T E C H N I C A L
R E P O R T

0552

LOCAL LINEAR SPATIAL QUANTILE REGRESSION

HALLIN M., LU Z., and K. YU

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Local Linear Spatial Quantile Regression *

Marc Hallin,
ISRO, ECARES, and Département de Mathématique
Université Libre de Bruxelles, Belgium

Zudi Lu
Institute of Systems Science
Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing, China,
and
Department of Statistics
London School of Economics, U.K.

and

Keming Yu
Department of Mathematics and Statistics
University of Plymouth, UK

August 11, 2005

Abstract

Let $\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}), \mathbf{i} \in \mathbb{Z}^N\}$ be a stationary real-valued $(d+1)$-dimensional spatial processes. Denote by $\mathbf{x} \mapsto q_p(\mathbf{x})$, $p \in (0,1)$, $\mathbf{x} \in \mathbb{R}^d$, the spatial quantile regression function of order p, characterized by $\mathrm{P}\{Y_{\mathbf{i}} \leq q_p(\mathbf{x}) | \mathbf{X}_{\mathbf{i}} = \mathbf{x}\} = p$. Assume that the process has been observed over a N-dimensional rectangular domain of the form $\mathcal{I}_{\mathbf{n}} := \{\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{Z}^N | 1 \leq i_k \leq n_k, k = 1, \ldots, N\}$, $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{Z}^N$. We propose a local linear kernel estimator of q_p, which extends to random fields with unspecified and possibly highly complex spatial dependence structure the weighted quantile regression methods considered in the context of independent samples or time series. Under mild regularity assumptions, we obtain a Bahadur representation for the estimators of q_p and its derivatives, from which we establish consistency and asymptotic normality. The spatial process is assumed to satisfy some very general mixing conditions, generalizing classical time-series strong mixing concepts. The size of the rectangular domain $\mathcal{I}_{\mathbf{n}}$ is allowed to tend to infinity at different rates depending on the direction in $\mathbb{Z}^N$ (non-isotropic asymptotics). The method provides much richer information than the traditional mean regression approach.

AMS 1980 subject classification : Primary: 62G05; Secondary: 60J25, 62J02.
Key words and phrases : mixing random field, bandwidth choice, local linear kernel estimate, spatial quantile regression, Bahadur representation, asymptotic normality.

*Research supported by a PAI contract of the Belgian Federal Government, an Action de Recherche Concertée of the Communauté Française de Belgique, the National Natural Science Foundation of China, and the Leverhulme Trust.

1 Introduction

Since the pathbreaking paper by Koenker and Basset (1978), quantile regression methods have attracted considerable interest, basically in all domains of statistics: see Koenker (2000) or the recent monograph by Koenker (2005) for a review of regression and autoregression quantiles in a traditional setting of independent samples or time series data. To the best of our knowledge, and quite surprisingly so, quantile regression seldom has been considered in a spatial context. Recently, Koenker and Mizera (2004) have proposed, under the name of *penalized triograms*, a penalized spline method based on adaptively selected triangulations of the plane which allows for computing conditional quantiles. Their method however is limited to the case where regressors are the two-dimensional spatial coordinates, and does not take into account the spatial dependence structure of the data.

Let $\mathbb{Z}^N$, $N \geq 1$, denote the integer lattice points in the N-dimensional Euclidean space. A point $\mathbf{i} = (i_1, \ldots, i_N)$ in $\mathbb{Z}^N$ will be referred to as a *site*. Spatial data are modelled as finite realizations of vector stochastic processes indexed by $\mathbf{i} \in \mathbb{Z}^N$, also called *random fields*. In this paper, we will consider strictly stationary $(d+1)$-dimensional real random fields, of the form

$$\Big\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}})\ ;\ \mathbf{i} \in \mathbb{Z}^N\Big\}, \tag{1.1}$$

where $Y_{\mathbf{i}}$, with values in $\mathbb{R}$, and $\mathbf{X}_{\mathbf{i}}$, with values in $\mathbb{R}^d$, are defined over some probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Such spatial data arise in a variety of fields, including econometrics, environmental sciences, image analysis, oceanography, geostatistics, and many others. The statistical treatment of such data is the subject of an abundant literature, which cannot be reviewed here; for background reading, we refer the reader to the monographs by Anselin and Florax (1995), Cressie (1993), Guyon (1995), Possolo (1991), or Ripley (1981).

In a number of applications, a crucial problem consists in describing and analyzing the influence of a vector $\mathbf{X}_{\mathbf{i}}$ of covariates on some real-valued response variable $Y_{\mathbf{i}}$. In the present context, where the observations are made over a collection of sites, this study is more difficult, due to the complexity of the possible spatial dependence among the various sites—a dependence that typically cannot be modelled in any adequate way, and is to be treated as an unspecified nuisance. The traditional approach to this problem consists in assuming that $Y_{\mathbf{i}}$ has finite expectation, so that the *spatial mean regression function* $g\,:\,\mathbf{x} \mapsto g(\mathbf{x}) := \mathrm{E}\Big[Y_{\mathbf{i}} | \mathbf{X}_{\mathbf{i}} = \mathbf{x}\Big]$ is well defined and clearly carries relevant information on the dependence of Y on the covariates $\mathbf{X}$. This approach has been successfully considered in several papers, among which Hallin, Lu, and Tran (2004b). However, (conditional) expectations may not exist. And, even when they do, they only carry a limited information on the dependence under study. In most practical cases, for instance, we would expect different structural relationships for the higher (or lower) order quantiles than for the central ones: a regression analysis based on conditional means and spatial mean regression functions (or on conditional medians as well) overlooks such an essential feature of the dependence of Y on $\mathbf{X}$, which can be taken care of by Koenker and Bassett's more general conditional quantile analysis only.

In this paper, instead of spatial mean regression, we thus consider the *spatial quantile regression functions* $q_p\,:\,\mathbf{x} \mapsto q_p(\mathbf{x})$, $0 < p < 1$, characterized by $\mathrm{P}\{Y_{\mathbf{i}} \leq q_p(\mathbf{x}) | \mathbf{X}_{\mathbf{i}} = \mathbf{x}\} = p$. Although q_p (just as g) is only defined up to a P-null set of values of $\mathbf{x}$ (being a class of P-a.s. mutually equal functions rather than a function), we treat it, for the sake of simplicity, as a well-defined real-valued $\mathbf{x}$-measurable function, which has no implication on the probabilistic statements of this paper. In the particular case under which $\mathbf{X}_{\mathbf{i}}$ itself is measurable with respect to a subset of $Y_{\mathbf{j}}$'s, with $\mathbf{j}$ ranging over some neighborhood of $\mathbf{i}$, q_p is called a *spatial quantile* auto*regression function.* Parametric (linear) spatial *mean* autoregression models were considered as early as 1954 by Whittle (1954, 1963); see Besag (1974) for further developments in this context. Similarly, we could expect that spatial quantile autoregression would be of wide interests in robust modelling of spatial dependence (cf., for instance, Sections 3.3 and 3.5 of Cressie 1993) as well as in the construction of confidence (prediction) intervals.

Our objective consists in estimating the spatial quantile regression functions $q_p : \mathbf{x} \mapsto q_p(\mathbf{x})$; contrary to Whittle (1954), we adopt a nonparametric point of view, as in Hallin, Lu and Tran (2004b), avoiding any parametric specification, both for q_p as for the possibly extremely complex spatial dependence structure of the data.

For $N = 1$, this problem reduces to the classical problem of quantile (auto)regression for independent samples or serially dependent observations. This problem has received extensive attention in the literature: see, for instance, Koenker and Bassett (1978, 1982), Koenker and Portnoy (1987), Granger, White, and Kamstra (1989), Efron (1990), Portnoy (1991), Fan, Hu, and Truong (1994), Koenker and Zhao (1996), Koul and Mukherjee (1994), Welsh (1996), Yu and Jones (1997, 1998), Taylor and Bunn (1999), Honda (2000), Cai (2002), as well as Yu and Lu (2004), to quote only a few. Quite surprisingly, despite its importance for applications, the spatial version ($N > 1$) of the same problem remains essentially unexplored. Several recent papers (among which Tran 1990, Tran and Yakowitz 1993, Carbon, Hallin, and Tran 1996, Hallin, Lu, and Tran 2001 and 2004a, Yao 2003) are dealing with the related problem of estimating the density f of a random field of the form $\{\mathbf{X}_{\mathbf{i}} \ ; \ \mathbf{i} \in \mathbb{Z}^N\}$, whereas Hallin, Lu, and Tran (2004b), Lu and Chen (2002, 2004)consider the estimation of spatial mean regression functions. But, to the best of our knowledge, the only attempt that has been made to estimate spatial quantile regression functions is Koenker and Mizera (2004)'s penalized triogram method, which however restricts to $d = 2 = N$, with $\mathbf{X}_{\mathbf{i}} = \mathbf{i}$ and mutually independent observations—on the other hand, a regular grid is not assumed for the sites.

Our estimators of the spatial quantile regression functions naturally involve some spatial smoothing techniques. Among all these techniques, the Nadaraya-Watson method, in the traditional serial case ($N = 1$), is probably the most standard one; it has been well documented, however—see, for instance, Fan and Gijbels (1996)—that this approach suffers from several severe drawbacks, such as poor boundary performances, excessive bias and low efficiency, and that the local polynomial fitting methods developed by Stone (1977) and Cleveland (1979) are generally preferable. Such local polynomial methods, and more particularly local linear fitting have become increasingly popular in the light of recent work by Fan (1992), Fan and Gijbels (1996), Ruppert and Wand (1994), Yu and Jones (1997, 1998), Loader (1999), and several others. For $N = 1$, Honda (2000) has studied the asymptotics of local polynomial fitting for quantile regression under general mixing conditions. In this paper, we extend this approach to the context of spatial quantile regression ($N > 1$) by defining an estimator of q_p based on local linear regression quantiles.

Extending classical or time-series asymptotics ($N = 1$) to spatial asymptotics ($N > 1$) however is far from trivial. Due to the absence of any canonical ordering in the space, there is no obvious definition of tail sigma-fields, ergodicity, mixing, and other traditional time-domain concepts. And, little seems to exist about this in the literature, where only central limit results are well documented: see, e.g., Bolthausen (1982) or Nakhapetian (1980). Even the simple idea of a sample size $\mathbf{n}$ going to infinity (the sample size here is a domain in $\mathbb{Z}^N$) has to be clarified in this setting. The assumptions we are making in (A3, A3$'$, and A3$''$) are reasonable and flexible generalizations of traditional time series concepts.

The paper is organized as follows. In Section 2.1 we provide the notation and main assumptions. Sections 2.2 and 2.3 introduce the main ideas underlying local linear regression in the context of random fields, and sketch the main steps of the proofs to be developed in the sequel. Section 2.4, where asymptotic normality is stated under various types of asymptotics and various mixing assumptions, is the main theoretical section of the paper. Section 3 is devoted to a real-data application. Proofs and technical lemmas are concentrated in an Appendix (Section 4).

2 Local linear spatial quantile regression.

2.1 Notation and main assumptions.

For the sake of convenience, we are summarizing here the main assumptions we are making on the random field (1.1) and the kernel K to be used in the estimation method. Assumptions (A1)-(A3) are related to the random field itself.

(A1) (Conditions on densities) The random field (1.1) is strictly stationary, with densities (denote by f the joint density of $(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}})$, by $f_{\mathbf{X}}$ the marginal density of $\mathbf{X}$, by $f_{Y|\mathbf{X}=\mathbf{x}}$ the density of Y conditional on $\mathbf{X} = \mathbf{x}$) satisfying the following assumptions:

- (i) for all $\mathbf{x}$, $\mathbf{x} \mapsto f_{\mathbf{X}}(\mathbf{x})$ is strictly positive and continuous;
- (ii) for all $\mathbf{x}$, there exist a neighborhood B of $y = q_p(\mathbf{x})$ and a neighborhood $\mathbf{B}$ of $\mathbf{x}$ such that $y \mapsto f_{Y|\mathbf{X}=\mathbf{x}}(y)$ is strictly positive and continuous over B, uniformly over $\mathbf{B}$, and $\mathbf{x} \mapsto f_{Y|\mathbf{X}=\mathbf{x}}(y)$ is continuous over $\mathbf{B}$ for all $y \in B$;
- (iii) the joint density $f_{\mathbf{i},\mathbf{j}}(\mathbf{x}, \tilde{\mathbf{x}})$ of $(\mathbf{X}_{\mathbf{i}}, \mathbf{X}_{\mathbf{j}})$ is bounded uniformly in $\mathbf{i}$ and $\mathbf{j}$, that is, $\sup_{\mathbf{i},\mathbf{j}\in\mathbb{Z}^N} \sup_{\mathbf{x},\,\tilde{\mathbf{x}}\in\mathbb{R}^d} f_{\mathbf{i},\mathbf{j}}(\mathbf{x}, \tilde{\mathbf{x}}) \le C$ for some $C > 0$.

