INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus
Mathématique

Mètolidji Moquilas Raymond Affossogbe, Guy Martial Nkiet
and Carlos Ogouyandjou

Dimension reduction in spatial regression with kernel SAVE method

Volume 359, issue 4 (2021), p. 475-479

Published online: 17 June 2021

https://doi.org/10.5802/crmath.187

Les Comptes Rendus. Mathématique *sont membres du*
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org
e-ISSN : 1778-3569

Comptes Rendus
Mathématique
2021, 359, nº 4, p. 475-479
https://doi.org/10.5802/crmath.187

Statistics / *Statistiques*

Dimension reduction in spatial regression with kernel SAVE method

Réduction de la dimension en régression spatiale avec la méthode SAVE à noyau

Mètolidji Moquilas Raymond Affossogbe[a], **Guy Martial Nkiet**[*, b] **and Carlos Ogouyandjou**[a]

[a] Institut de Mathématiques et de Sciences Physiques,Porto Novo, Bénin

[b] Université des Sciences et Techniques de Masuku, Franceville, Gabon

E-mails: metolodji.affossogbe@imsp-uac.org, guymartial.nkiet@univ-masuku.org, ogouyandjou@imsp-uac.org

Abstract. We consider the smoothed version of sliced average variance estimation (SAVE) dimension reduction method for dealing with spatially dependent data that are observations of a strongly mixing random field. We propose kernel estimators for the interest matrix and the effective dimension reduction (EDR) space, and show their consistency.

Résumé. Nous considérons la version lisse de la méthode SAVE pour prendre en compte des observations spatialement dépendantes émanant d'un champ aléatoire fortement mélangeant. Nous proposons des estimateurs à noyau pour la matrice d'intérêt et l'espace de rédution de la dimension, et montrons leur convergence.

Manuscript received 22nd September 2019, accepted 7th February 2021.

Version française abrégée

Soit le modèle semi-paramétrique de régression (1) ; on cherche à estimer l'espace de réduction de la dimension, obtenu par la méthode SAVE, sur la base d'observations sur le domaine $\mathscr{I}_{\mathbf{n}} = \{\mathbf{i} = (i_1, i_2, \dots, i_L) \in \mathbb{Z}^L, 1 \le i_k \le n_k, k = 1, 2, \dots, L\}$ d'un champ aléatoire $\{W_{\mathbf{i}}, \mathbf{i} \in (\mathbb{N}^*)^L\}$, où $W_{\mathbf{i}} = (Z_{\mathbf{i}}, Y_{\mathbf{i}})$ a la même loi que (Z, Y), avec $Z = \Sigma^{-1/2}(X - \mathbb{E}(X))$, la matrice Σ étant la matrice de covariance de X supposée inversible. La méthode SAVE étant basée sur l'analyse spectrale de la matrice Γ définie en (2), l'estimation recherchée est obtenue à partir de l'analyse spectrale de l'estimateur à noyau $\widehat{\Gamma}_{\mathbf{n}}$ de Γ défini en (3). Posant $\widehat{\mathbf{n}} = n_1 \times n_2 \cdots \times n_L$ et $\mathbf{n} = (n_1, \dots, n_L) \in (\mathbb{N}^*)^L$, on écrit $\mathbf{n} \to +\infty$ lorsque $\min\{n_i, i = 1, 2, \dots, L\} \to +\infty$. Soit une suite $(b_{\mathbf{n}})$ de réels strictement positifs

* Corresponding author.

tendant vers 0 lorsque $\mathbf{n} \to +\infty$, constituant la fenêtre de l'estimateur à noyau précédent, on pose $\phi_{\mathbf{n}} = b_{\mathbf{n}}^k + \frac{1}{b_{\mathbf{n}}}\sqrt{\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}$, et on obtient :

Théorème 8. *Sous les hypothèses 2–6 avec* $\chi(t) = O(t^{-\theta})$, $t > 0$, $\theta > 2L$ *et* $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^3(\log\widehat{\mathbf{n}})^{-1} \to 0$, $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^{\theta_1}(\log\widehat{\mathbf{n}})^{-1} \to +\infty$ *où* $\theta_1 = \frac{4L+\theta}{\theta-2L}$, *on a :*

$$\widehat{\Gamma}_{\mathbf{n}} - \Gamma = O_p\left(\frac{1}{\sqrt{\widehat{\mathbf{n}}}}\right) + O_p\left(\frac{\phi_{\mathbf{n}}}{e_{\mathbf{n}}}\right) + O_p\left(\frac{\phi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right) + O_p\left(\frac{\phi_{\mathbf{n}}^3}{e_{\mathbf{n}}^3}\right) + O_p\left(\frac{\phi_{\mathbf{n}}^4}{e_{\mathbf{n}}^4}\right) + O_p\left(b_{\mathbf{n}}^k + \frac{\phi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right).$$

