spatial information included. In both cases, the proposed estimator achieves reasonable prediction accuracy. The details of the simulation studies can be found in Appendix B.

In this section, we present a real data analysis, by considering the Thames Valley Vision (TVV) data presented in [Jacob, Neves and Vukadinović Greetham \(2020\)](#). The dataset was collected as a part of Scottish and Southern Electricity Network (SSEN) of TVV project, funded by UK gas and electricity regulator Ofgem through the Low Carbon Networks Fund and Network Innovation Competition. The project’s overall aim was to monitor and model a typical low voltage network using monitoring in households and substations in order to simulate future realistic demand scenarios. Bracknell, a moderate-sized town west of London, was chosen as it hosts many large companies and the local network, with its urban and rural parts, is representative of much of Britain’s electricity network.

This data set contains profiles for 226 households on half-hourly resolution between 19th July 2015 and 20th September 2015, spanning 9 weeks of data. In our analysis, the first 8 weeks of data are used as a training set while we use our proposed estimator to predict different quantiles of last week’s electricity demand for each of the household. Considering the temporal resolution of the data, we note that the number of observations for every location in the training set is  $n = 56 \times 48 = 2688$ . It follows from the described theory of the proposed estimator, as in Theorem 4.1, that the number of spatial points (226) well exceeds the allowed range of  $\sqrt{n}$  for the estimator to remain asymptotically consistent. Thus, we decide to cluster the households based on the daily average demand profiles into groups of size at most  $52 \approx \sqrt{2688}$ . To determine these clusters, we calculate the pairwise dynamic time warping (DTW) distance ([Jacob, Neves and Vukadinović Greetham, 2020](#)) between any two household’s average daily demand profiles using the FastDTW algorithm ([Gold and Sharir, 2018](#)), and use the Hierarchical clustering to cluster these households. In Figure 1, we show the heatmap for the pairwise DTW distances (log-transformed) between the households. With the restriction of 52 being the maximum size of any cluster, we divide the 226 households into 12 clusters and the households in the figure are arranged according to their cluster assignments. Note that the sizes of the clusters are 21, 52, 42, 17, 10, 42, 28, 4, 2, 4, 2 and 2.

One can clearly see that the pairwise within-cluster distances are relatively smaller than the between-cluster distances. Subsequently, for the main analysis, we assume that the electricity demand profiles of two households residing in two different clusters are independent. This allows us to apply the proposed quantile estimator separately on each of the cluster of households and obtain predictions for the quantile profiles for the 9-th week across all 226 households. For all households within a cluster, we use the same quantile level for ease of implementation and interpretation, and we use the levels of 0.5, 0.75, 0.8, 0.9, 0.95, 0.99 for every case. After estimating the quantile functions from the training data, the quantile levels for different households in the 12 clusters are predicted and Figure 2 demonstrates these predicted values.

It is evident that the estimated quantile has a nice smoothness property due to the continuous and smooth nature of the kernel function, as well as it does not have any quantile crossing issue. As expected, the two quantile curves are closer when the general trend demand is low, but during the peak hours, the quantiles can become extremely different from the usual median trend level. The systematic cyclical pattern which is present in electricity demand due to the daily day-night cycle is also captured in the prediction. Some clusters like clusters 2, 4, 8 and 12 indicate high electricity loads during the weekends. While for most of the clusters, the quantile curves vary, the quantile curves for cluster 9 are very tightly packed together. This is because there are only two households, namely Household 106 and Household 207 lying in



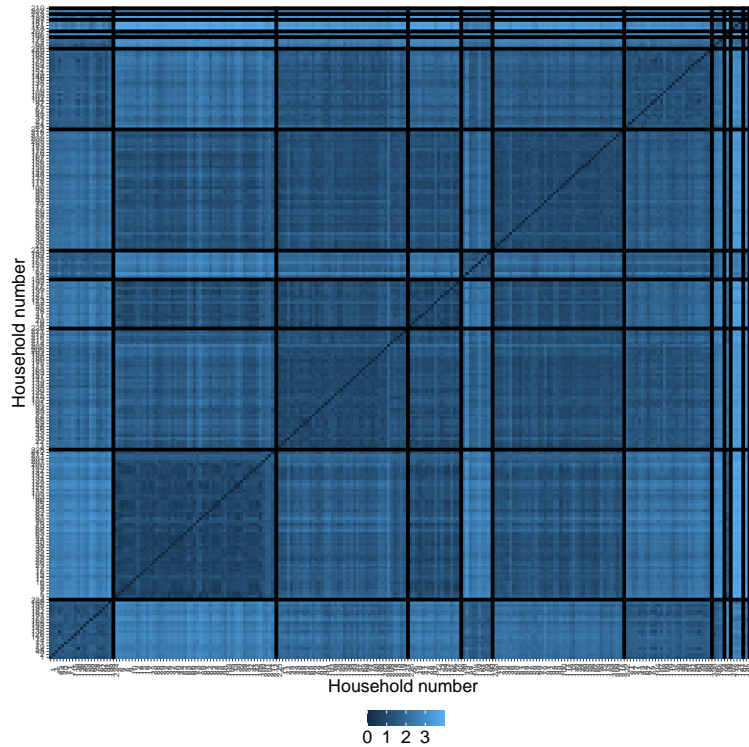


Fig 1: Heatmap of DTW distances (transformed to log-scale) between average daily household profiles, arranged according to clusters.

this cluster, making the number of spatial points insufficient to produce a reliable estimate of the variations in the quantile. Also, the pattern of the cyclical nature of the electricity demand for Cluster 12 is different from the other households, with the peak value occurring at the beginning of the day only. These correspond to Households 202, 210, and others.