(A2) (Conditions on the spatial quantile regression function) The spatial quantile regression function $\mathbf{x} \mapsto q_p(\mathbf{x})$ is twice differentiable. Denoting by $\dot{q}_p(\mathbf{x})$ and $\ddot{q}_p(\mathbf{x})$ its gradient and the matrix of its second derivatives (at $\mathbf{x}$), respectively, $\mathbf{x} \mapsto \ddot{q}_p(\mathbf{x})$ is continuous at all $\mathbf{x}$.

Conditions similar to Assumption (A1) have been considered in the literature, in the *i.i.d.* setting (cf. Fan *et al.*, 1994). Assumption (A2) is standard. Throughout, we denote by C a generic positive constant, the value of which may vary according to the context.

Besides (A1) and (A2), we need some appropriate assumption of spatial mixing. For any collection $\mathcal{S} \subset \mathbb{Z}^N$ of sites, denote by $\mathcal{B}(\mathcal{S})$ the Borel σ-field generated by $\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}})|\ \mathbf{i} \in \mathcal{S}\}$. For each couple $\mathcal{S}', \mathcal{S}''$, let $d(\mathcal{S}', \mathcal{S}'') := \min\{\|\mathbf{i}' - \mathbf{i}''\| \,|\, \mathbf{i}' \in \mathcal{S}',\ ,\mathbf{i}'' \in \mathcal{S}''\}$ be the distance between $\mathcal{S}'$ and $\mathcal{S}''$, where $\|\mathbf{i}\| := (i_1^2 + \ldots + i_N^2)^{1/2}$ stands for the Euclidean norm. Finally, write $\mathrm{Card}(\mathcal{S})$ for the cardinality of $\mathcal{S}$. As in Hallin, Lu, and Tran (2004b), two distinct forms (either (A3) and (A3$'$) or (A3) and (A3$''$)) of spatial mixing are considered.

(A3) There exist two functions, $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\varphi(t) \downarrow 0$ as $t \to \infty$, and $\psi : \mathbb{N}^2 \to \mathbb{R}^+$ symmetric and decreasing in each of its two arguments, such that

$$\begin{aligned} \alpha\left(\mathcal{B}(\mathcal{S}'), \mathcal{B}(\mathcal{S}'')\right) &:= \sup\left\{|\mathrm{P}(AB) - \mathrm{P}(A)\mathrm{P}(B)|, A \in \mathcal{B}(\mathcal{S}'),\ B \in \mathcal{B}(\mathcal{S}'')\right\} \\ &\le \psi(\mathrm{Card}(\mathcal{S}'), \mathrm{Card}(\mathcal{S}''))\varphi(d(\mathcal{S}', \mathcal{S}'')). \end{aligned} \tag{2.1}$$

for any $\mathcal{S}', \mathcal{S}'' \subset \mathbb{Z}^N$. The function φ moreover is such that

$$\lim_{m\to\infty} m^a \sum_{j=m}^{\infty} j^{N-1}\{\varphi(j)\} = 0 \quad \text{for some constant } a > N.$$

The assumptions we are making on the function ψ are either

(A3$'$) $\psi(n', n'') \leq \min(n', n'')$

or

(A3$''$) $\psi(n', n'') \leq C(n' + n'' + 1)^\kappa$ for some $C > 0$ and $\kappa > 1$.

In case (2.1) holds with $\psi \equiv 1$, the random field $\{(Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}})\}$ is called *strongly mixing*. In the serial case ($N = 1$), many stochastic processes and time series are known to be strongly mixing; cf. Fan and Yao (2003). Guyon (1987) has shown that, under certain conditions, linear random fields of the form $\mathbf{X}_{\mathbf{n}} = \sum_{\mathbf{j}\in\mathbb{Z}^N} \mathbf{g}_{\mathbf{j}}\mathbf{Z}_{\mathbf{n}-\mathbf{j}}$, where the $\mathbf{Z}_{\mathbf{j}}$'s are independent random variables, are strongly mixing. Assumptions (A3$'$) and (A3$''$) are the same as the mixing conditions used by Neaderhouser (1980) and Takahata (1983), respectively, and are weaker than the uniform strong mixing condition considered by Nakhapetyan (1980). They are satisfied by many spatial models, as shown by Neaderhouser (1980), Rosenblatt (1985), and Guyon (1987).

Throughout, we assume that the random field (1.1) is observed over a rectangular region of the form $\mathcal{I}_{\mathbf{n}} := \{\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{Z}^N | \, 1 \leq i_k \leq n_k, \; k = 1, \ldots, N\}$, for $\mathbf{n} = (n_1, \cdots, n_N) \in \mathbb{Z}^N$ with strictly positive coordinates $n_1, \ldots, n_N$. The total sample size is thus $\hat{\mathbf{n}} := \prod_{k=1}^N n_k$. We write $\mathbf{n} \to \infty$ as soon as $\min_{1\leq k\leq N}\{n_k\} \to \infty$. A more demanding way for $\mathbf{n}$ to tend to infinity is the one considered in Tran (1990): we use the notation $\mathbf{n} \implies \infty$ if $\mathbf{n} \to \infty$ and moreover $|n_j/n_k| < C$ for some $0 < C < \infty$, $1 \leq j, \, k \leq N$. In this latter case, all components of $\mathbf{n}$ tend to infinity at the same rate.

Assumption (A4) deals with the kernel function $K : \mathbb{R}^d \to \mathbb{R}$ to be used in the estimation method. For any $\mathbf{c} := (c_0, \mathbf{c}_1')' \in \mathbb{R}^{d+1}$, define $K_{\mathbf{c}}(\mathbf{u}) := (c_0 + \mathbf{c}_1'\mathbf{u})K(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^d$.

(A4)(i) For any $\mathbf{c} \in \mathbb{R}^{d+1}$, $|K_{\mathbf{c}}(\mathbf{u})|$ is uniformly bounded by some constant $K_{\mathbf{c}}^+$, and is integrable, i.e., $\displaystyle\int_{\mathbb{R}^{d+1}} |K_{\mathbf{c}}(\mathbf{x})| d\mathbf{x} < \infty$;

(ii) for any $\mathbf{c} \in \mathbb{R}^{d+1}$, $|K_{\mathbf{c}}|$ has an integrable second order radial majorant, that is, $Q_{\mathbf{c}}^K(\mathbf{x}) := \sup_{\|\mathbf{y}\|\geq\|\mathbf{x}\|} \left[\|\mathbf{y}\|^2 K_{\mathbf{c}}(\mathbf{y})\right]$ is integrable;

(iii) the kernel function K is a bounded density function with compact support $C_K \subset \mathbb{R}^d$ such that $\int \mathbf{u}K(\mathbf{u})d\mathbf{u} = \mathbf{0}$ and $\int \mathbf{u}\mathbf{u}'K(\mathbf{u})d\mathbf{u}$ is positive definite.

Finally, for convenient reference, we are listing here some conditions on the asymptotic behavior, as $\mathbf{n} \to \infty$, of the bandwidth $h_{\mathbf{n}}$ that will be used in the sequel.

(B1) The bandwith $h_{\mathbf{n}}$ tends to zero in such a way that $\hat{\mathbf{n}}h_{\mathbf{n}}^d \to \infty$ as $\mathbf{n} \to \infty$.

(B2) Same as B1, but moreover $\hat{\mathbf{n}}h_{\mathbf{n}}^{4+d} = O(1)$ as $\mathbf{n} \to \infty$.

Also, we denote by $F_{Y|\mathbf{X}}(y|\mathbf{x}) := P(Y_{\mathbf{i}} < y|\mathbf{X}_{\mathbf{i}} = \mathbf{x})$ and $f_{Y|\mathbf{X}}(y|\mathbf{x})$ the conditional distribution and conditional density functions of $Y_{\mathbf{i}}$ given $\mathbf{X}_{\mathbf{i}} = \mathbf{x}$, respectively. The notation $\mathbf{M}'$ is used for the transpose of a matrix or vector $\mathbf{M}$.

2.2 Local fitting of the spatial quantile regression function.

In this section we extend the traditional local linear fitting approach to estimate the spatial quantile regression function.

Write $\dot{q}_p(\mathbf{x}) = (\partial q_p(\mathbf{x})/\partial x_1, \cdots, \partial q_p(\mathbf{x})/\partial x_d)'$ for the vector of the first order partial derivatives of the quantile regression function $q_p(\mathbf{x})$ at $\mathbf{x} = (x_1, \cdots, x_d)' \in \mathbb{R}^d$. The general idea of local linear fitting, which can be found in Fan and Gijbels (1996) and Loader (1999), consists in approximating the unknown quantile regression function $q_p(\mathbf{z})$ by a linear function in a neighborhood of $\mathbf{x}$ (cf. Fan *et al.*1994, and Yu and Jones 1998):

$$q_p(\mathbf{z}) \approx q_p(\mathbf{x}) + (\dot{q}_p(\mathbf{x}))'(\mathbf{z} - \mathbf{x}) \equiv a_0 + \mathbf{a}_1'(\mathbf{z} - \mathbf{x}). \tag{2.2}$$

Therefore, estimating $(q_p(\mathbf{x}), \dot{q}_p(\mathbf{x}))$ is locally equivalent to estimating $(a_0, \mathbf{a}_1)$. The classical theory of quantile regression suggests estimating a_0 and $\mathbf{a}_1$ as

$$(\widehat{a}_0, \widehat{\mathbf{a}}_1) := \arg\min_{(a_0, \mathbf{a}_1)} \sum_{\substack{i_k=1 \\ k=1,\ldots,N}}^{n_k} \rho_p(Y_{\mathbf{i}} - a_0 - \mathbf{a}_1'(\mathbf{X}_{\mathbf{i}} - \mathbf{x}))K_h(\mathbf{X}_{\mathbf{i}} - \mathbf{x}), \tag{2.3}$$

where $\rho_p(y) := y(p - I_{\{y<0\}})$ stands for the traditional check function $\rho_p(y) := y(p - I_{\{y<0\}})$, I_A is the indicator function of set A, and $K_h(\mathbf{x}) := h_{\mathbf{n}}^{-d} K(\mathbf{x}/h_{\mathbf{n}})$, with a kernel function K defined on $\mathbb{R}^d$, and a bandwidth $h = h_{\mathbf{n}} > 0$ tending to 0 as $\mathbf{n} \to \infty$. This motivates the choice of $\widehat{q}_p(\mathbf{x}) \equiv \widehat{a}_0$ and $\widehat{\dot{q}}_p(\mathbf{x}) \equiv \widehat{\mathbf{a}}_1$ as estimators of $q_p(\mathbf{x})$ and $\dot{q}_p(\mathbf{x})$, respectively.

Note that the definition (2.3) of the estimator does not require the regular grid structure we assuming throughout. It seems intuitively clear that "nearly regular grids" will not harm the results of this paper. However, the asymptotic treatment of irregular grids (essentially, a definition of a "nearly regular grid") is a delicate and problematic issue that we will not consider here.

2.3 Bahadur representation

The definition in (2.3) looks simple, but, unlike the local linear fitting estimator for spatial mean regression proposed in Hallin, Lu and Tran (2004b), (2.3) does not allow for an explicit solution, which creates additional difficulties in developing the asymptotic theory. We overcome these difficulties by obtaining a Bahadur representation for for $\widehat{q}_p$ and $\widehat{\dot{q}}_p$. Note that the following only requires $\mathbf{x} \mapsto q_p(\mathbf{x})$ to be continuously differentiable.