Corollaire 9. *Sous les hypothèses 2–7 avec* $\chi(t) = O(t^{-\theta})$, $t > 0$, $\theta > 2L$ *et* $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^3(\log\widehat{\mathbf{n}})^{-1} \to 0$, $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^{\theta_1}(\log\widehat{\mathbf{n}})^{-1} \to +\infty$ *où* $\theta_1 = \frac{4L+\theta}{\theta-2L}$, *on a* $\widehat{\Gamma}_{\mathbf{n}} - \Gamma = O_p\left(\widehat{\mathbf{n}}^{-1/2}\right)$.

Supposons que $\tau_1, \tau_2, \dots, \tau_N$ sont des vecteurs propres orthonormaux de Γ associés aux valeurs propres $\lambda_1, \dots, \lambda_N$ respectivement, telles que $\lambda_1 > \lambda_2 > \cdots > \lambda_N > 0$. Soient $\widehat{\tau}_1, \widehat{\tau}_2, \dots, \widehat{\tau}_N$ des vecteurs propres orthonormaux de $\widehat{\Gamma}_{\mathbf{n}}$ associés aux valeurs propres $\widehat{\lambda}_1, \dots, \widehat{\lambda}_N$ respectivement, telles que $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_N > 0$. Pour tout $j \in \{1, \dots, N\}$, on a $\beta_j = \Sigma^{-1/2}\tau_j$ et on pose $\widehat{\beta}_j = \widehat{\Sigma}_{\mathbf{n}}^{-1/2}\widehat{\tau}_j$, où $\widehat{\Sigma}_{\mathbf{n}} = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}}\left(X_{\mathbf{i}} - \overline{X}\right)\left(X_{\mathbf{i}} - \overline{X}\right)^T$ et $\overline{X} = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} X_{\mathbf{i}}$.

Corollaire 10. *Sous les hypothèses du Théorème 8, on a pour tout* $j \in \{1, \dots, N\}$, $\|\widehat{\beta}_j - \beta_j\| = o_p(1)$.

1. Introduction

Let us consider the semiparametric regression model introduced by Li [8] and defined as

$$Y = g(\beta_1^T X, \beta_2^T X, \dots, \beta_N^T X, \varepsilon), \tag{1}$$

where Y (resp. X) is a random variable with values in $\mathbb{R}$ (resp. $\mathbb{R}^d$, $d \geq 2$), N is an integer such that $N < d$, the parameters $\beta_1, \beta_2, \dots, \beta_N$ are d-dimensional linearly independent vectors, ε is a random variable that is independent of X, and g is an arbitrary unkown function. The estimation of the space spanned by the β_k's, called the effective dimension reduction (EDR) space, is a crucial issue for achieving reduction dimension. For this problem, Li [8] introduced the Sliced Inverse Regression (SIR) method whereas an alternative method, called sliced average variance estimation (SAVE), that is more comprehensive since it uses first and second moments was proposed in [4]. Smoothed versions of these methods, based on kernel estimators, have been proposed later in [13] and [14]. Recently, nonparametric statistical methods have evolved with the existence of spatially dependent data. So, kernel nonparametric estimation of the spatial regression function have been studied (cf. [1, 2, 6, 7, 10, 11]). For dimension reduction in spatial context, Loubes and Yao [9] investigated the kernel SIR method under strong mixing conditions. In this note, we study the case of kernel SAVE, which had never been done before. In Section 2, we introduce a kernel estimate of SAVE based on spatially dependent observations. Then, assumptions and consistency results are given in Section 3. An outline of the proofs of theorems is postponed in Section 4.