## REFERENCES

- BELLONI, A. and OLIVEIRA, R. I. (2018). A high dimensional central limit theorem for martingales, with applications to context tree models. *arXiv preprint arXiv:1809.02741*.
- CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association* **91** 862–872.
- CHAUDHURI, P., DOKSUM, K. and SAMAROV, A. (1997). On average derivative quantile regression. *The Annals of Statistics* **25** 715 – 744. <https://doi.org/10.1214/aos/1031833670>
- CHEN, Z. (1996). Conditional Lp-quantiles and their application to the testing of symmetry in non-parametric regression. *Statistics & Probability Letters* **29** 107-115. [https://doi.org/10.1016/0167-7152\(95\)00163-8](https://doi.org/10.1016/0167-7152(95)00163-8)
- CHOWDHURY, J. and CHAUDHURI, P. (2019). Nonparametric depth and quantile regression for functional data. *Bernoulli* **25** 395 – 423. <https://doi.org/10.3150/17-BEJ991>
- DAS, P. and GHOSAL, S. (2017). Analyzing ozone concentration by Bayesian spatio-temporal quantile regression. *Environmetrics* **28** e2443.
- DAS, D. and LAHIRI, S. (2020). Central Limit Theorem in High Dimensions : The Optimal Bound on Dimension Growth Rate. <https://doi.org/10.48550/ARXIV.2008.04389>
- DE GOOIJER, J. G., GANNOUN, A. and ZEROM, D. (2006). A multivariate quantile predictor. *Communications in Statistics—Theory and Methods* **35** 133–147.
- DEB, S. and JANA, K. (2021). Nonparametric quantile regression for time series with replicated observations and its application to climate data. *arXiv preprint arXiv:2107.02091*.
- DUAN, Q., MCGRORY, C. A., BROWN, G., Mengersen, K. and WANG, Y.-G. (2021). Spatio-temporal quantile regression analysis revealing more nuanced patterns of climate change: a study of long-term daily temperature in Australia. *arXiv preprint arXiv:2103.05791*.

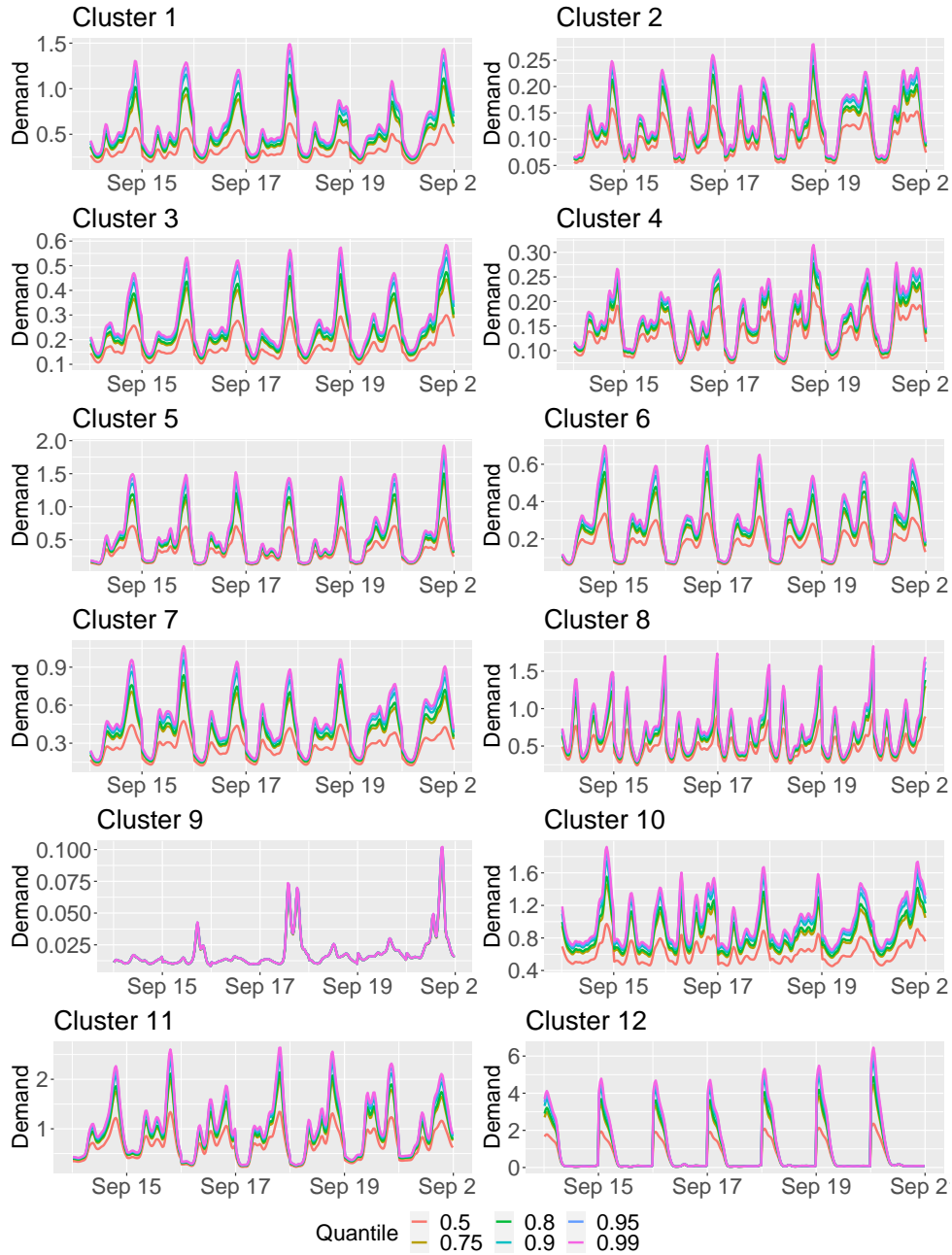


Fig 2: Predicted quantile curves for the demand during the last week of the dataset, for different households in the 12 clusters.