Theorem 2.1 (Bahadur representation) *Let Assumptions A1, A3, A4 (for some $a > N$), and B1 hold, and assume that the quantile function q_p has a continuous first order derivative at $\mathbf{x}$. Then,*

$$(\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{1/2} \begin{bmatrix} \widehat{q}_p(\mathbf{x}) - q_p(\mathbf{x}) \\ h_{\mathbf{n}}(\widehat{\dot{q}}_p(\mathbf{x}) - \dot{q}_p(\mathbf{x})) \end{bmatrix} = \eta_p(\mathbf{x}) \frac{1}{\sqrt{\widehat{\mathbf{n}} h_{\mathbf{n}}^d}} \sum_{\substack{i_k=1 \\ k=1,\ldots,N}}^{n_k} \psi_p(Y_{\mathbf{i}}^*) \begin{bmatrix} 1 \\ \frac{\mathbf{X}_{\mathbf{i}} - \mathbf{x}}{h_{\mathbf{n}}} \end{bmatrix} K\left(\frac{\mathbf{X}_{\mathbf{i}} - \mathbf{x}}{h_{\mathbf{n}}}\right) + o_P(1),$$

as $\mathbf{n} \to \infty$, where $\psi_p(y) := p - I_{\{y<0\}}$, $Y_{\mathbf{i}}^* := Y_{\mathbf{i}}^*(p) = Y_{\mathbf{i}} - q_p(\mathbf{x}) - (\dot{q}_p(\mathbf{x}))'(\mathbf{X}_{\mathbf{i}} - \mathbf{x})$, and $\eta_p(\mathbf{x}) := (f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}))^{-1}$.

The proof of Theorem 2.1 is postponed to Section 4.2.

2.4 Asymptotic normality

Using the powerful tool of the Bahadur representation, we can establish the consistency and asymptotic distribution of the local linear quantile regression estimates under weak conditions.

First, we consider the case where the sample size tends to ∞ in the manner of Tran (1990), that is, $\mathbf{n} \Longrightarrow \infty$.

Assuming that Assumption A2 holds, so that q_p is twice differentiable, let

$$B_0(\mathbf{x}) := \{f_{\mathbf{X}}(x)\}^{-1} \operatorname{tr}\left[\ddot{q}_p(\mathbf{x}) \int \mathbf{u}\mathbf{u}' K(\mathbf{u}) d\mathbf{u}\right] \quad \text{and} \quad \mathbf{B}_1(\mathbf{x}) := (B_{11}(\mathbf{x}), \cdots, B_{1d}(\mathbf{x}))',$$

with $B_{1j}(\mathbf{x}) := f_{\mathbf{X}}^{-1}(x) \operatorname{tr}\left[\ddot{q}_p(\mathbf{x}) \int \mathbf{u}\mathbf{u}' u_j K(\mathbf{u}) d\mathbf{u}\right]$, $j = 1, \cdots, d$, $\sigma_0^2(\mathbf{x}) := \eta^*(\mathbf{x}) \int K^2(\mathbf{u}) d\mathbf{u}$, and $\boldsymbol{\sigma}_1^2(\mathbf{x}) := \eta^*(\mathbf{x}) \int \mathbf{u}\mathbf{u}' K^2(u)\, d\mathbf{u}$, where $\eta^*(\mathbf{x}) := \eta_p^2(\mathbf{x}) p(1-p) f_{\mathbf{X}}(\mathbf{x}) = \dfrac{p(1-p)}{f_{\mathbf{X}}(\mathbf{x}) f_{Y|\mathbf{X}}^2(q_p(\mathbf{x})|\mathbf{x})}$

Theorem 2.2 *Let Assumptions A1, A2, A3′, A4 (with $\varphi(x) = O(x^{-\mu})$ as $x \to \infty$ for some $\mu > 2N$), and B2 hold. Suppose that there exists a sequence of positive integers $q = q_{\mathbf{n}}$ such that $q_{\mathbf{n}} \to \infty$, $q_{\mathbf{n}} = o\left((\widehat{\mathbf{n}} h_{\mathbf{n}}^d)^{1/2N}\right)$, and $\widehat{\mathbf{n}} q^{-\mu} \to 0$ as $\mathbf{n} \Longrightarrow \infty$. Moreover, let the bandwidth $h_{\mathbf{n}}$ tend to zero in such a manner that*

$$\liminf_{\mathbf{n} \Longrightarrow \infty} q h_{\mathbf{n}}^{d/a} > 1 \quad \text{for some } N < a < \mu - N. \tag{2.4}$$

Then, for any $\mathbf{x}$ and $0 < p < 1$, as $\mathbf{n} \Longrightarrow \infty$,

$$\sqrt{\widehat{\mathbf{n}} h_{\mathbf{n}}^d} \left[\begin{pmatrix} \widehat{q}_p(\mathbf{x}) - q_p(\mathbf{x}) \\ h_{\mathbf{n}}(\widehat{\dot{q}}_p(\mathbf{x}) - \dot{q}_p(\mathbf{x})) \end{pmatrix} - \frac{1}{2} \begin{pmatrix} B_0(\mathbf{x}) \\ \mathbf{B}_1(\mathbf{x}) \end{pmatrix} h_{\mathbf{n}}^2 \right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma_0^2(\mathbf{x}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}_1^2(\mathbf{x}) \end{pmatrix}\right),$$

so that $\widehat{q}_p(\mathbf{x})$ and $\widehat{\dot{q}}_p(\mathbf{x})$ are asymptotically independent.

The asymptotic normality results in Theorem 2.2 are stated for $\widehat{q}_p(\mathbf{x})$ and $\widehat{\dot{q}}_p(\mathbf{x})$ at a given site $\mathbf{x}$. They are easily extended, via the traditional Cramér-Wold device, into a joint asymptotic normality result for any couple $(\mathbf{x}_1, \mathbf{x}_2)$ (or any finite collection) of sites; the asymptotic covariance terms (between $\widehat{q}_p(\mathbf{x}_1)$ and $\widehat{q}_p(\mathbf{x}_2)$, $\widehat{q}_p(\mathbf{x}_1)$ and $\widehat{\dot{q}}_p(\mathbf{x}_2)$, etc.) all are equal to zero (cf. Hallin, Lu and Tran, 2004b, page 2478). The same remark also holds for Theorems 2.3-2.6 below.

One of the important advantages of local polynomial (and linear) fitting over the more traditional Nadaraya-Watson approach is that it has much better boundary behavior. This advantage often has been emphasized in the usual regression and time-series settings when the regressors take values on a compact subset of $\mathbb{R}^d$. For example, as Fan and Gijbels (1996) and Fan and Yao (2003) illustrate for mean regression, for a univariate ($d = 1$) regressor X with bounded support ([0, 1], say), it can be proved, using an argument similar to the one developed in the proof of Theorem 3.1 of Hallin *et al.* (2004b), that asymptotic normality still holds at the origin, but with asymptotic bias and variances

$$B_0 = \{f_X(0^+)\}^{-1} \left[\ddot{q}_p(0^+) \int_{-c}^{\infty} u^2 K(u) du\right] \quad \text{and} \quad \sigma_0^2 = \eta^*(0^+) \int_{-c}^{\infty} K^2(u) du,$$

and

$$B_1 = \{f_X(0^+)\}^{-1} \left[\ddot{q}_p(0^+) \int_{-c}^{\infty} u^3 K(u) du\right] \quad \text{and} \quad \sigma_1^2 = \eta^*(o^+) \int_{-c}^{\infty} u^2 K^2(u)\, du,$$

respectively, where $\eta^*(0^+) = \eta_p^2(0^+) p(1-p) f_X(0^+) = \dfrac{p(1-p)}{f_X(0^+) f_{Y|X}^2(q_p(0^+)|0^+)}$. As pointed out in Hallin *et al.* (2004b), this advantage is likely to be much more substantial as N is growing.

In the important particular case under which $\varphi(x)$ tends to zero at exponential rate, the same results are obtained under milder conditions.

Theorem 2.3 *Let Assumptions A1, A2, A3′, and A4 hold, with $\varphi(x) = O(e^{-\xi x})$ as $x \to \infty$ for some $\xi > 0$. Then, if $h_{\mathbf{n}}$ tends to zero as $\mathbf{n} \Longrightarrow \infty$ in such a manner that $(\widehat{\mathbf{n}} h_{\mathbf{n}}^{d(1+2N/a)})^{1/2N} (\log \widehat{\mathbf{n}})^{-1} \to \infty$ for some $a > N$, the conclusions of Theorem 2.2 still hold.*

Note that, for $N = 1$, and for "large" values of a, this condition is "close" to the classical condition (for independent observations) that $nh_n^d \to \infty$.

Next, we consider the situation under which the sample size tends to ∞ in the "weak" sense (that is, $\mathbf{n} \to \infty$ instead of $\mathbf{n} \Longrightarrow \infty$).

Theorem 2.4 *Let Assumptions A1, A2, A3′, and A4 hold, with $\varphi(x) = O(x^{-\mu})$ as $x \to \infty$ for some $\mu > 2N$. Let the sequence of positive integers $q = q_{\mathbf{n}} \to \infty$, and let the bandwidth $h_{\mathbf{n}}$ factorize into $h_{\mathbf{n}} := \prod_{i=1}^{N} h_{n_i}$, such that $\widehat{\mathbf{n}} q^{-\mu} \to 0$, $q = o((\min_{1 \le k \le N}(n_k h_{n_k}^d))^{1/2})$, and $\liminf_{\mathbf{n} \to \infty} q h_{\mathbf{n}}^{d/a} > 1$ for some $N < a < \mu - N$. Then the conclusions of Theorem 2.2 hold as $\mathbf{n} \to \infty$.*

In the important case that $\varphi(x)$ tends to zero at an exponential rate, we have the following result, which parallels Theorem 2.3.

Theorem 2.5 *Let Assumptions A1, A2, A3′, and A4 hold, with $\varphi(x) = O(e^{-\xi x})$ as $x \to \infty$ for some $\xi > 0$. Let the bandwidth $h_{\mathbf{n}}$ factorize into $h_{\mathbf{n}} := \prod_{i=1}^{N} h_{n_i}$ in such a way that, as $\mathbf{n} \to \infty$, $\min_{1 \le k \le N}\{(n_k h_{n_k}^d)^{1/2}\} h_{\mathbf{n}}^{d/a} (\log \widehat{\mathbf{n}})^{-1} \to \infty$ for some $a > N$. Then the conclusions of Theorem 2.2 hold as $\mathbf{n} \to \infty$.*

Under (A3″), we then have the following counterpart of Theorem 2.2.

Theorem 2.6 *Let Assumptions A1, A2, A3″, and A4 hold, with $\varphi(x) = O(x^{-\mu})$ as $x \to \infty$ for some $\mu > 2N$. Suppose that there exists a sequence of positive integers $q = q_{\mathbf{n}} \to \infty$ such that $q_{\mathbf{n}} = o((\widehat{\mathbf{n}} b_{\mathbf{n}}^d)^{1/2N})$ and $\widehat{\mathbf{n}}^{\kappa+1} q^{-\mu-N} \to 0$ as $\mathbf{n} \Longrightarrow \infty$, and that the bandwidth $h_{\mathbf{n}}$ tends to zero in such a manner that (2.4) is satisfied as $\mathbf{n} \Longrightarrow \infty$. Then the conclusions of Theorem 2.2 hold as $\mathbf{n} \Longrightarrow \infty$.*

Analogues of Theorems 2.3, 2.4, and 2.5 can also be obtained under Assumption (A3″); details are left to the reader. The proofs of Theorems 2.2-2.6 are given in Section 4.3.