2. Kernel estimation of SAVE based on spatial data

In all of the paper, we assume that $\mathbb{E}\left(\|X\|^2\right) < +\infty$, where $\|\cdot\|$ is the usual Euclidean norm of $\mathbb{R}^d$, and that the covariance matrix Σ of X is invertible. Putting $Z = \Sigma^{-1/2}(X - \mathbb{E}(X))$ and denoting by $\mathrm{Cov}(Z|Y)$ the conditional covariance matrix of Z conditionally to Y, it is shown in [4] that the EDR space is fully obtained from the spectral analysis of the matrix

$$\Gamma := \mathbb{E}[(I_d - \mathrm{Cov}(Z|Y))^2] = \mathbf{I}_d - 2\mathbb{E}(C(Y)) + \mathbb{E}\left[C(Y)^2\right], \tag{2}$$

where $\mathbf{I}_d$ is the $d\times d$ identity matrix and $C(Y) := \mathrm{Cov}(Z|Y) = R(Y) - r(Y)r(Y)^T$, where $R(Y) = \mathbb{E}(ZZ^T|Y)$ and $r(Y) = \mathbb{E}(Z|Y)$. From the variance decomposition theorem, we have that $\mathbf{I}_d = \mathbb{E}(C(Y)) + \Psi$, where $\Psi = \mathrm{Cov}(\mathbb{E}(r(Y)) = \mathbb{E}[r(Y)r(Y)^T]$. Therefore, $\Gamma = -\mathbf{I}_d + 2\Psi + \Lambda$, where $\Lambda = \mathbb{E}[C(Y)^2]$, and the estimation of Γ boils down to that of the matrices Ψ and Λ. From now on, we assume that Y admits a density such that $f(y) > 0$ for all $y \in \mathbb{R}$. Let us consider a stationary random field $\{W_{\mathbf{i}},\ \mathbf{i} \in (\mathbb{N}^*)^L\}$ where $W_{\mathbf{i}} = (Z_{\mathbf{i}}, Y_{\mathbf{i}})$ has the same distribution than (Z,Y). We suppose that this process is observed on a region $\mathscr{I}_{\mathbf{n}} = \{\mathbf{i} = (i_1, i_2, \dots, i_L) \in \mathbb{Z}^L, 1 \le i_k \le n_k, k = 1,2,\dots,L\}$, where $\mathbf{n} = (n_1,\dots,n_L) \in (\mathbb{N}^*)^L$. We put $\widehat{\mathbf{n}} = n_1 \times n_2 \cdots \times n_L$ and write $\mathbf{n} \to +\infty$ if $\min\{n_i, i = 1,2,\dots,L\} \to +\infty$. For defining our estimators, we consider a sequence $(b_{\mathbf{n}})$ of strictly positive real numbers converging to zero as $\mathbf{n} \to +\infty$, and a kernel function K defined on $\mathbb{R}$. An estimator of f is then given by $\widehat{f}_{e_{\mathbf{n}}}(y) = \max\{e_{\mathbf{n}}, \widehat{f}_{\mathbf{n}}(y)\}$, where $(e_{\mathbf{n}})$ is a sequence of strictly positive real numbers such that $\lim_{\mathbf{n}\to+\infty} e_{\mathbf{n}} = 0$, and

$$\widehat{f}_{\mathbf{n}}(y) = \frac{1}{\widehat{\mathbf{n}} b_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} K\left(\frac{y - Y_{\mathbf{i}}}{b_{\mathbf{n}}}\right).$$

Then, we consider

$$\widehat{m}_{\mathbf{n}}(y) = \frac{1}{\widehat{\mathbf{n}} b_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} K\left(\frac{y - Y_{\mathbf{i}}}{b_{\mathbf{n}}}\right) Z_{\mathbf{i}}, \quad \widehat{M}_{\mathbf{n}}(y) = \frac{1}{\widehat{\mathbf{n}} b_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} K\left(\frac{y - Y_{\mathbf{i}}}{b_{\mathbf{n}}}\right) Z_{\mathbf{i}} Z_{\mathbf{i}}^T,$$

$$\widehat{r}_{\mathbf{n}}(y) = \frac{\widehat{m}_{\mathbf{n}}(y)}{\widehat{f}_{e_{\mathbf{n}}}(y)}, \quad \widehat{R}_{\mathbf{n}}(y) = \frac{\widehat{M}_{\mathbf{n}}(y)}{\widehat{f}_{e_{\mathbf{n}}}(y)},$$

and we take as estimator of Γ the random matrix

$$\widehat{\Gamma}_{\mathbf{n}} = -\mathbf{I}_d + 2\widehat{\Psi}_{\mathbf{n}} + \widehat{\Lambda}_{\mathbf{n}}, \tag{3}$$