- EDDY, W. F. (1985). Ordering of multivariate data. *Computer Science and Statistics: The Interface*. L. Billard, Amsterdam: North-Holland 25–30.
- EINMAHL, J. H. and MASON, D. M. (1992). Generalized quantile processes. *The Annals of Statistics* 1062–1078.
- EZZAHRIQUI, M. and OULD-SAÏD, E. (2008). Asymptotic results of a nonparametric conditional quantile estimator for functional time series. *Communications in Statistics—Theory and Methods* **37** 2735–2759.
- EZZAHRIQUI, M. and OULD-SAÏD, E. (2008). Asymptotic normality of the kernel estimators of the conditional quantile in the normed space. *Far East J. Theoretical Statist* **25** 15–38.
- FERRATY, F., RABHI, A. and VIEU, P. (2005). Conditional quantiles for dependent functional data with application to the climatic "el niño" phenomenon. *Sankhyā: The Indian Journal of Statistics* 378–398.

- FIGALLI, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Analysis* **177** 413–421.
- GOLD, O. and SHARIR, M. (2018). Dynamic time warping and geometric edit distance: Breaking the quadratic barrier. *ACM Transactions on Algorithms (TALG)* **14** 1–17.
- HUBER, P. J. (2011). Robust statistics. In *International encyclopedia of statistical science* 1248–1251. Springer.
- JACOB, M., NEVES, C. and VUKADINOVIĆ GREETHAM, D. (2020). *Forecasting and assessing risk of individual electricity peaks*. Springer Nature.
- KOENKER, R. and BASSETT JR, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society* 33–50.
- KOENKER, R., NG, P. and PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680. <https://doi.org/10.1093/biomet/81.4.673>
- KONEN, D. and PAINDAVEINE, D. (2022). Multivariate  $\rho$ -quantiles: A spatial approach. *Bernoulli* **28** 1912 – 1934. <https://doi.org/10.3150/21-BEJ1404>
- KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511802256>
- POLLARD, D. (1991). Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory* **7** 186–199.
- PORTNOY, S. (1986). On the central limit theorem in  $R^p$  when  $p \rightarrow \infty$ . *Probability Theory and Related Fields* **73** 571–583. <https://doi.org/10.1007/BF00324853>
- REICH, B. J. (2012). Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **61** 535–553.
- REICH, B. J., FUENTES, M. and DUNSON, D. B. (2011). Bayesian Spatial Quantile Regression. *Journal of the American Statistical Association* **106** 6–20. PMID: 23459794. <https://doi.org/10.1198/jasa.2010.ap09237>
- RESNICK, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer.
- TAKEUCHI, I., LE, Q., SEARS, T., SMOLA, A. et al. (2006). Nonparametric quantile estimation.
- TAO, T. (2012). *Topics in random matrix theory* **132**. American Mathematical Soc.
- VERSHYNIN, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.
- ZHAO, Z. and WU, W. B. (2008). Confidence bands in nonparametric time series regression. *The Annals of Statistics* **36** 1854–1878.
- ZHOU, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *The Annals of Statistics* **38** 2187.
- ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics* **37** 2696 – 2729. <https://doi.org/10.1214/08-AOS636>

## APPENDIX A: PROOF OF PROPOSITION 1

### A.1. Proof of (P1).

$$\begin{aligned}
 -a_n \nabla \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}_0) &= \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} + \mathbf{u} \right) + \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \mathbf{V}_2(t_i) \mathbf{V}_1^{1/2}(t_i) \boldsymbol{\xi}(t_i) \\
 &\quad + \frac{a_n}{n} \sum_{i=1}^n \mathcal{O}_p(\|\mathbf{K}(t_i)\| \|\boldsymbol{\eta}(t_i)\|^{-3} \|\mathbf{V}_1^{1/2}(t_i)\| \|\boldsymbol{\xi}(t_i)\|^2)
 \end{aligned}$$

Let us start with the third term. Due to the sub-gaussian nature of  $\boldsymbol{\xi}(t)$  for any  $t \in \mathcal{T}$ ,  $\|\boldsymbol{\xi}(t)\| = \mathcal{O}_p(\sqrt{p})$ ,  $\|\boldsymbol{\xi}(t)\|^2 = \mathcal{O}_p(p + \sqrt{p})$  and  $\|\boldsymbol{\xi}(t)\|^3 = \mathcal{O}_p(p\sqrt{p} + p)$ , with probability at least  $(1 - e^{-Cp})$ . Also, since each element of the kernel matrix is  $\mathcal{O}(b_n)$ , its norm is bounded by  $\mathcal{O}(\sqrt{p}b_n)$  asymptotically almost surely (Tao, 2012). Therefore, it follows that

$$\frac{a_n}{n} \sum_{i=1}^n \mathcal{O}_p(\|\mathbf{K}(t_i)\| \|\boldsymbol{\eta}(t_i)\|^{-3} \|\mathbf{V}_1^{1/2}(t_i)\| \|\boldsymbol{\xi}(t_i)\|^2) = \mathcal{O} \left( \frac{a_n b_n^2 p^{3/4} (p + \sqrt{p})}{b_n^3 (p\sqrt{p} - p)^2} \right) = \mathcal{O} \left( b_n^{-2} p^{-9/4} \right).$$

Since  $p^{9/8} b_n \rightarrow \infty$  as  $n \rightarrow \infty$ , it follows that this third term goes to 0 asymptotically almost surely.