3 Numerical example

The wheat data (Mercer and Hall, 1911) were collected in 1910 at the Rothamsted Experimental Station. They consist of 500 yields obtained on a 20×25 lattice of plots approximately 1 acre in total area; the 20 rows run in the East-West direction, and the 25 columns in the North-South direction. This data set has been analyzed via parametric spatial regression models by several authors, among which Besag (1974), McBratney and Webster (1981), and Cressie (1993, Section 4.51). Besag (1974) implemented his *first-order auto-normal scheme* method, McBratney and Webster (1981) and Cressie (1993) their *two-way parametric effect analysis.* These methods all model the expected wheat yield by a linear regression model involving two regressors: the sum of the two row-neighbours and that of the two column-neighbours, thus artificially imposing East-West and North-South symmetric (isotropic) spatial influences. They moreover assume the homogeneous dependence of wheat yield for spatial variation: Besag (1974) for instance assumes that, denoting by $Y_{i,j}$ the yield in the ith row and jth column ($i = 1, ..., 20$; $j = 1, ..., 25$), the

$Y_{i,j}$'s are multinormal, and that the conditional expectation of $Y_{i,j}$, given all other site values, is of the form $\gamma_0 + \gamma_1(Y_{i-1,j} + Y_{i+1,j}) + \gamma_2(Y_{i,j-1} + Y_{i,j+1})$ ("first-order scheme"). Besag's analysis leads to the estimates $\hat{\gamma}_0 = 0.16$, $\hat{\gamma}_1 = 0.34$, $\hat{\gamma}_2 = 0.14$: row- (East-West-) neighbours thus are better predictors than column- (North-South-) ones. The surface (Figure 1b) going through the estimated conditional means $\hat{y}_{ij} := \hat{\gamma}_0 + \hat{\gamma}_1(Y_{i-1,j} + Y_{i+1,j}) + \hat{\gamma}_2(Y_{i,j-1} + Y_{i,j+1})$ provides an estimation of the spatial variations of expected yields across the grid. This surface exhibits a visible North-South trend, with a crest, followed by a trough, both parallel to the East-West direction. This North-South trend is a finding of the analysis, which is by no means apparent in the data themselves (which look quite "horizontal"). As we shall see, it is probably a consequence of the model choice, rather than an actual spatial variation of expected yields.

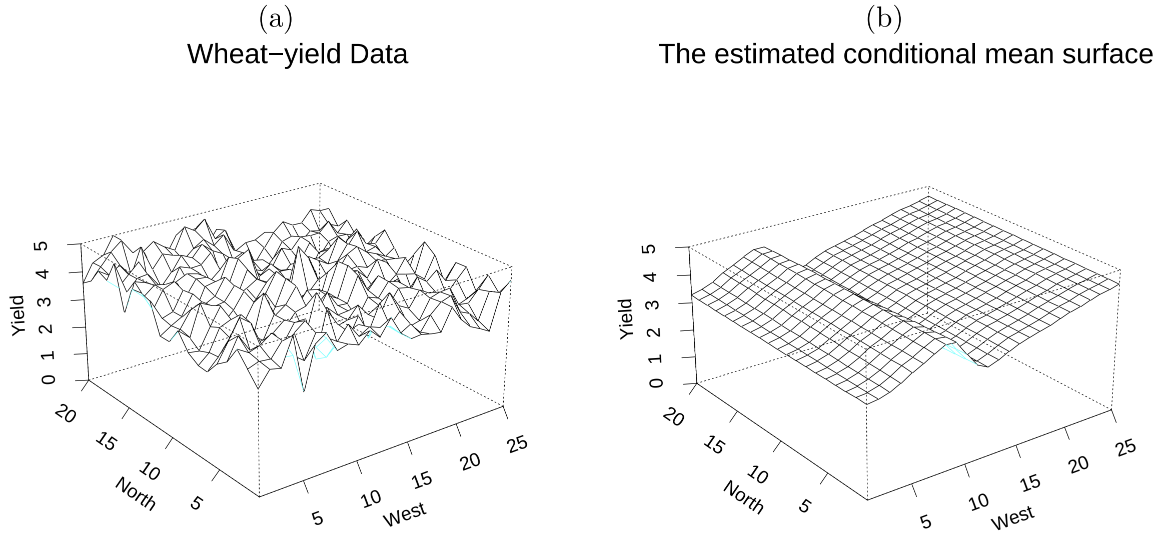


Figure 1: **Wheat-yield data.** *Three-dimensional perspective of (a) the observations, (b) the smoothed surface based on expected yields estimated via Besag's first order autonormal scheme.*

Besag himself in his concluding remarks indicates that his autonormal schemes (a "second-order scheme" appears to be equally unsatisfactory) fail to provide a fully convincing fit of the data. Our approach, which avoids parametric (linear) specification of the conditional mean function, constancy of conditional variances and Gaussian assumptions, is more flexible, and provides a more complete picture of the process under study.

First, our method can be used to check whether the assumptions of Besag's autonormal scheme are plausible. For example, using our method for estimating the pth quantile ($0 < p < 1$) of Y_{ij} conditional on $X_{i,j} := 0.34(Y_{i-1,j} + Y_{i+1,j}) + 0.14(Y_{i,j-1} + Y_{i,j+1})$, then plotting these quantiles against $X_{i,j}$, one should obtain, under sufficient smoothing, roughly parallel horizontal lines.

Figure 2 displays the observations (scatter plot) of $Y_{i,j}$ against X_{ij}, along with the estimated conditional quantile curves for p=10%, 50%, and 90%, respectively, along with their asymptotic 90% confidence intervals. Although no rigorous testing procedure is performed, an informal inspection of this figure suggests a strong nonlinear pattern in the three curves, indicating that the assumptions of the model are unlikely to hold, hence that the conclusions (Figure 1(b)) should be considered with caution.

Breaking with Besag's approach, we performed two different quantile analyses:

(a) maintaining Besag's assumption of East-West and North-South isotropic spatial influences, we estimated the quantiles of order p=10%, 50%, and 90% of $Y_{i,j}$ conditional on the two-dimensional covariate $\mathbf{X}_{i,j} =: (X_{ij}^{(1)}, X_{ij}^{(2)})$, where $X_{ij}^{(1)} := Y_{i-1,j} + Y_{i+1,j}$ and $X_{ij}^{(2)} := Y_{i,j-1} + Y_{i,j+1}$;

(b) abandoning the isotropy assumption, we estimated the quantiles of order p=10%, 50%, and 90% of $Y_{i,j}$ conditional on $\mathbf{X}_{i,j} =: (Y_{i-1,j}, Y_{i+1,j}, Y_{i,j-1}, Y_{i,j+1})$ (a four-dimensional covariate).

The results are presented in Figures 3 and 4, respectively. In order to facilitate the comparison with Besag, we provide smoothed three-dimensional perspectives of the estimated 10%, median, and 90% conditional quantile surfaces computed at the observations: at each site (i, j), the surface provides an estimation of the quantiles of Y_{ij} conditional on $\mathbf{X}_{ij}$ taking the observed value $\mathbf{x}_{ij}$. All three surfaces in Figure 3 exhibit globally increasing North-South profiles, with a peak followed by a trough; this peak is hardly detected in the 10% surface, and sharper in the 90% than in the median surface. East-West profiles, on the other hand, look quite flat. The median surface is pretty much the same as that obtained by Besag, despite the fact that his model assumptions clearly are not satisfied.

Interpretation of Figure 3 might be as follows: on the Northern boundary of the grid, where median yields are low, conditional distributions are also more spread out (mainly, on the left-hand side of the median) than on the Southern boundary, where yields are better: high left-tail variability, thus, with low median yields in the North, stable high yields in the South. East-West spatial variations seem negligible (hence East-West isotropy is not implausible), but North-South isotropy is clearly unreasonable.

These conclusions however are strongly invalidated by the non-isotropic analysis. Turning to Figure 4, indeed, the three estimated quantile surfaces appear to be much less parallel than in Figure 3. While an East-West trend is still present, and quite marked for the 10% surface, it is hardly present anymore in the median and 90% ones. Again, the left-tail conditional spread is much larger on the Northern than on the Southern boundary of the grid, whereas East-West spatial variations remain negligible. The right conclusion thus is that median yields are roughly stable over the grid, and that the North-South mean-trend detected in Besag (1974), as well as the median-trend in the isotropic analysis of Figure 3 are essentially due to conditional heteroskedasticity (more precisely, to left-tail spread): this variation in the spread seems to be the main spatial issue in the data set. The difference between Figure 3 and Figure 4 are entirely imputable to the isotropy assumption made in the first one, which artificially attributes the same influence to a Northern neighbor with larger left-tail spread and a Southern one, for which this left-tail spread is considerably less. Such a phenomenon is undetectable in classical approaches, where only conditional location (typically, conditional mean) is modelled.

4 Appendix: Proofs

4.1 A preliminary lemma

The following lemma is an improved version of the cross-term inequality of Lemma 5.2 of Hallin *et al.*(2004b), adapted to the quantile regression context, and plays a crucial role in the subsequent sections. For the sake of generality, and in order for this lemma to apply beyond the specific framework of this paper, we do not necessarily assume that the mixing coefficient α take the form imposed in Assumption (A3).

Lemma 4.1 (Cross-term Lemma) *Let* $\{(Y_{\mathbf{j}}, \mathbf{X}_{\mathbf{j}}); \mathbf{j} \in \mathbb{Z}^N\}$ *denote a stationary spatial process with general mixing coefficient*

$$\varphi(\mathbf{j}) = \varphi(j_1, \ldots, j_N) := \sup\Big\{|\mathrm{P}(AB) - \mathrm{P}(A)\mathrm{P}(B)| : A \in \mathcal{B}(\{Y_{\mathbf{i}}, \mathbf{X}_{\mathbf{i}}\}),\, B \in \mathcal{B}(\{Y_{\mathbf{i+j}}, \mathbf{X}_{\mathbf{i+j}})\}\Big\}.$$

Let $(y, \mathbf{x}) \mapsto \tilde{b}(y, \mathbf{x})$ *be a bounded Borel-measurable function defined on* $\mathbb{R}^1 \times \mathbb{R}^d$*. Set*

$$\eta_{\mathbf{j}}(\mathbf{x}) := \tilde{b}(Y_{\mathbf{j}}, \mathbf{X}_{\mathbf{j}})K((\mathbf{x} - \mathbf{X}_{\mathbf{j}})/h_{\mathbf{n}}), \qquad \Delta_{\mathbf{j}}(\mathbf{x}) := \eta_{\mathbf{j}}(\mathbf{x}) - E\eta_{\mathbf{j}}(\mathbf{x}),$$

and $\tilde{R}(\mathbf{x}) := (\hat{\mathbf{n}}h_{\mathbf{n}}^d)^{-1} \sum_{\{\mathbf{i},\mathbf{j} \in \mathcal{I}_{\mathbf{n}} | \exists k : i_k \neq j_k\}} \sum \mathrm{E}\left[\Delta_{\mathbf{i}}(\mathbf{x})\Delta_{\mathbf{j}}(\mathbf{x})\right]$*. For any* $\mathbf{c_n} := (c_{\mathbf{n}1}, \cdots, c_{\mathbf{n}N}) \in \mathbb{Z}^N$ *with* $1 < c_{\mathbf{n}k} < n_k$ *for all* $k = 1, \cdots, N$*, define* $\tilde{J}_1(\mathbf{x}) := h_{\mathbf{n}}^{2d} \prod_{k=1}^N (n_k c_{\mathbf{n}k})$ *and*

$$\tilde{J}_2(\mathbf{x}) := \hat{\mathbf{n}} \sum_{k=1}^N \left(\sum_{\substack{|j_s|=1 \\ s=1,\cdots,k-1}}^{n_s} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{\substack{|j_s|=1 \\ s=k+1,\cdots,N}}^{n_s} \{\varphi(j_1, \cdots, j_N)\} \right).$$

Then, under Assumptions A1, A2, and A4, there exists a constant $C > 0$ *such that*

$$|\tilde{R}(\mathbf{x})| \leq C(\hat{\mathbf{n}}h_{\mathbf{n}}^d)^{-1}\left[\tilde{J}_1(\mathbf{x}) + \tilde{J}_2(\mathbf{x})\right]. \tag{4.5}$$

If furthermore $\varphi(j_1, \ldots, j_N)$ *takes the form* $\varphi(\|\mathbf{j}\|)$*, then*

$$\tilde{J}_2(\mathbf{x}) \leq C\hat{\mathbf{n}} \sum_{k=1}^N \left(\sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1}\varphi(t) \right). \tag{4.6}$$