where

$$\widehat{\Psi}_{\mathbf{n}} = \frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} \widehat{r}_{\mathbf{n}}(Y_{\mathbf{i}}) \widehat{r}_{\mathbf{n}}(Y_{\mathbf{i}})^T - \overline{Z}\,\overline{Z}^T,\ \widehat{\Lambda}_{\mathbf{n}} = \frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} \widehat{C}_{\mathbf{n}}(Y_{\mathbf{i}})^2$$

with $\overline{Z} = \frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} Z_{\mathbf{i}}$ and $\widehat{C}_{\mathbf{n}}(Y_{\mathbf{i}}) = \widehat{R}_{\mathbf{n}}(Y_{\mathbf{i}}) - \widehat{r}_{\mathbf{n}}(Y_{\mathbf{i}}) \widehat{r}_{\mathbf{n}}(Y_{\mathbf{i}})^T$.

3. Assumptions and asymptotic results

In order to establish the asymptotic results, the following assumptions will be considered.

Assumption 1. Γ is a positive-definite matrix.

Assumption 2. The kernel K is a density function with compact support, is of order k (where $k \ge 3$) and satisfies $\int |u|^k K(u)\,\mathrm{d}u = 1$ and $|K(x) - K(y)| \le C|x - y|$ for some $C > 0$.

Assumption 3. The functions f, r and R belong to $C^k(\mathbb{R})$ and $\sup_{y\in\mathbb{R}} |f^{(k)}(y)|$, $\sup_{y\in\mathbb{R}} \|m^{(k)}(y)\|$ and $\sup_{y\in\mathbb{R}} \|M^{(k)}(y)\|$ are bounded, where $m(y) = f(y)\,r(y)$ and $M(y) = f(y)\,R(y)$.

Assumption 4. $\sqrt{\widehat{\mathbf{n}}}\,\mathbb{E}[\|R(Y)\|^2 \mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}] = o(1)$, $\sqrt{\widehat{\mathbf{n}}}\,\mathbb{E}[\|r(Y)\|^4 \mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}] = o(1)$ and $\sqrt{\widehat{\mathbf{n}}}\,\mathbb{E}[\|R(Y)\| \times \|r(Y)\|^2 \mathbf{1}_{\{f(Y)\le e_{\mathbf{n}}\}}] = o(1)$.

Assumption 5. $\|Z\| \le D$, where D is a strictly positive constant.

Assumption 6. The process $\{W_{\mathbf{i}},\ \mathbf{i} \in (\mathbb{Z}^*)^L\}$ is strongly mixing, i.e. there exists a function χ from $\mathbb{R}_+$ to itself satisfying $\chi(t) \downarrow 0$ as $t \to +\infty$, such that for all subsets S and S' of $(\mathbb{Z}^*)^L$,

$$\alpha(\mathscr{B}(S), \mathscr{B}(S')) := \sup_{A\in\mathscr{B}(S),\, B\in\mathscr{B}(S')} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \le \chi(\delta(S, S'))$$

where $\mathscr{B}(S)$ (resp. $\mathscr{B}(S')$) denotes the Borel σ-fields generated by $\{W_{\mathbf{i}},\ \mathbf{i} \in S\}$ (resp. $\{W_{\mathbf{i}},\ \mathbf{i} \in S'\}$) and $\delta(S, S')$ denotes the Euclidean distance between S and S'.

Assumption 7. $b_{\mathbf{n}} \sim \widehat{\mathbf{n}}^{-c_1}$ and $e_{\mathbf{n}} \sim \widehat{\mathbf{n}}^{-c_2}$, where c_1 and c_2 are real numbers satisfying $c_1 > 0$, $0 < c_2 < \frac{2k-1}{4(2k+1)}$ and $\frac{c_2}{k} + \frac{1}{4k} < c_1 < \frac{1}{2} - 2c_2$.