Turning to the first term, we note that  $\left\| \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} \right\| \leq 1$  and  $\|\mathbf{u}\| \leq 1$ . Consequently, we can rewrite it as

$$\begin{aligned}
& \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} + \mathbf{u} \right) \\
&= \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} - \mathbb{E}_{\mathbf{Y}(t)|\mathbf{X}(t)=\mathbf{x}} \left( \frac{\mathbf{Y}(t) - \mathbf{q}_0}{\|\mathbf{Y}(t) - \mathbf{q}_0\|} \right) \right) + \\
& \quad \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \mathbb{E}_{\mathbf{Y}(t)|\mathbf{X}(t)=\mathbf{x}} \left( \frac{\mathbf{Y}(t) - \mathbf{q}_0}{\|\mathbf{Y}(t) - \mathbf{q}_0\|} + \mathbf{u} \right) \\
&= \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} - \mathbb{E}_{\mathbf{Y}(t)|\mathbf{X}(t)=\mathbf{x}} \left( \frac{\mathbf{Y}(t) - \mathbf{q}_0}{\|\mathbf{Y}(t) - \mathbf{q}_0\|} \right) \right), \\
& \quad \text{as the second summand is equal to 0 by Lemma 3.1.} \\
&= \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} - \frac{\boldsymbol{\mu}(t) - \mathbf{q}_0}{\|\boldsymbol{\mu}(t) - \mathbf{q}_0\|} + O(p^{-1/2}) \right) \\
&= \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} - \frac{\boldsymbol{\mu}(t) - \mathbf{q}_0}{\|\boldsymbol{\mu}(t) - \mathbf{q}_0\|} \right) + O(b_n^{-1} p^{-3/2})
\end{aligned}$$

where we again apply Taylor series decomposition for  $\mathbb{E}_{\mathbf{Y}(t)|\mathbf{X}(t)=\mathbf{x}} \left( \frac{\mathbf{Y}(t) - \mathbf{q}_0}{\|\mathbf{Y}(t) - \mathbf{q}_0\|} \right)$  and bound the norms using standard sub-gaussian tail bounds. Clearly, as  $b_n^{-2} p^{-9/4} \rightarrow 0$ , it readily follows that  $b_n^{-1} p^{-3/2} \rightarrow 0$  as  $n \rightarrow \infty$ . In view of (15), the first term here converges to  $\boldsymbol{\eta}_t(\mathbf{q}_0)$  in the limit.

We also note that the variance of  $\sum_{i=1}^n \mathbf{K}(t_i) (\boldsymbol{\eta}(t_i)/\|\boldsymbol{\eta}(t_i)\| + \mathbf{u})$  is bounded through

$$\begin{aligned}
\left\| \text{Var} \left( \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} + \mathbf{u} \right) \right) \right\| &\leq \frac{a_n^2}{n^2} \sum_{i=1}^n \|\mathbf{K}(t_i)\|^2 \left\| \text{Var} \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} \right) \right\| \\
&= O \left( \frac{a_n^2}{n} (p + \sqrt{p}) b_n^2 \right) = O((np)^{-1})
\end{aligned}$$

where  $\|\text{Var}(\boldsymbol{\eta}(t_i)/\|\boldsymbol{\eta}(t_i)\|)\| = \mathcal{O}(1)$  as the norm of the random variable is uniformly bounded by 1. Clearly, as  $n \rightarrow \infty$ , the variance tends to 0 **which, given the assumptions in the proposition, implies**

$$\left\| \frac{a_n}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left( \frac{\boldsymbol{\eta}(t_i)}{\|\boldsymbol{\eta}(t_i)\|} + \mathbf{u} \right) - \boldsymbol{\eta}_t(\mathbf{q}_0) \right\| \xrightarrow{P} 0.$$

Finally, we consider the second term. Note that,  $\boldsymbol{\eta}(t_i)$ ,  $\mathbf{V}_1(t_i)$  and  $\mathbf{V}_2(t_i)$  are all predictable with respect to the filtration  $\{\mathcal{F}_t : t \in \mathcal{T}\}$ . Therefore,  $\mathbf{K}(t_i) \mathbf{V}_2(t_i) \mathbf{V}_1^{1/2}(t_i) \boldsymbol{\xi}(t_i)$  is a martingale difference with respect to the same filtration. Letting  $\zeta(t_i) := \frac{a_n}{n} \mathbf{K}(t_i) \mathbf{V}_2(t_i) \mathbf{V}_1^{1/2}(t_i) \boldsymbol{\xi}(t_i)$ , we obtain

$$\Sigma_{\zeta,i}^2 = \mathbb{E} [\zeta(t_i) \zeta(t_i)^\top | \mathcal{F}_{t_{i-1}}] = \frac{a_n^2}{n^2} \mathbf{K}(t_i) \mathbf{V}_2(t_i) \mathbf{V}_1(t_i) \mathbf{V}_2(t_i)^\top \mathbf{K}(t_i)$$