Proof. The main idea of the proof is similar to that of Lemma 5.2 of Hallin *et al.* (2004b), but details are different. We only sketch the proof here. Writing $Z_{\mathbf{j}}$ for $\tilde{b}(Y_{\mathbf{j}}, \mathbf{X}_{\mathbf{j}})$, we have $\eta_{\mathbf{j}}(\mathbf{x}) = Z_{\mathbf{j}}K((\mathbf{x} - \mathbf{X}_{\mathbf{j}})/h_{\mathbf{n}})$, where $|Z_{\mathbf{j}}|$ is bounded by some $L > 0$. For $\mathbf{i} \neq \mathbf{j}$, letting $K_{\mathbf{n}}(\mathbf{x}) := (1/h_{\mathbf{n}}^d)K(\mathbf{x}/h_{\mathbf{n}})$,

$$\begin{aligned} &h_{\mathbf{n}}^{-d}\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x}) \\ &\quad = h_{\mathbf{n}}^d \int\!\!\int K_{\mathbf{n}}(\mathbf{x} - \mathbf{u})K_{\mathbf{n}}(\mathbf{x} - \mathbf{v})\{g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v})f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - g_1^{(1)}(\mathbf{u})g_1^{(1)}(\mathbf{v})f(\mathbf{u})f(\mathbf{v})\}d\mathbf{u}d\mathbf{v}, \end{aligned}$$

where $g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v}) := \mathrm{E}(Z_{\mathbf{i}}Z_{\mathbf{j}} | \mathbf{X}_{\mathbf{i}} = \mathbf{u}, \mathbf{X}_{\mathbf{j}} = \mathbf{v})$, and $g_1^{(1)}(\mathbf{u}) := \mathrm{E}(Z_{\mathbf{i}} | \mathbf{X}_{\mathbf{i}} = \mathbf{u})$. Since $|Z_{\mathbf{i}}|$ is bounded by L, we have that $|g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v})| \leq L^2$ and $|g_1^{(1)}(\mathbf{u})g_1^{(1)}(\mathbf{v})| \leq L^2$. Thus,

$$\begin{aligned} &|g_{1\mathbf{ij}}(\mathbf{u}, \mathbf{v})f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - g_1^{(1)}(\mathbf{u})g_1^{(1)}(\mathbf{v})f(\mathbf{u})f(\mathbf{v})| \\ &\qquad \leq \; L^2|f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - f(\mathbf{u})f(\mathbf{v})| + 2L^2 f(\mathbf{u})f(\mathbf{v}). \end{aligned}$$

It then follows from Assumption (A1) and the Lebesgue density theorem (see Chapter 2 of Devroye and Györfi 1985) that

$$\begin{aligned} h_{\mathbf{n}}^{-d}|\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})| \;&\leq\; h_{\mathbf{n}}^d \int\!\!\int K_{\mathbf{n}}(\mathbf{x} - \mathbf{u})K_{\mathbf{n}}(\mathbf{x} - \mathbf{v})L^2|f_{\mathbf{i},\mathbf{j}}(\mathbf{u}, \mathbf{v}) - f(\mathbf{u})f(\mathbf{v})|\, d\mathbf{u}d\mathbf{v} \\ &\qquad + h_{\mathbf{n}}^d \int\!\!\int K_{\mathbf{n}}(\mathbf{x} - \mathbf{u})K_{\mathbf{n}}(\mathbf{x} - \mathbf{v})2L^2 f(\mathbf{u})f(\mathbf{v})\}\, d\mathbf{u}d\mathbf{v} \\ &\leq\; Ch_{\mathbf{n}}^d L^2 = Ch_{\mathbf{n}}^d. \end{aligned} \tag{4.7}$$

Let $c_{\mathbf{n}} = (c_{\mathbf{n}1}, \cdots, c_{\mathbf{n}N}) \in \mathbb{R}^N$ be a sequence of vectors with positive components. Define

$$\mathcal{S}_1 := \{\mathbf{i} \neq \mathbf{j} \in \mathcal{I}_{\mathbf{n}} : |j_k - i_k| \leq c_{\mathbf{n}k}, \text{ for all } k = 1, \cdots, N\},$$

and

$$\mathcal{S}_2 := \{\mathbf{i}, \mathbf{j} \in \mathcal{I}_{\mathbf{n}} : |j_k - i_k| > c_{\mathbf{n}k}, \text{ for some } k = 1, \cdots, N\}.$$

Clearly, $\text{Card}(\mathcal{S}_1) \leq 2^N \hat{\mathbf{n}} \prod_{k=1}^N c_{\mathbf{n}k}$. Splitting $\tilde{R}(\mathbf{x})$ into $(\hat{\mathbf{n}} h_{\mathbf{n}}^d)^{-1}(J_1 + J_2)$, with $J_\ell := \sum_{\mathbf{i},\mathbf{j}} \sum_{\in \mathcal{S}_\ell} \mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})$, $\ell = 1, 2$, it follows from (4.7) that

$$|J_1| \leq C h_{\mathbf{n}}^{2d} \text{Card}(\mathcal{S}_1) \leq 2^N C h_{\mathbf{n}}^{2d} \hat{\mathbf{n}} \prod_{k=1}^N c_{\mathbf{n}k}. \tag{4.8}$$

Turning to J_2, we have $|J_2| \leq \sum_{\mathbf{i},\mathbf{j}} \sum_{\in \mathcal{S}_2} |\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})|$. Davydov's inequality (cf. Lemma 2.1 of Tran 1990) and the boundedness of $\Delta_{\mathbf{i}}(\mathbf{x})$ yield $|\mathrm{E}\Delta_{\mathbf{j}}(\mathbf{x})\Delta_{\mathbf{i}}(\mathbf{x})| \leq C\varphi(\mathbf{j} - \mathbf{i})$. Hence,

$$|J_2| \leq C \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}_2} \sum \varphi(\mathbf{j} - \mathbf{i}) =: C\Sigma_2, \quad \text{say.}$$

We now analyze Σ_2 in detail. For any N-tuple $\mathbf{0} \neq \boldsymbol{\ell} = (\ell_1, \ldots, \ell_N) \in \{0,1\}^N$, set

$$\mathcal{S}(\ell_1, \ldots, \ell_N) := \{\mathbf{i}, \mathbf{j} \in \mathcal{I}_{\mathbf{n}} : |j_k - i_k| > c_{\mathbf{n}k} \text{ if } \ell_k = 1 \text{ and } |j_k - i_k| \leq c_{\mathbf{n}k} \text{ if } \ell_k = 0,\ k = 1, \cdots, N\}$$

and $V(\ell_1, \ldots, \ell_N) := \sum_{\mathbf{i},\mathbf{j} \in \mathcal{S}(\ell_1, \ldots, \ell_N)} \sum \varphi(\mathbf{j} - \mathbf{i})$. Then, $\Sigma_2 = \sum_{\mathbf{i},\mathbf{j}} \sum_{\in \mathcal{S}_2} \varphi(\mathbf{j} - \mathbf{i}) = \sum_{\mathbf{0} \neq \boldsymbol{\ell} \in \{0,1\}^N} V(\ell_1, \ldots, \ell_N)$ where, as in (5.11) of Hallin *et al.* (2004b),

$$V(\ell_1, \ell_2, \ldots, \ell_N) \leq \hat{\mathbf{n}} \sum_{|j_1|} \ldots \sum_{|j_k|} \ldots \sum_{|j_N|} \varphi(j_1, \ldots, j_N),$$

with the sums $\sum_{|j_k|}$ running over all values of j_k such that $1 \leq |j_k| \leq n_k$ if $\ell_k = 0$, such that $c_{\mathbf{n}1} \leq |j_k| \leq n_k$ if $\ell_k = 1$. Since the summands are nonnegative, for $1 \leq c_{\mathbf{n}k} \leq n_k$, we have $\sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \cdots \leq \sum_{|j_k|=1}^{n_k} \cdots$, and

$$|J_2| \leq C\hat{\mathbf{n}} \sum_{k=1}^N \left(\sum_{|j_1|=1}^{n_1} \cdots \sum_{|j_{k-1}|=1}^{n_{k-1}} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{|j_{k+1}|=1}^{n_{k+1}} \cdots \sum_{|j_N|=1}^{n_N} \varphi(j_1, \ldots, j_N) \right). \tag{4.9}$$

Thus (4.5) is a consequence of (4.8) and (4.9). If furthermore $\varphi(j_1, \ldots, j_N)$ depends on $\|\mathbf{j}\|$ only, then (4.6) follows from the fact that

$$\sum_{|j_1|=1}^{n_1} \cdots \sum_{|j_{k-1}|=1}^{n_{k-1}} \sum_{|j_k|=c_{\mathbf{n}k}}^{n_k} \sum_{|j_{k+1}|=1}^{n_{k+1}} \cdots \sum_{|j_N|=1}^{n_N} \varphi(\|\mathbf{j}\|) \leq \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} \sum_{|j_1|=1}^{t} \cdots \sum_{|j_{N-1}|=1}^{t} \varphi(t) \leq \sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1}\varphi(t). \quad \square$$

4.2 Proof of the Bahadur representation result

We first introduce some notation. Throughout the proof, let C denote a generic positive constant. Set

$$\mathbf{X}_{h\mathbf{i}} := (\mathbf{X}_{\mathbf{i}} - \mathbf{x})/h_{\mathbf{n}}, \quad \mathcal{X}_{h\mathbf{i}} := (1,\, \mathbf{X}'_{h\mathbf{i}})', \quad K_{\mathbf{i}} := K(\mathbf{X}_{h\mathbf{i}}), \quad H_{\mathbf{n}} = \sqrt{\widehat{\mathbf{n}} h_{\mathbf{n}}^d},$$

$$\overline{\theta}_{\mathbf{n}} := H_{\mathbf{n}}\left(\widehat{a}_0 - q_p(\mathbf{x}), \quad h_{\mathbf{n}}(\widehat{\mathbf{a}}_1 - \dot{q}(\mathbf{x}))'\right)', \quad \theta := H_{\mathbf{n}}\left(a_0 - q_p(\mathbf{x}), \quad h_{\mathbf{n}}(\mathbf{a}_1 - \dot{q}(\mathbf{x}))'\right)',$$

and $\tilde{\theta} := H_{\mathbf{n}}\left(\tilde{a}_0 - q_p(\mathbf{x}),\, h_{\mathbf{n}}(\tilde{\mathbf{a}}_1 - \dot{q}(\mathbf{x}))'\right)'$, where $(a_0,\, \mathbf{a}'_1)'$, $(\tilde{a}_0,\, \tilde{\mathbf{a}}'_1)' \in \mathbb{R}^{1+d}$. With $Y^*_{\mathbf{i}}$ defined in Theorem 2.1, put $Y^*_{\mathbf{ni}}(\theta) := Y^*_{\mathbf{i}} - \theta'\mathcal{X}_{h\mathbf{i}}/H_{\mathbf{n}}$, $T_{\mathbf{ni}} := (\dot{q}_p(\mathbf{x}))'\mathbf{X}_{h\mathbf{i}} h_{\mathbf{n}}$, and $U_{\mathbf{ni}} := U_{\mathbf{ni}}(\theta) = T_{\mathbf{ni}} + \theta'\mathcal{X}_{h\mathbf{i}}/H_{\mathbf{n}}$. Based on these notations, $Y^*_{\mathbf{i}} = Y_{\mathbf{i}} - q_p(\mathbf{x}) - T_{\mathbf{ni}}$ and $Y^*_{\mathbf{ni}}(\theta) = Y_{\mathbf{i}} - q_p(\mathbf{x}) - U_{\mathbf{ni}}(\theta) = Y_{\mathbf{i}} - a_0 - \mathbf{a}'_1(\mathbf{X}_{\mathbf{i}} - \mathbf{x})$. Since K is a bounded function with bounded support,

$$\|\mathbf{X}_{h\mathbf{i}}\| \le C, \quad \|\mathcal{X}_{h\mathbf{i}}\| \le C \text{ when } K_{\mathbf{i}} > 0. \tag{4.10}$$

When $\|\theta\| \le M$ and $K_{\mathbf{i}} > 0$, $|T_{\mathbf{ni}}| \le C h_{\mathbf{n}}$ and $|U_{\mathbf{ni}}| \le C h_{\mathbf{n}} + C H_{\mathbf{n}}^{-1} \to 0$ as $\mathbf{n} \to \infty$. It follows from (2.3) that

$$\overline{\theta}_{\mathbf{n}} = \operatorname{argmin}_{\theta \in \mathbb{R}^{1+d}} \sum_{\substack{i_k=1\\ k=1,\dots,N}}^{n_k} \rho_p(Y^*_{\mathbf{ni}}(\theta)) K_{\mathbf{i}}. \tag{4.11}$$

Finally, define $V_{\mathbf{n}}(\theta) := H_{\mathbf{n}}^{-1} \sum_{\substack{i_k=1\\ k=1,\dots,N}}^{n_k} \psi_p(Y^*_{\mathbf{ni}}(\theta))\mathcal{X}_{h\mathbf{i}} K_{\mathbf{i}}$. The following lemma provides an asymptotic representation result for sequences $\theta_{\mathbf{n}}$ of solutions of $V_{\mathbf{n}}(\theta) = 0$ or, more generally, for any sequence $\theta_{\mathbf{n}}$ such that $V_{\mathbf{n}}(\theta) = o_P(1)$ as $\mathbf{n} \to \infty$. This result is a spatial version of Lemma A.4 of Koenker and Zhao (1996, page 809), and plays a key role in the proof of Theorem 2.1.