Putting $\phi_{\mathbf{n}} = b_{\mathbf{n}}^k + \frac{1}{b_{\mathbf{n}}}\sqrt{\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}$, we have:

Theorem 8. *Under Assumptions 2–6, if* $\chi(t) = O(t^{-\theta})$, $t > 0$, $\theta > 2L$ *and* $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^3 (\log\widehat{\mathbf{n}})^{-1} \to 0$, $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^{\theta_1} (\log\widehat{\mathbf{n}})^{-1} \to +\infty$ *with* $\theta_1 = \frac{4L+\theta}{\theta-2L}$, *then we have:*

$$\widehat{\Gamma}_{\mathbf{n}} - \Gamma = O_p\left(\frac{1}{\sqrt{\widehat{\mathbf{n}}}}\right) + O_p\left(\frac{\phi_{\mathbf{n}}}{e_{\mathbf{n}}}\right) + O_p\left(\frac{\phi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right) + O_p\left(\frac{\phi_{\mathbf{n}}^3}{e_{\mathbf{n}}^3}\right) + O_p\left(\frac{\phi_{\mathbf{n}}^4}{e_{\mathbf{n}}^4}\right) + O_p\left(b_{\mathbf{n}}^k + \frac{\phi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right).$$

Corollary 9. *Under Assumptions 2–7, if* $\chi(t) = O(t^{-\theta})$, $t > 0$, $\theta > 2L$ *and* $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^3 (\log\widehat{\mathbf{n}})^{-1} \to 0$, $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^{\theta_1} (\log\widehat{\mathbf{n}})^{-1} \to +\infty$ *with* $\theta_1 = \frac{4L+\theta}{\theta-2L}$, *then we have* $\widehat{\Gamma}_{\mathbf{n}} - \Gamma = O_p\left(\widehat{\mathbf{n}}^{-1/2}\right)$.

For dealing with the $\widehat{\beta}_j$'s we assume that $\tau_1, \tau_2, \ldots, \tau_N$ are orthonormal eigenvectors of Γ associated with eigenvalues $\lambda_1, \ldots, \lambda_N$ respectively, such that $\lambda_1 > \lambda_2 > \cdots > \lambda_N > 0$. Let $\widehat{\tau}_1, \widehat{\tau}_2, \ldots, \widehat{\tau}_N$ be orthonormal eigenvectors of $\widehat{\Gamma}_{\mathbf{n}}$ associated with the eigenvalues $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N$ respectively, such that $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_N > 0$. For $j \in \{1, \ldots, N\}$, we have $\beta_j = \Sigma^{-1/2}\tau_j$ and we put $\widehat{\beta}_j = \widehat{\Sigma}_{\mathbf{n}}^{-1/2}\widehat{\tau}_j$, where $\widehat{\Sigma}_{\mathbf{n}} = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}}\left(X_{\mathbf{i}} - \overline{X}\right)\left(X_{\mathbf{i}} - \overline{X}\right)^T$ and $\overline{X} = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} X_{\mathbf{i}}$. Then, we have:

Corollary 10. *Under Assumptions 1–6, if* $\chi(t) = O(t^{-\theta})$, $t > 0$, $\theta > 2L$ *and* $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^3 (\log\widehat{\mathbf{n}})^{-1} \to 0$, $\widehat{\mathbf{n}}\, b_{\mathbf{n}}^{\theta_1} (\log\widehat{\mathbf{n}})^{-1} \to +\infty$ *with* $\theta_1 = \frac{4L+\theta}{\theta-2L}$, *then we have for any* $j \in \{1, \ldots, N\}$, $\|\widehat{\beta}_j - \beta_j\| = o_p(1)$.

4. Outline of proofs

We first have to prove that, under the assumptions of Theorem 8, we have $\sup_{y\in\mathbb{R}} \|\widehat{M}_{\mathbf{n}}(y) - M(y)\| = O_p\left(\phi_{\mathbf{n}}\right)$. For doing that, we consider the inequality $\sup_{y\in\mathbb{R}} \|\widehat{M}_{\mathbf{n}}(y) - M(y)\| \le \sup_{y\in\mathbb{R}} |\widehat{M}_{\mathbf{n}}(y) - \mathbb{E}[\widehat{M}_{\mathbf{n}}(y)]\| + \sup_{y\in\mathbb{R}} \|\mathbb{E}[\widehat{M}_{\mathbf{n}}(y)] - M(y)\|$. From the equality $\mathbb{E}[\widehat{M}_{\mathbf{n}}(y)] - M(y) = \int K(v)\left[M(y - vb_{\mathbf{n}}) - M(y)\right]\mathrm{d}v$ we obtain $\sup_{y\in\mathbb{R}} \|\mathbb{E}[\widehat{M}_{\mathbf{n}}(y)] - M(y)\| = O_p(b_{\mathbf{n}}^k)$ by using a Taylor expansion of M and the properties of K given in Assumption 2. For dealing with the other term, we consider a real $\varepsilon > 0$ and a sequence $(a_{\mathbf{n}})$ of non-negative real numbers converging to $+\infty$, and we obtain