In view of (16),

$$\Delta_n := \left\| \sum_{i=1}^n \Sigma_{\zeta,i}^2 - \Omega_t(\mathbf{q}_0) \right\| \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Let,  $f \in \mathcal{F}$ , the set of three-times differentiable functions with bounded first, second and third derivative, and let  $c_2$  and  $c_3$  denote the upper bound of the second and third derivative of  $f$  respectively. Then, application of the quantitative central limit theorem for the multidimensional martingale (Belloni and Oliveira, 2018) ensures that

$$\left\| \mathbb{E} \left[ f \left( \frac{a_n}{n} \sum_{i=1}^n \zeta(t_i) \right) \right] - \mathbb{E} [f(\mathbf{Z}^*(\mathbf{q}_0))] \right\| \leq 2c_2 \Delta_n + c_3 \sum_{i=1}^n \mathbb{E} [\|\zeta(t_i)\|^3],$$

where  $\mathbf{Z}^*(\mathbf{q}_0)$  is a mean-zero Gaussian vector with dispersion matrix  $\Omega_t(\mathbf{q}_0)$ . Since  $\Delta_n \rightarrow 0$ , then

$$\sum_{i=1}^n \mathbb{E} [\|\zeta(t_i)\|^3] = \mathcal{O} \left( \frac{a_n^3}{n^3} \sum_{i=1}^n (p\sqrt{p} + p)^3 b_n^3 \right) = \mathcal{O}(n^{-2} p^{3/2}).$$

Additionally, it follows from  $p^2/n \rightarrow 0$  that  $n^{-2} p^{3/2} \rightarrow 0$ , as  $n \rightarrow \infty$ .

Combining all these using Slutsky's theorem, we have for any fixed  $\mathbf{u} \in B^{p-1}$  and for sufficiently large  $n$ ,

$$\sup_{f \in \mathcal{F}} \left\| \mathbb{E} [f(a_n(\hat{q}_n(\mathbf{u}, \mathbf{x}) - Q(\mathbf{u}, \mathbf{x})))] - \mathbb{E} [f(-\Omega_t^{-1/2}(\mathbf{q}_0) \mathbf{Z}(\mathbf{q}_0))] \right\| = \mathcal{O}(\max\{p^2/n, b_n^{-1} p^{-9/8}\})$$

where  $\mathbf{Z}(\mathbf{q}_0)$  is given as in the statement of proposition (P1). Finally, we can apply a standard  $\epsilon$ -net argument (see Section 2.3.1 of Tao (2012)) with union bounds over the unit ball  $B^{p-1}$  to take a supremum over the choice of  $\mathbf{u}$ , which transforms the above into uniform convergence as needed in the proposition (P1).

**A.2. Proof of (P2).** First, we express the term  $\nabla^2 \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$  by computing the second order derivative of the objective function  $\mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$  as follows

$$\nabla^2 \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) = \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left[ \frac{I_p}{\|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}(t_i)}} - \frac{\mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q})(\mathbf{Y}(t_i) - \mathbf{q})^\top \mathbf{K}(t_i)}{\|\mathbf{Y}(t_i) - \mathbf{q}\|_{\mathbf{K}(t_i)}^3} \right] \mathbf{K}(t_i).$$

We perform an M-R decomposition technique on  $\nabla^2 \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q})$  as in Zhao and Wu (2008) and rewrite

$$\nabla^2 \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) = M_{n,p} + R_{n,p}$$

where

$$\begin{aligned} \mathbf{a}(t_i) &= \mathbf{K}(t_i)(\mathbf{Y}(t_i) - \mathbf{q}), \\ R_{n,p} &= \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \mathbb{E} \left[ \frac{I_p}{\|\mathbf{a}(t_i)\|} - \frac{\mathbf{a}(t_i) \mathbf{a}(t_i)^\top}{\|\mathbf{a}(t_i)\|^3} \mid \mathcal{G}_{t_i-1} \right] \mathbf{K}(t_i), \\ M_{n,p} &= \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left[ \frac{I_p}{\|\mathbf{a}(t_i)\|} - \frac{\mathbf{a}(t_i) \mathbf{a}(t_i)^\top}{\|\mathbf{a}(t_i)\|^3} \right] \mathbf{K}(t_i) - R_{n,p}. \end{aligned}$$

We note that  $M_{n,p}$  is the partial average of a martingale difference sequence with respect to the filtration  $\{\mathcal{G}_t : t \in \mathcal{T}\}$ . Applying the quantitative central limit theorem for multidimensional

martingale (Belloni and Oliveira, 2018), it follows that  $M_{n,p} = \mathcal{O}(pb_n)$  which goes to 0 as  $n \rightarrow 0$ . The conditions for applying the quantitative central limit theorem are similar to the ones shown in the appendix A.1 and can be verified easily.

To deal with  $R_{n,p}$ , we expand  $\mathbf{a}(t_i) = \boldsymbol{\eta}(t_i) + V_1^{1/2}(t_i)\boldsymbol{\xi}(t_i)$  and apply Taylor series expansion of the normalized errors about the origin. Hence,

$$R_{n,p} = \frac{1}{n} \sum_{i=1}^n \mathbf{K}(t_i) \left[ \frac{I_p}{\|\boldsymbol{\eta}(t_i)\|} - \frac{\boldsymbol{\eta}(t_i)\boldsymbol{\eta}(t_i)^\top}{\|\boldsymbol{\eta}(t_i)\|^3} \right] \mathbf{K}(t_i) + \mathcal{O}(\sqrt{pb_n}),$$

where we use the fact that  $\mathbb{E}[\boldsymbol{\xi}(t_i) \mid \mathcal{G}_{t_i-1}] = 0$  by independence of the errors and the filtration  $\{\mathcal{G}_t : t \in \mathcal{T}\}$  and mean-zero property of the errors. Note that, the matrix

$$\|\boldsymbol{\eta}(t_i)\|^2 I_p - \boldsymbol{\eta}(t_i)\boldsymbol{\eta}(t_i)^\top$$

is positive definite for all  $n, p$  and  $\mathbf{q} \notin \{\mathbf{Y}(t_i) : t_i \in \mathcal{T}_n\}$ . Therefore,

$$\left\| \nabla^2 \mathcal{M}_{\mathbf{u},n}^{(p)}(\mathbf{q}) - \boldsymbol{\Psi}_t(\mathbf{q}_0) \right\| = \mathcal{O}(pb_n, \sqrt{pb_n}),$$

and the precise result in (P2) follows readily since  $pb_n \rightarrow 0$  as  $n \rightarrow 0$ .