Lemma 4.2 *Let $\Delta \mapsto V_{\mathbf{n}}(\Delta)$ satisfy*

(i) $-\Delta' V_{\mathbf{n}}(\lambda\Delta) \ge -\Delta' V_{\mathbf{n}}(\Delta), \quad \lambda \ge 1,$

(ii) $\sup_{\|\Delta\| \le M} \|V_{\mathbf{n}}(\Delta) + f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x})\mathbf{D}\Delta - \mathbf{A}_{\mathbf{n}}\| = o_P(1)$ *as* $\mathbf{n} \to \infty$, *where* $\|\mathbf{A}_{\mathbf{n}}\| = O_P(1)$, $0 < M < \infty$, $f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x}) > 0$, *and* $\mathbf{D}$ *is a positive definite matrix.*

Suppose that $\Delta_{\mathbf{n}}$ *is such that* $\|V_{\mathbf{n}}(\Delta_{\mathbf{n}})\| = o_P(1)$. *Then,* $\|\Delta_{\mathbf{n}}\| = O_P(1)$, *and*

$$\Delta_{\mathbf{n}} = [f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x})]^{-1}\mathbf{D}^{-1}\mathbf{A}_{\mathbf{n}} + o_P(1) \quad \textit{as} \quad \mathbf{n} \to \infty.$$

Proof. The proof follows along the same lines as in Koenker and Zhao (1996, page 809); details are left to the reader. □

In order to establish the Bahadur representation result of Theorem 2.1, it is now sufficient to check that the assumptions of Lemma 4.2 are satisfied. To do this, we will repeatedly use the next lemma, the proof of which is essentially the same as in the time series case (cf. Lu, Hui, and Zhao 1998) and hence is omitted.

Lemma 4.3 *Let Assumptions A1(ii)-(iii) and A2 hold. Then, for* $\mathbf{n}$ *large enough,*

$$\begin{aligned} E|\psi_p(Y^*_{\mathbf{ni}}(\theta)) - \psi_p(Y^*_{\mathbf{ni}}(\tilde{\theta}))|K_{\mathbf{i}} &\le CEI_{(|Y^*_{\mathbf{ni}}(\tilde{\theta})| < C\|\theta - \tilde{\theta}\|/H_{\mathbf{n}})}\, K_{\mathbf{i}} \le C\|\theta - \tilde{\theta}\| h_{\mathbf{n}}^d / H_{\mathbf{n}}, \\ E|\psi_p(Y^*_{\mathbf{ni}}(\theta)) - \psi_p(Y^*_{\mathbf{ni}}(\tilde{\theta}))|^2 K_{\mathbf{i}}^2 &\le CEI_{(|Y^*_{\mathbf{ni}}(\tilde{\theta})| < C\|\theta - \tilde{\theta}\|/H_{\mathbf{n}})}\, K_{\mathbf{i}}^2 \le C\|\theta - \tilde{\theta}\| h_{\mathbf{n}}^d / H_{\mathbf{n}} \end{aligned}$$

for any $\theta,\ \tilde{\theta} \in \{\theta :\ \|\theta\| \le M\}$.

Lemma 4.4 *Under the conditions of Theorem 2.1,*

$$\sup_{\|\theta\|\le M} \|V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(0) - E(V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(0))\| = o_P(1).$$

Proof. The proof is divided into two steps. The first step consists in proving that

$$\|V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(0) - E(V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(0))\| = o_P(1). \tag{4.12}$$

for any fixed θ such that $\|\theta\| \le M$. Note that

$$V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(0) = H_{\mathbf{n}}^{-1} \sum_{\substack{i_k=1\\k=1,\dots,N}}^{n_k} [\psi_p(Y^*_{\mathbf{ni}}(\theta)) - \psi_p(Y^*_{\mathbf{i}})]\mathcal{X}_{h\mathbf{i}}K_{\mathbf{i}} =: H_{\mathbf{n}}^{-1} \sum_{\substack{i_k=1\\k=1,\dots,N}}^{n_k} V_{\mathbf{ni}}(\theta), \tag{4.13}$$

where $V_{\mathbf{ni}}(\theta) = (V^0_{\mathbf{ni}}(\theta), (V^1_{\mathbf{ni}}(\theta))')'$, with

$$V^0_{\mathbf{ni}}(\theta) = [\psi_p(Y^*_{\mathbf{ni}}(\theta)) - \psi_p(Y^*_{\mathbf{i}})]K_{\mathbf{i}} \quad \text{and} \quad V^1_{\mathbf{ni}}(\theta) = [\psi_p(Y^*_{\mathbf{ni}}(\theta)) - \psi_p(Y^*_{\mathbf{i}})]\mathbf{X}_{h\mathbf{i}}K_{\mathbf{i}}.$$

Then, from (4.13), the left-hand side of (4.12) is bounded by

$$H_{\mathbf{n}}^{-1}\left| \sum_{\substack{i_k=1\\k=1,\dots,N}}^{n_k} (V^0_{\mathbf{ni}}(\theta) - EV^0_{\mathbf{ni}}(\theta))\right| + H_{\mathbf{n}}^{-1}\left\| \sum_{\substack{i_k=1\\k=1,\dots,N}}^{n_k} (V^1_{\mathbf{ni}}(\theta) - EV^1_{\mathbf{ni}}(\theta))\right\| =: V^0_{\mathbf{n}} + V^1_{\mathbf{n}}, \quad \text{say.} \tag{4.14}$$

It follows from stationarity together with Lemma 4.1 that

$$\begin{aligned} E(V^0_{\mathbf{n}})^2 &= (\widehat{\mathbf{n}}h^d_{\mathbf{n}})^{-1}\left[\sum_{\substack{i_k=1\\k=1,\dots,N}}^{n_k} \mathrm{var}(V^0_{\mathbf{ni}}(\theta)) + \sum_{\{\mathbf{i},\mathbf{j}\in\mathcal{I}_{\mathbf{n}}|\exists k\,:\, i_k\ne j_k\}}\sum \mathrm{cov}(V^0_{\mathbf{ni}}(\theta), V^0_{\mathbf{nj}}(\theta))\right] \\ &\le h_{\mathbf{n}}^{-d}\mathrm{var}(V^0_{\mathbf{n}1}(\theta)) + (\widehat{\mathbf{n}}h^d_{\mathbf{n}})^{-1}[\tilde{J}_1(\mathbf{x}) + \tilde{J}_2(\mathbf{x})], \end{aligned} \tag{4.15}$$

where $\tilde{J}_1(\mathbf{x}) \le C\widehat{\mathbf{n}}h^{2d}_{\mathbf{n}}\prod_{k=1}^N c_{\mathbf{n}k}$ and $\tilde{J}_2(\mathbf{x}) \le C\widehat{\mathbf{n}}\sum_{k=1}^N\sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1}\varphi(t)$, as implied by Lemma 4.1. Here $c_{\mathbf{n}k}$, $k = 1, \cdots, N$, are positive integers depending on $\mathbf{n}$, to be specified later on. In order to bound (4.15), we apply Lemma 4.3 with $\tilde{\theta} = 0$; for $\|\theta\| \le M$,

$$\mathrm{var}(V^0_{\mathbf{n}1}(\theta)) \le E(V^0_{\mathbf{n}1})^2 = E|\psi_p(Y^*_{\mathbf{ni}}(\theta)) - \psi_p(Y^*_{\mathbf{i}})|^2K^2_{\mathbf{i}} \le Ch^d_{\mathbf{n}}/H_{\mathbf{n}}.$$

Then it follows from (4.15) with $c_{\mathbf{n}k} = h_{\mathbf{n}}^{-d/a}$ for $k = 1, \cdots, N$, that

$$\begin{aligned} E(V^0_{\mathbf{n}})^2 &\le CH_{\mathbf{n}}^{-1} + Ch^d_{\mathbf{n}}\prod_{k=1}^N c_{\mathbf{n}k} + Ch_{\mathbf{n}}^{-d}\sum_{k=1}^N\sum_{t=c_{\mathbf{n}k}}^{\|\mathbf{n}\|} t^{N-1}\varphi(t) \\ &= CH_{\mathbf{n}}^{-1} + Ch_{\mathbf{n}}^{(1-N/a)d} + C\sum_{k=1}^N c^a_{\mathbf{n}k}\sum_{t=c_{\mathbf{n}k}}^{\infty} t^{N-1}\varphi(t) = o(1), \end{aligned} \tag{4.16}$$

in view of Assumption A3 and the fact that $h_{\mathbf{n}} \to 0$, $\widehat{\mathbf{n}}h^d_{\mathbf{n}} \to \infty$, and $a > N$. Similarly to (4.16), we have $E(V^1_{\mathbf{n}})^2 = o(1)$ which, together with (4.14) and (4.16), implies (4.12).

The second step consists in establishing the uniform consistency with respect to $\|\theta\| \le M$ by a chaining argument. Decompose $\{\|\theta\| \le M\}$ into cubes based on the grid $(j_1\gamma M, \cdots, j_{d+1}\gamma M)$, $j_i = 0, \pm 1, \cdots, \pm[1/\gamma]+1$, where $[1/\gamma]$ denotes the integer part of $1/\gamma$, and γ is a small positive number that does not depend on $\mathbf{n}$. Let $R(\theta)$ be the lower vertex of the cube that contains θ. Clearly, $\|R(\theta) - \theta\| \le C\gamma$ and the number of elements of $\{R(\theta) : \|\theta\| \le M\}$ is finite. Then

$$\sup_{\|\theta\|\le M} \|V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(0) - E(V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(0))\| \le V^*_{\mathbf{n}1} + V^*_{\mathbf{n}2} + V^*_{\mathbf{n}3} \tag{4.17}$$

where, following (4.12), $V^*_{\mathbf{n}1} := \sup_{\|\theta\|\le M} \|V_{\mathbf{n}}(R(\theta)) - V_{\mathbf{n}}(0) - E(V_{\mathbf{n}}(R(\theta)) - V_{\mathbf{n}}(0))\|$ is $o_P(1)$, $V^*_{\mathbf{n}2} := \sup_{\|\theta\|\le M} \|V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(R(\theta))\|$, and $V^*_{\mathbf{n}3} := \sup_{\|\theta\|\le M} \|E(V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(R(\theta)))\|$. Using (4.10) and, for $\|\theta\| \le M$, applying Lemma 4.3 with $\tilde{\theta} = R(\theta)$ for $\mathbf{n}$ large enough,

$$V^*_{\mathbf{n}3} \le CH^{-1}_{\mathbf{n}} n \sup_{\|\theta\|\le M} E|\psi_p(Y^*_{\mathbf{ni}}(\theta)) - \psi_p(Y^*_{\mathbf{ni}}(R(\theta)))|K_{\mathbf{i}} \le C \sup_{\|\theta\|\le M} \|\theta - R(\theta)\| \le C\gamma. \tag{4.18}$$

Therefore, letting $\mathbf{n} \to \infty$ and $\gamma \to 0$, we have $V^*_{\mathbf{n}3} = o(1)$.