$$\mathbb{P}\left(\sup_{y\in\mathbb{R}}\left\|\widehat{M}_{\mathbf{n}}(y) - \mathbb{E}\left[\widehat{M}_{\mathbf{n}}(y)\right]\right\| > \varepsilon; \|ZZ^T\| \le a_{\mathbf{n}}\right) \le \mathbb{P}(S_{\mathbf{n}} > \varepsilon),$$

where $S_{\mathbf{n}} = \sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} \Theta_{\mathbf{i},\mathbf{n}}$ with $\Theta_{\mathbf{i},\mathbf{n}} = \frac{C_2}{\widehat{\mathbf{n}} b_{\mathbf{n}}}\{\|Z_{\mathbf{i}} Z_{\mathbf{i}}^T\| + \mathbb{E}[\|Z_{\mathbf{i}} Z_{\mathbf{i}}^T\|]\}\mathbf{1}_{\{\|Z_{\mathbf{i}} Z_{\mathbf{i}}^T\| \le a_{\mathbf{n}}\}}$, C_2 being a constant that bound K. Then, using the spatial block decomposition of [12] together with Lemma 3.6 in [3], Markov inequality and Bernstein inequality, and taking $\varepsilon = \varepsilon_{\mathbf{n}} = \frac{1}{b_{\mathbf{n}}}\sqrt{\frac{\log\widehat{\mathbf{n}}}{\widehat{\mathbf{n}}}}$ and $a_{\mathbf{n}} = (\log\widehat{\mathbf{n}})^{1/4}$ we show that $\mathbb{P}(S_{\mathbf{n}} > \varepsilon) \to 0$ and $\mathbb{P}(\|ZZ^T\| > a_{\mathbf{n}}) \to 0$ as $\mathbf{n} \to +\infty$. Hence $\mathbb{P}\left(\sup_{y\in\mathbb{R}}\left\|\widehat{M}_{\mathbf{n}}(y) - \mathbb{E}\left[\widehat{M}_{\mathbf{n}}(y)\right]\right\| > \varepsilon_{\mathbf{n}}\right) \to 0$ as $\mathbf{n} \to +\infty$, that is $\sup_{y\in\mathbb{R}}\left\|\widehat{M}_{\mathbf{n}}(y) - \mathbb{E}\left[\widehat{M}_{\mathbf{n}}(y)\right]\right\| = O_p(\varepsilon_{\mathbf{n}})$. We have $\widehat{\Gamma}_{\mathbf{n}} - \Gamma = 2(\widehat{\Sigma}_{\mathbf{n}} - \Sigma) + (\widehat{\Lambda}_{\mathbf{n}} - \Lambda)$; it is shown in [9] that $\widehat{\Sigma}_{\mathbf{n}} - \Sigma = O_p\left(b_{\mathbf{n}}^k + \frac{\phi_{\mathbf{n}}^2}{e_{\mathbf{n}}^2}\right)$, then for proving Theorem 8 it remains to treat the second term. This is done from the decomposition $\widehat{\Lambda}_{\mathbf{n}} - \Lambda = A_{1\mathbf{n}} + A_{2\mathbf{n}} + A_{3\mathbf{n}}$, where $A_{1\mathbf{n}} = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} C(Y_{\mathbf{i}})^2 - \mathbb{E}\left[\mathrm{Cov}(Z|Y)^2\right]$, $A_{2\mathbf{n}} = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} C_{e,\mathbf{n}}(Y_{\mathbf{i}})^2 - \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} C(Y_{\mathbf{i}})^2$ and $A_{3\mathbf{n}} = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} \widehat{C}_{\mathbf{n}}(Y_{\mathbf{i}})^2 - \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_{\mathbf{n}}} C_{e,\mathbf{n}}(Y_{\mathbf{i}})^2$. Using Lemma 6.3 of [9], we show that $A_{1\mathbf{n}} = O_p(1/\widehat{\mathbf{n}}) = O_p\left(\frac{1}{\sqrt{\widehat{\mathbf{n}}}}\right)$. The second term is decomposed as $A_{2\mathbf{n}} = A_{21\mathbf{n}} - A_{22\mathbf{n}}$ with $A_{21\mathbf{n}} = A_{211\mathbf{n}} + A_{212\mathbf{n}}$, where