## APPENDIX B: SIMULATION STUDY

In this section, we focus on the finite-sample performance of the proposed spatio-temporal quantile estimation methodology in two distinct simulation frameworks. The first takes place in a full-fledged spatio-temporal setting, where the locations are treated in terms of a multivariate time series framework, without explicitly incorporating information on the spatial coordinates. In contrast, the second case has been designed so as to draw on the inter-connectedness of spatial coordinates to emulate typical features in spatio-temporal data. Both settings were chosen upon careful consideration for illustrating the overall potential and efficacy of the proposed nonparametric quantile estimation as part of this paper. In this exercise, in order to evaluate efficiency, we use two popular metrics – the mean absolute error (MAE) and the mean absolute percentage error (MAPE).

**B.1. Simulation setup 1.** We use  $p$  to denote the number of spatial points, and let  $s_1, \dots, s_p$  denote the positions of these points. In this setting, we assume that the exact coordinates of the locations are unknown. Therefore, to simplify the computations, we consider a regular set of the form  $\{1, 2, \dots, p\}$  to denote the set of the spatial points. Throughout this experiment, the number of time points is fixed at  $T = 300$ . The covariate series  $\{\mathbf{X}(t)\}_{1 \leq t \leq T}$  is assumed to be bivariate and we simulate it following a vector-valued autoregressive (VAR) series of order 1. In particular,  $\mathbf{X}(t)$  is assumed to follow a VAR(1) process with mean  $\mathbf{0}$ , satisfying the equation

$$\mathbf{X}(t) = \begin{bmatrix} 0.2 & 0.1 \\ -0.3 & 0.4 \end{bmatrix} \mathbf{X}(t-1) + \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix},$$

where  $w_{1t}, w_{2t}$  are uncorrelated white noise processes. We shall use  $\mathbf{X}_1(t)$  and  $\mathbf{X}_2(t)$  to denote the two coordinates of the vector.

Next, the mean function associated with the response variable  $\mathbf{Y}(t)$  (see (2) for reference) is assumed to take the form

$$\boldsymbol{\mu}(\mathbf{X}(t)) = \begin{bmatrix} \alpha_1 + \beta_{1,1}t + \beta_{2,1} \sin(2\pi \mathbf{X}_1(t)) + \beta_{3,1} \cos(2\pi \mathbf{X}_2(t)) \\ \alpha_2 + \beta_{1,2}t + \beta_{2,2} \sin(2\pi \mathbf{X}_1(t)) + \beta_{3,2} \cos(2\pi \mathbf{X}_2(t)) \\ \vdots \\ \alpha_p + \beta_{1,p}t + \beta_{2,p} \sin(2\pi \mathbf{X}_1(t)) + \beta_{3,p} \cos(2\pi \mathbf{X}_2(t)) \end{bmatrix}.$$



Note that the above is a nonlinear function in  $\mathbf{X}(t)$ , and this formulation is useful to understand if the specified nonparametric method is effective in a general scenario. To generate the intercept terms  $(\alpha_i)$  in the above expression, for  $i = 1, 2, \dots, p$ , we use the standard normal distribution. The  $\beta_{1,i}$  coefficients indicate a linear trend pattern, and they are generated independently from uniform distribution between  $-1$  and  $1$ . Finally, the parameters  $\beta_{2,i}$  and  $\beta_{3,i}$ , for all  $i$ , are obtained independently from a uniform distribution in the interval  $(-20, 20)$ . Note that these choices ensure a greater degree of dependence on  $\mathbf{X}(t)$ , and are helpful to assess the accuracy of the proposed technique.

For the covariance structure in the model (2), we consider

$$\Sigma(\mathbf{X}(t)) = 0.1 \|\mu(\mathbf{X}(t))\| [\exp\{-0.1(s - s')^2\}]_{s,s'=1}^p,$$

the second term indicating a matrix with  $(s, s')^{th}$  element given by the exponentially decaying covariance function. We point out that this specification maintains the signal to the noise ratio close to 0.1 throughout the process.

Based on the above simulation setup, we generate  $B = 1000$  datasets, each with 300 time-points and 10 spatial coordinates. Then, using Monte Carlo simulation, we can approximate the expectation in (4) for any choice of  $\mathbf{x}$  and  $\mathbf{u}$ , and, therefore, are able to optimize the objective function given in (4) to obtain  $Q(\mathbf{u}, \mathbf{x})$ . To assess the accuracy of the proposed technique, we consider this value as the true quantile which we wish to estimate. Now, we randomly choose 100 datasets among those 1000, and for each of them, we compute the proposed estimator  $\hat{q}(\mathbf{u}, \mathbf{x})$  as in (6). The errors are computed according to different sizes of the dataset, in terms of the number of time-points. Corresponding RMSE and MAPE values are summarized in Table 1. Here, we choose a Gaussian kernel over the temporal and spatial domain, and restrict our attention to the choice of  $\mathbf{u}$  as  $(2\alpha - 1)\mathbf{1}_p/\sqrt{p}$ , where  $0 < \alpha < 1$  is the quantile of interest.