Set $B_{\mathbf{i}}(\theta) := I_{(|Y^*_{\mathbf{ni}}(\theta)| < C\gamma/H_{\mathbf{n}})} \|\mathcal{X}_{h\mathbf{i}}\| K_{\mathbf{i}}$. Noting that $|I_{(y<a)} - I_{(y<0)}| \le I_{(|y|\le|a|)}$, we obtain

$$V^*_{\mathbf{n}2} \le \sup_{\|\theta\|\le M} \|V_{\mathbf{n}}(\theta) - V_{\mathbf{n}}(R(\theta))\| \le C \sup_{\|\theta\|\le M} H^{-1}_{\mathbf{n}} \sum_{\substack{i_k=1\\k=1,\ldots,N}}^{n_k} B_{\mathbf{i}}(R(\theta)) \le B_{\mathbf{n}1} + B_{\mathbf{n}2},$$

where, by an argument similar to (4.18), $B_{n1} := C \sup_{\|\theta\|\le M} H^{-1}_{\mathbf{n}} \sum_{\substack{i_k=1\\k=1,\ldots,N}}^{n_k} EB_{\mathbf{i}}(R(\theta)) = o(1)$, and, similarly to (4.16), $B_{n2} := C \sup_{\|\theta\|\le M} |H^{-1}_{\mathbf{n}} \sum_{\substack{i_k=1\\k=1,\ldots,N}}^{n_k} (B_{\mathbf{i}}(R(\theta)) - EB_{\mathbf{i}}(R(\theta)))| = o_P(1)$. Thus, $V^*_{\mathbf{n}2} = o_P(1)$, and Lemma 4.4 follows from (4.17). □

Lemma 4.5 *Let* $\mathbf{D} = f_{\mathbf{X}}(\mathbf{x})\mathrm{diag}(1, \int \mathbf{uu}'K(\mathbf{u})d\mathbf{u})$. *Under Assumptions A1(iii) and A2,* $\sup_{\|\theta\|\le M} \|E(V_{\mathbf{n}}(\theta)$ $V_{\mathbf{n}}(0)) + f_{Y|\mathbf{X}}(q_p(\mathbf{x})|\mathbf{x})\mathbf{D}\theta\| = o(1)$.

Proof. The proof is similar to that in the time series case (cf. Lu, Hui, and Zhao 1998). □

Lemma 4.6 *Denote by* $\overline{\theta}_{\mathbf{n}}$ *the minimizer defined in (4.11). Then,* $\|V_{\mathbf{n}}(\overline{\theta}_{\mathbf{n}})\| = o_P(H^{-1}_{\mathbf{n}})$.

Proof. The proof is similar as that of Lemma A.2 of Ruppert and Carroll (1980). □

Lemma 4.7 *Under Assumptions A1 and A2, if* $a \ge N$ *and* $h_{\mathbf{n}} \to 0$,

$$E\left[(\mathbf{c}'V_{\mathbf{n}}(0) - \mathbf{c}'EV_{\mathbf{n}}(0))^2\right] \to p(1-p)f_{\mathbf{X}}(\mathbf{x}) \int (c_0 + \mathbf{c}'_1\mathbf{u})^2 K^2(\mathbf{u})d\mathbf{u}$$

as $\mathbf{n} \to \infty$, *where* $\mathbf{c} = (c_0, \mathbf{c}'_1)' \in \mathbb{R}^{1+d}$.

Proof. Set $v_{\mathbf{i}} := \psi_p(Y_{\mathbf{i}}^*)(c_0 + c_1'\mathbf{X}_{h\mathbf{i}})K_{\mathbf{i}}$. Then, Lemma 4.1 with $c_{\mathbf{n}k} = h_{\mathbf{n}}^{-d/a}$ for $k = 1, \cdots, N$ leads to

$$E\left[(c'V_{\mathbf{n}}(0) - c'EV_{\mathbf{n}}(0))^2\right] = (\widehat{\mathbf{n}}h_{\mathbf{n}}^d)^{-1}\left[\sum_{\substack{i_k=1\\k=1,\ldots,N}}^{n_k} \text{var}(v_{\mathbf{i}}) + \sum_{\{\mathbf{i},\mathbf{j}\in\mathcal{I}_{\mathbf{n}}|\exists k\,:\,i_k\neq j_k\}}\sum \text{cov}(v_{\mathbf{i}}, v_{\mathbf{j}})\right]$$
$$= h_{\mathbf{n}}^{-d}\text{var}(v_{\mathbf{1}}) + O(1)h_{\mathbf{n}}^{(1-N/a)d} + O(1)\sum_{k=1}^{N} c_{\mathbf{n}k}^a \sum_{t=c_{\mathbf{n}k}}^{\infty} t^{N-1}\varphi(t)$$
$$=: v_{n1} + v_{n2} + v_{n3}, \quad \text{say.}$$

Theorem 3 of Devroye and Györfi (1984, page 8) entails

$$E\left[I_{(Y_1^*<0)}(c_0 + \mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})^2 K_{\mathbf{1}}^2\right] = E\left[F_{Y|\mathbf{X}}(q_p(\mathbf{x}) + \dot{q}_p(\mathbf{X_1} - \mathbf{x})|\mathbf{X_1})(c_0 + \mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})^2 K_{\mathbf{1}}^2\right]$$
$$\to pf_X(\mathbf{x})\int (c_0 + \mathbf{c}_1'\mathbf{u})^2 K^2(\mathbf{u})d\mathbf{u}$$

and $E\left[I_{(Y_1^*<0)}(c_0 + \mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})K_{\mathbf{1}}\right] \to pf_X(\mathbf{x})\int (c_0 + \mathbf{c}_1'\mathbf{u})K(\mathbf{u})d\mathbf{u}$. This in turn implies

$$h_{\mathbf{n}}^{-d}E\left[v_{\mathbf{1}}^2\right] = E\left[\left(p^2 - 2pI_{(Y_{\mathbf{1}}^*<0)} + I_{(Y_{\mathbf{1}}^*<0)}\right)(c_0 + \mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})^2 K_{\mathbf{1}}^2\right]$$
$$\to p(1-p)f_X(\mathbf{x})\int (c_0 + \mathbf{c}_1'\mathbf{u})^2 K^2(\mathbf{u})d\mathbf{u},$$

and

$$h_{\mathbf{n}}^{-d}E[v_{\mathbf{1}}] = E\left[\left(p - I_{(Y_{\mathbf{1}}^*<0)}\right)(c_0 + \mathbf{c}_1'\mathbf{X}_{h\mathbf{1}})K_{\mathbf{1}}\right] \to (p-p)f_X(\mathbf{x})\int (c_0 + \mathbf{c}_1'\mathbf{u})K(\mathbf{u})d\mathbf{u} = 0.$$

Hence, $v_{n1} = h_{\mathbf{n}}^{-d}E\left[v_1^2\right] - h_{\mathbf{n}}^{-d}(Ev_1)^2 \to p(1-p)f_X(\mathbf{x})\int (c_0 + \mathbf{c}_1'\mathbf{u})^2 K^2(\mathbf{u})d\mathbf{u}$. On the other hand, it clearly follows, from the fact that $h_{\mathbf{n}} \to 0$ and Assumption (A3) with $a > N$, that $|v_{n2} + v_{n3}| = O(1)h_{\mathbf{n}}^{(1-N/a)d} + O(1)\sum_{k=1}^{N} c_{\mathbf{n}k}^a \sum_{t=c_{\mathbf{n}k}}^{\infty} t^{N-1}\varphi(t) \to 0$. The desired result follows.□

Proof of Theorem 2.1. As already mentioned, it is sufficient to check that the conditions of Lemma 4.2 are fulfilled. First we note that Lemmas 4.4 and 4.5 lead to (ii) of Lemma 4.2. Also, it follows from Lemma 4.6 together with Assumptions A2 and A3 that $\|V_{\mathbf{n}}(\overline{\theta}_{\mathbf{n}})\| = o_P(1)$. Take $A_{\mathbf{n}} = V_{\mathbf{n}}(\mathbf{0})$. Then it is clear from Lemma 4.7 that $A_{\mathbf{n}} = O_P(1)$. Since $\psi_p(y)$ is an increasing function of y, the function

$$\lambda \mapsto -\theta' V_{\mathbf{n}}(\lambda\theta) = H_{\mathbf{n}}^{-1}\sum_{\substack{i_k=1\\k=1,\ldots,N}}^{n_k} \psi_p(y_{\mathbf{i}}^* - \lambda\theta'\mathcal{X}_{h\mathbf{i}}/H_{\mathbf{n}})(-\theta'\mathcal{X}_{h\mathbf{i}})K_{\mathbf{i}}$$

is increasing with λ. Therefore, condition (i) of Lemma 4.2 holds. The theorem follows. □

4.3 Proof of asymptotic normality

On the basis of the Bahadur representation of Theorem 2.1, the asymptotic normality of our estimators in Theorems 2.2–2.6 follows exactly as in the corresponding proofs for mean regression in Hallin *et al.* (2004b), with the cross-term lemma, Lemma 4.1 replacing the corresponding

Lemma 5.2 in that paper, yielding the asymptotic normality with the bias (i.e. expectation) of the first term on the right-hand side of (2.5) as

$$E\left\{\eta_p(\mathbf{x})\frac{1}{\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}}\sum_{\substack{i_k=1\\k=1,\ldots,N}}^{n_k}\psi_p(Y_{\mathbf{i}}^*)\left[\begin{array}{c}1\\ \frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\end{array}\right]K\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\right)\right\}$$

$$=\eta_p(\mathbf{x})\frac{1}{\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}}\widehat{\mathbf{n}}E\left\{\psi_p(Y_{\mathbf{i}}^*)\left[\begin{array}{c}1\\ \frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\end{array}\right]K\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\right)\right\}$$

$$=\eta_p(\mathbf{x})\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}\;h_{\mathbf{n}}^{-d}$$
$$\times E\left\{\left(F_{Y|\mathbf{X}}\left(q_p(\mathbf{X_i})|\mathbf{X_i}\right)-F_{Y|\mathbf{X}}\left(q_p(\mathbf{x})+(\dot{q}_p(\mathbf{x}))'(\mathbf{X_i}-\mathbf{x})|\mathbf{X_i}\right)\right)\left[\begin{array}{c}1\\ \frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\end{array}\right]K\left(\frac{\mathbf{X_i}-\mathbf{x}}{h_{\mathbf{n}}}\right)\right\}$$

$$=\sqrt{\widehat{\mathbf{n}}h_{\mathbf{n}}^d}\left[(1+o(1))\frac{1}{2}\left(\begin{array}{c}B_0(\mathbf{x})\\ \mathbf{B}_1(\mathbf{x})\end{array}\right)h_{\mathbf{n}}^2\right],$$

where the last equality is derived via a first order Taylor expansion of $y\mapsto F_{Y|\mathbf{X}}(y|\cdot)$ and a second order Taylor expansion of $\mathbf{x}\mapsto q_p(\mathbf{x})$ (these expansions exist in view of Assumptions A1(ii) and A2). The $(1+o(1))$ factor is got rid of in Theorems 2.2–2.6 by using Assumption B2. Details are omitted. □

References

[1] Lu, Z., Hui, Y.V., and Zhao, Q. (1998). Local linear quantile regression under dependence: Bahadur representation and application. Discussion paper, Department of Management Sciences, City University of Hong Kong.

[2] Anselin, L. and Florax, R.J.G.M. (1995). *New Directions in Spatial Econometrics*. Berlin: Springer.

[3] Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B* **36**, 192-225.

[4] Bolthausen, E. (1982). On the central limit theorem for stationary random fields. *Annals of Probability* **10**, 1047-1050.

[5] Cai, Z.W. (2002). Regression quantiles for time series, *Econometric Theory* **18**, 169-192.

[6] Carbon, M., Hallin, M., and Tran, L.T. (1996). Kernel density estimation for random fields: the L_1 theory, *Journal of Nonparametric Statistics* **6**, 157-170.

[7] Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829-836.

[8] Cressie, N.A.C. (1993). *Statistics for Spatial Data*, revised ed. New York: John Wiley.

[9] Devroye, L., and Györfi, L. (1985). *Nonparametric Density Estimation: The L1 View.* New York: John Wiley.

[10] Efron, B. (1991). Regression percentiles using asymmetric squared error loss, *Statistica Sinica* **1**, 93-125.

[11] Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* **87**, 998-1004.

[12] Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications.* London: Chapman and Hall.

[13] Fan, J., Hu, T. C., and Truong, Y. K. (1994), Robust nonparametric function estimation. *Scandinavian Journal of Statistics* **21**, 433-446.

[14] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: nonparametric and parametric methods.* New York: Springer-Verlag.

[15] Granger, C.W.J., White, H., and Kamstra, M. (1989). Interval forecasting: an analysis based upon ARCH-quantile estimators. *Journal of Econometrics* **40**, 87-96.

[16] Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblable conditionelle: étude asymptotique et application au cas Markovien. In J.-J. Droesbeke et al., Eds, *Spatial Processes and Spatial Time Series Analysis*, *Proceedings of the 6th Franco-Belgian Meeting of Statisticians.* Brussels: FUSL.

[17] Guyon, X. (1995). *Random Fields on a Network.* New York: Springer Verlag.