TABLE 1  
MAE (MAPE given in parentheses) in estimating the  $\mathbf{u}$ -quantile at different time-points ( $t$ ), where  $\mathbf{u} = (2\alpha - 1)\mathbf{1}_p/\sqrt{p}$ , under simulation setup 1.

$\alpha$	$t = 10$	$t = 50$	$t = 150$	$t = 250$	$t = 290$
0.50	3.93 (10.75)	0.69 (3.72)	1.23 (4.04)	2.46 (3.42)	3.65 (1.24)
0.75	5.16 (15.97)	2.12 (4.29)	1.39 (5.17)	3.72 (2.71)	4.21 (1.87)
0.80	5.26 (20.44)	2.24 (4.58)	1.46 (5.20)	4.07 (2.61)	4.36 (2.07)
0.85	5.50 (22.96)	2.50 (5.12)	1.63 (5.35)	4.09 (2.52)	5.20 (2.45)
0.90	7.83 (25.44)	3.07 (9.03)	2.91 (6.89)	4.78 (2.97)	6.56 (3.37)
0.95	14.22 (27.87)	3.23 (12.89)	4.32 (10.70)	6.50 (4.09)	9.91 (4.86)

From the above table, we can observe that the percentage errors reduce substantially for larger sizes of the dataset. In estimating the median ( $\alpha = 0.5$ ), our proposed method makes only about 1.2% error if 290 observations per time series are allowed. The MAPE typically remains less than 3% for moderate quantiles, and for more than 250 observations per spatial point. The errors are slightly higher for more extreme quantiles. For instance, at  $\alpha = 0.95$  and for  $t \geq 250$ , our technique incurs 4 to 5 percent error on average.

In order to demonstrate the predictive performance of the method, we use the observations for  $t \leq 290$  as the training set, and then forecast the quantiles for the next time-points. The RMSE and MAPE thus computed, across 100 different datasets, are reported in Table 2. We can see that the errors vary between 3% to 9.2%, and stays around 4 to 5 percent in most of the cases. The method is robust in predicting 1 to 5 steps ahead. For different quantiles though, we see slightly different behaviors. For central quantiles, typically the errors are more, whereas for the extreme quantiles, our approach registers better accuracy with the MAPE of around 3 to 4.5 percent.

TABLE 2  
MAE (MAPE given in parentheses) in predicting the  $\mathbf{u}$ -quantile, where  $\mathbf{u} = (2\alpha - 1)\mathbf{1}_p/\sqrt{p}$ , for simulation setup 1.

$\alpha$	1-step ahead	2-step ahead	3-step ahead	4-step ahead	5-step ahead
0.50	11.653 (6.361)	11.876 (6.496)	9.305 (4.997)	16.723 (9.18)	9.57 (5.16)
0.75	10.686 (5.776)	10.496 (5.666)	7.003 (3.693)	14.093 (7.704)	7.489 (3.999)
0.80	10.129 (5.478)	9.882 (5.318)	6.473 (3.4)	13.413 (7.322)	7.012 (3.736)
0.85	10.689 (5.839)	9.342 (5.048)	5.879 (3.076)	12.588 (6.857)	6.481 (3.449)
0.90	9.847 (5.362)	9.722 (5.287)	6.009 (3.184)	11.527 (6.261)	5.835 (3.108)
0.95	8.735 (4.666)	8.166 (4.387)	5.493 (2.876)	11.629 (6.331)	5.855 (3.095)

**B.2. Simulation setup 2.** In the second simulation scenario, unlike the previous multivariate setup where only partial spatial information is assumed to be known, we focus on a spatio-temporal framework to generate the data where the actual spatial information is used in an explicit manner. We follow the work of [Das and Ghosal \(2017\)](#), where the authors considered the spatio-temporal estimation problem from a Bayesian paradigm.

To develop this model,  $p = 15$  spatial locations  $\mathbf{s}_1, \dots, \mathbf{s}_{15}$  are first generated uniformly in the unit square  $[0, 1]$ . Thus,  $\mathcal{S}_p = \{\mathbf{s}_1, \dots, \mathbf{s}_{15}\}$ , and  $\mathbf{s}_i = (s_{i1}, s_{i2})$  with  $0 \leq s_{i1}, s_{i2} \leq 1$ . Next, we take 200 equidistant points in the interval  $[0, 1]$  as the time-points. In other words,  $\mathcal{T}_n = \{t_1, t_2, \dots, t_n\}$  with  $t_1 = 0, t_n = 1$  and  $(t_i - t_{i-1}) = 1/(n - 1)$  for all  $i = 2, \dots, n$ .

We now assume that  $\mathbf{X}(t) = t$  for all  $t \in \mathcal{T} = [0, 1]$ . This can be construed as a linear trend function. The conditional quantile-curve of level  $\tau \in [0, 1]$ , conditioned on the temporal variable  $t$  and spatial variable  $\mathbf{s}$ , takes the form

$$(28) \quad Q(\tau | t, \mathbf{s}) = t\xi_1(\tau, \mathbf{s}) + (1 - t)\xi_2(\tau, \mathbf{s}),$$

so that

$$\xi_1(\tau, \mathbf{s}_i) = \left(1 - \frac{s_{i1} + s_{i2}}{2}\right) \tau^2 + s_{i1} \frac{\log(1 + \tau)}{2 \log 2} + \frac{s_{i2}}{2} \tau^3$$

and

$$\xi_2(\tau, \mathbf{s}_i) = (1 - s_{i2}^2) \sin(\tau\pi/2) + s_{i2} \frac{(e^\tau - 1)}{(e - 1)}.$$

Owing to the inverse relationship between the quantile and the distribution function, we can generate observations from the above model by simulating uniform  $[0, 1]$  random variables and evaluating the quantile function at those generated values. More relevant details of this approach can be found in [Das and Ghosal \(2017\)](#).