[18] Hallin, M., Lu, Z., and Tran, L. T. (2001). Density estimation for spatial linear processes. *Bernoulli* **7**, 657-668.

[19] Hallin, M., Lu, Z., and Tran, L. T. (2004a). Density estimation for spatial processes: the L_1 theory. *Journal of Multivariate Analysis* **88**, 61-75.

[20] Hallin, M., Lu, Z., and Tran, L. T. (2004b). Local linear spatial regression. *Annals of Statistics* **32**, 2469-2500. .

[21] Honda T. (2000), Nonparametric estimation of a conditional quantile for α-mixing processes, *Annals of the Institute of Statistical Mathematics* **52**, 459-470.

[22] Koenker, R. (2000). Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics. *Journal of Econometrics* **95**, 347-374.

[23] Koenker, R. (2005). *Quantile Regression* Cambridge University Press, Cambridge, U.K.

[24] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

[25] Koenker, R. and Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50**, 43-61.

[26] Koenker, R. and Mizera, I. (2004). Penalized triograms: total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society Series B* **66**, 145-164.

[27] Koenker, R. and Portnoy, S. (1987). L-estimation for linear models. *Journal of the American Statistical Association* **82**, 851-857.

[28] Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* **12**, 793-813.

[29] Koul, H. L. and Mukherjee, K. (1994). Regression quantiles and related processes under long range dependence, *Journal of Multivariate Analysis* **51**, 318-337.

[30] Loader, C. (1999). *Local Regression and Likelihood*, Springer, New York.

[31] Lu, Z. and Chen, X. (2002). Spatial nonparametric regression estimation: non-isotropic case. *ACta Mathematicae Applicatae Sinica*, English Series **18**, 641-656.

[32] Lu, Z. and Chen, X. (2004). Spatial kernel regression estimation: weak consistency. *Statistics and Probability Letters* **68**, 125-136

[33] McBratney, A. B. and Webster, R. (1981). Detection of ridge and furrow pattern by spectral analysis of crop yield. *Intern. Statist. Rev.* **49**, 45-52.

[34] Mercer, W. B. and Hall, A. D. (1911). The experimental error field trials. *J. Agricultural Science* **4**, 107-132.

[35] Nakhapetyan, B.S. (1980). The central limit theorem for random fields with mixing conditions. In R. L. Dobrushin and Ya. G. Sinai, Eds., *Advances in Probability* **6**, *Multicomponent Systems*, 531-548. New York: Marcel Dekker.

[36] Neaderhouser, C. C. (1980). Convergence of blocks spins defined on random fields. *Journal of Statistical Physics* **22**, 673-684.

[37] Portnoy, S. L. (1991). Asymptotic behavior of regression quantiles in non-stationary dependent cases. *Journal of Multivariate Analysis* **38**, 100-113.

[38] Possolo, A. (1991). *Spatial Statistics and Imaging.* IMS Lecture Notes-Monograph Series. Hayward: Institute of Mathematical Statistics.

[39] Ripley, B. (1981). *Spatial Statistics*. New York: Wiley.

[40] Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Boston: Birkhauser.

[41] Ruppert, D. and Carroll, R. J. (1980). Trimmed least square estimation in the linear model. *Journal of the American Statistical Association* **75**, 828-838.

[42] Ruppert D. and Wand, M P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics* **22**, 1346-1370.

[43] Stone, C.J. (1977). Consistent nonparametric regression. *Annals of Statistics* **5**, 595-620.

[44] Takahata, H. (1983). On the rates in the central limit theorem for weakly dependent random fields. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64**, 445-456.

[45] Taylor, J.W., and Bunn, D.W. (1999). A quantile regression approach to generating prediction intervals. *Management Science* **45**, 225-237.

[46] Tran, L. T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis* **34**, 37-53.

[47] Tran, L. T. and Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *Journal of Multivariate Analysis* **44**, 23-46.

[48] Welsh, A. H. (1996). Robust estimation of smooth regression and spread functions and their derivatives. *Statistica Sinica* **6**, 347-366.

[49] Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434-449.

[50] Whittle, P. (1963). Stochastic process in several dimensions. *Bulletin of the International Statistical Institute* **40**, 974-985.

[51] Yao, Q. (2003). Exponential inequalities for spatial processes and uniform convergence rates for density estimation. In *Development of Modern Statistics and Related Topics—In Celebration of Professor Yaoting Zhang's 70th Birthday*, H. Zhang and J. Huang (edit.), 118-128. World Scientific, Singapore.

[52] Yu, K. and Jones, M.C. (1997). A comparison of local constant and local linear regression quantile estimation. *Computational Statistics & Data Analysis* **25**, 159-166.

[53] Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *Journal of the American Statistical Association* **93**, 228-237.

[54] Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics* **31**, 333 - 346.

[55] Yu, K., Lu, Z., and Standar, J. (2003). Quantile regression: applications and current research areas. *Journal of Royal Statistical Society Series D* **52**, 331-350.

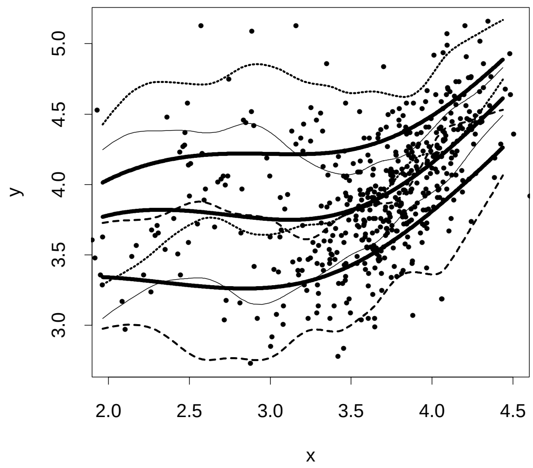
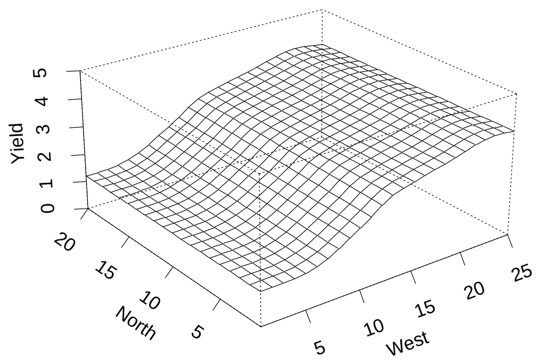


Figure 2: **Wheat-yield data.** *Three estimated conditional quantile regression functions (solid lines) for p =10%, 50% and 90%, of wheat yield based on Besag's model and their asymptotic 90% confidence intervals (thin solid lines for 50%, dashed lines for 10%, and dotted lines for 90% conditional quantile).*

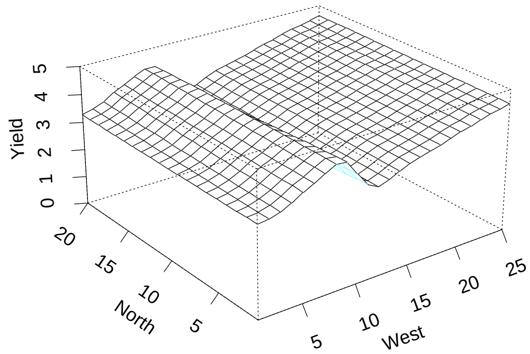
(a)

The estimated 10% quantile surface



(b)

The estimated median surface



(c)

The estimated 90% quantile surface

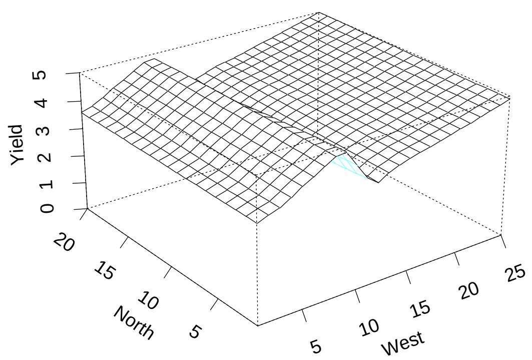


Figure 3: **Wheat-yield data.** *Three-dimensional perspective of the smoothed surfaces based on quantile regression conditional on row and column averaged neighbours, for (a)* $p = 10\%$*; (b)* $p = 50\%$*; (c)* $p = 90\%$.

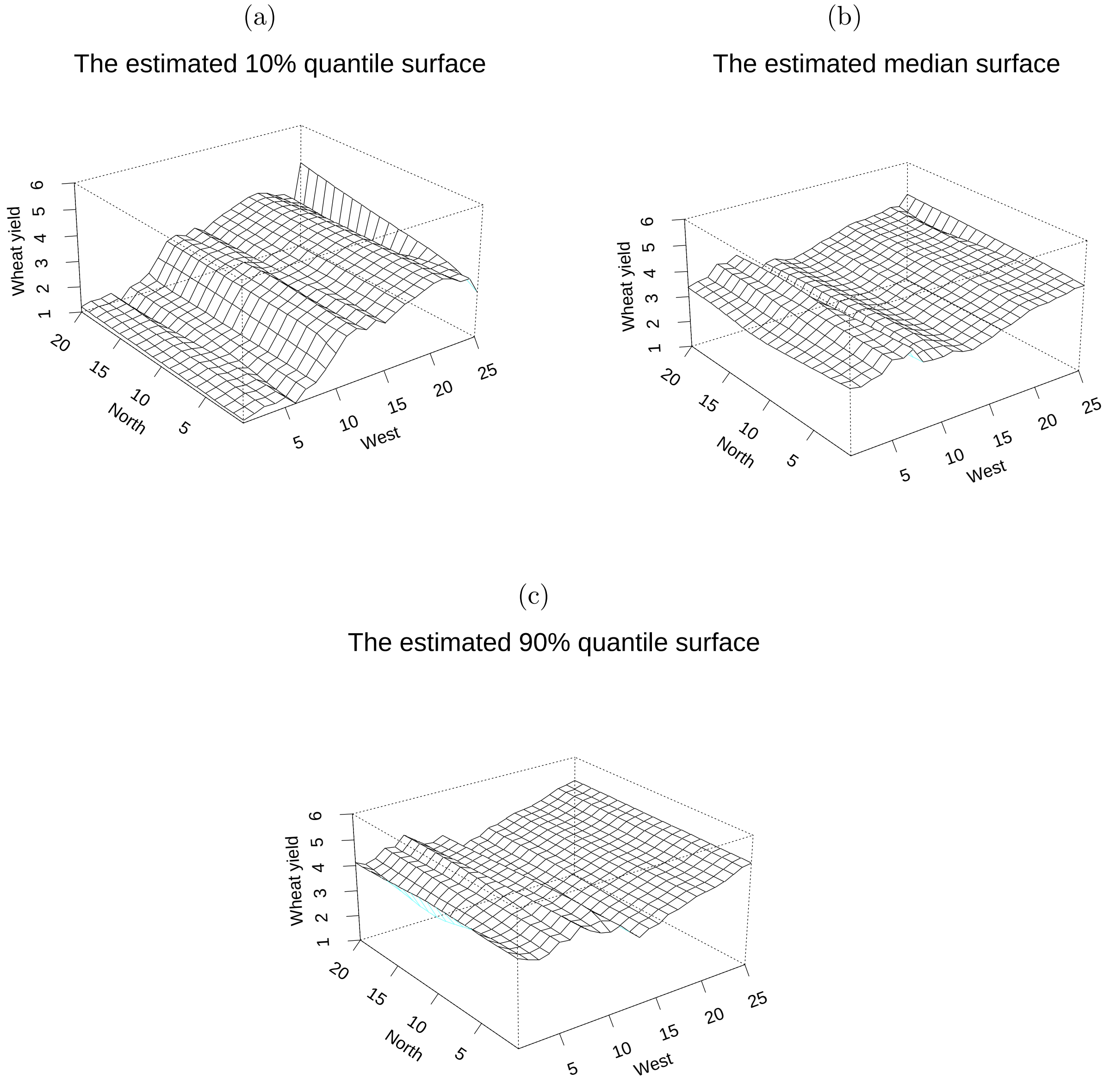


Figure 4: **Wheat-yield data.** *Three-dimensional perspective of the smoothed surfaces based on quantile regression conditional on four-dimensional neighbour sites, for (a)* $p = 0.10$*; (b)* $p = 0.50$*; (c)* $p = 0.90$.