TABLE 3  
MAE (MAPE given in parentheses) in estimating the  $\tau$ -quantile at different time-points ( $t$ ) under simulation setup 2.

$\tau$	$t = 0$	$t = 0.25$	$t = 0.5$	$t = 0.75$	$t = 1$
0.50	0.129 (1.44)	0.037 (0.47)	0.019 (0.27)	0.119 (2.07)	0.388 (8.28)
0.75	0.135 (1.51)	0.033 (0.42)	0.045 (0.66)	0.211 (3.67)	0.547 (11.66)
0.8	0.155 (1.73)	0.035 (0.44)	0.056 (0.82)	0.238 (4.14)	0.587 (12.51)
0.85	0.171 (1.92)	0.037 (0.47)	0.068 (1)	0.263 (4.58)	0.629 (13.4)
0.9	0.185 (2.08)	0.037 (0.48)	0.08 (1.18)	0.291 (5.06)	0.67 (14.29)
0.95	0.205 (2.29)	0.042 (0.54)	0.096 (1.42)	0.323 (5.62)	0.721 (15.37)
0.99	0.314 (3.52)	0.082 (1.05)	0.111 (1.63)	0.348 (6.06)	0.758 (16.16)

TABLE 4

MAE (MAPE given in parentheses) in estimating the  $\tau$ -quantile at different time-points ( $t$ ) under simulation setup 2 with 90 time-points and 10 locations in training set.

$\tau$	$t = 0.91$	$t = 0.92$	$t = 0.93$	$t = 0.94$	$t = 0.95$
0.5	0.343 (6.64)	0.355 (6.93)	0.367 (7.23)	0.38 (7.54)	0.393 (7.86)
0.75	0.158 (1.7)	0.152 (1.64)	0.146 (1.58)	0.14 (1.52)	0.134 (1.46)
0.8	0.352 (3.41)	0.343 (3.34)	0.335 (3.26)	0.326 (3.19)	0.318 (3.11)
0.85	0.695 (6.1)	0.685 (6.03)	0.675 (5.95)	0.664 (5.87)	0.654 (5.8)
0.9	1.214 (9.69)	1.204 (9.63)	1.194 (9.56)	1.184 (9.5)	1.174 (9.43)
0.95	1.922 (14.01)	1.916 (13.97)	1.909 (13.93)	1.902 (13.89)	1.896 (13.85)
0.99	2.601 (17.65)	2.599 (17.64)	2.597 (17.63)	2.596 (17.62)	2.594 (17.61)

$\tau$	$t = 0.96$	$t = 0.97$	$t = 0.98$	$t = 0.99$	$t = 1$
0.5	0.406 (8.19)	0.419 (8.53)	0.433 (8.89)	0.447 (9.25)	0.461 (9.63)
0.75	0.128 (1.4)	0.123 (1.35)	0.117 (1.29)	0.112 (1.24)	0.107 (1.19)
0.8	0.309 (3.04)	0.301 (2.97)	0.293 (2.9)	0.285 (2.83)	0.277 (2.76)
0.85	0.644 (5.72)	0.634 (5.64)	0.625 (5.57)	0.615 (5.49)	0.605 (5.42)
0.9	1.164 (9.36)	1.155 (9.3)	1.145 (9.23)	1.135 (9.17)	1.126 (9.11)
0.95	1.889 (13.81)	1.882 (13.77)	1.876 (13.73)	1.869 (13.7)	1.863 (13.66)
0.99	2.593 (17.6)	2.591 (17.59)	2.589 (17.58)	2.588 (17.58)	2.586 (17.57)

We have generated 100 datasets aimed at illustrating the finite sample behaviour of the quantile function for the above case. Then, the accuracy of the model is computed by checking the errors of these estimates with respect to the true quantile given by (28). In table 3, we demonstrate the mean absolute errors (MAE) between the true quantile values and the fitted quantiles. The mean absolute percentage error (MAPE) is shown in the brackets. Clearly, the maximum error is merely 16% in case of extreme quantiles, for the quantiles closer to the median the error is less than 5%. Also, as expected, the errors tend to increase as we move towards the extreme quantile or the right hand of the temporal domain due to the boundary effects.

We repeat the simulation exercise to measure the predictive performance of our proposed estimator as well. In this case, we use the first 90 temporal points and 12 locations as the training data, and use that to predict the behaviour of the quantiles for the rest of the 3 locations for time-points  $t = 0.91$  to  $t = 1$ . The summary measures MAE and MAPE are presented in table 4. It highlights that the error increases slowly for multiple step ahead predictions, the percentage error in prediction is 17.5% for extreme quantiles.

The first simulation setup imitates the data generating mechanism presented in this paper. Thus, the results present in simulation setup 1 corroborate our theoretical findings. On the other hand, simulation setup 2 describes the data generation mechanism through the specification of the quantile function. Therefore, the generated data may not satisfy the model (2). Since, this particular simulation illustrates actual spatio-temporal data, it is important to verify if the proposed estimator works equally well under this setup. The aforementioned simulation illustrates this point. Due to the violation of these model assumptions, the errors in simulation setup 2 are more than the errors for setup 1, though the percentage error remains at an acceptable level for non-extreme quantiles.