$$\mathbb{E}\left(\sqrt{\widehat{\mathbf{n}}}\|A_{211\mathbf{n}}\|\right) \le \sqrt{\widehat{\mathbf{n}}}\mathbb{E}\left(\|R(Y)\|^2\mathbf{1}_{\{f(Y)<e_{\mathbf{n}}\}}\right) + \sqrt{\widehat{\mathbf{n}}}\mathbb{E}\left(\|R(Y)\|\,\|r(Y)\|^2\mathbf{1}_{\{f(Y)<e_{\mathbf{n}}\}}\right)$$

and

$$\mathbb{E}\left(\sqrt{\widehat{\mathbf{n}}}\|A_{22\mathbf{n}}\|\right) \le \sqrt{\widehat{\mathbf{n}}}\mathbb{E}\left(\|R(Y)\|^2\mathbf{1}_{\{f(Y)<e_{\mathbf{n}}\}}\right) + 2\sqrt{\widehat{\mathbf{n}}}\mathbb{E}\left(\|R(Y)\|\,\|r(Y)\|^2\mathbf{1}_{\{f(Y)<e_{\mathbf{n}}\}}\right) + \sqrt{\widehat{\mathbf{n}}}\mathbb{E}\left(\|r(Y)\|^4\mathbf{1}_{\{f(Y)<e_{\mathbf{n}}\}}\right).$$

From Assumption 4 and Markov inequality we deduce that$A_{21\mathbf{n}} = O_p\left(1/\sqrt{\widehat{\mathbf{n}}}\right)$, $A_{22\mathbf{n}} = O_p\left(1/\sqrt{\widehat{\mathbf{n}}}\right)$, and consequently that $A_{2\mathbf{n}} = O_p\left(1/\sqrt{\widehat{\mathbf{n}}}\right)$. The third term has a decomposition $A_{3\mathbf{n}} = A_{31\mathbf{n}} - A_{32\mathbf{n}} + A_{33\mathbf{n}}$ with $A_{31\mathbf{n}} = A_{311\mathbf{n}} + A_{312\mathbf{n}} + A_{313\mathbf{n}} + A_{314\mathbf{n}}$, where $\|A_{311\mathbf{n}}\| \le \frac{2D^4}{e_\mathbf{n}}\|f-\widehat{f}_\mathbf{n}\|_\infty + \frac{2D^2}{e_\mathbf{n}}\|M-\widehat{M}_\mathbf{n}\|_\infty$, and

$$\|A_{312\mathbf{n}}\| \le 2D^4\frac{\|f-\widehat{f}_\mathbf{n}\|_\infty^2}{e_\mathbf{n}^2} + 2D^2\frac{\|m-\widehat{m}_\mathbf{n}\|_\infty^2}{e_\mathbf{n}^2} + 2D^3\frac{\|m-\widehat{m}_\mathbf{n}\|_\infty\,\|f-\widehat{f}_\mathbf{n}\|_\infty}{e_\mathbf{n}^2}.$$

Then from Lemma 6.5 in [9] and the above result $\|\widehat{M}_\mathbf{n} - M\|_\infty = O_p\left(\phi_\mathbf{n}\right)$, we obtain $A_{311\mathbf{n}} = O_p\left(\frac{\phi_\mathbf{n}}{e_\mathbf{n}}\right)$ and $A_{312\mathbf{n}} = O_p\left(\frac{\phi_\mathbf{n}^2}{e_\mathbf{n}^2}\right)$. Similar arguments lead to $A_{313\mathbf{n}} = O_p\left(\frac{\phi_\mathbf{n}}{e_\mathbf{n}}\right)$ and $A_{314\mathbf{n}} = O_p\left(\frac{\phi_\mathbf{n}}{e_\mathbf{n}}\right)$, and we conclude that $A_{31\mathbf{n}} = O_p(\frac{\phi_\mathbf{n}}{e_\mathbf{n}}) + O_p(\frac{\phi_\mathbf{n}^2}{e_\mathbf{n}^2})$. It can be noticed that $\|A_{33\mathbf{n}}\| = \|A_{31\mathbf{n}}\|$. Thus, $A_{33\mathbf{n}} = O_p(\frac{\phi_\mathbf{n}}{e_\mathbf{n}}) + O_p(\frac{\phi_\mathbf{n}^2}{e_\mathbf{n}^2})$. Similarly, we also obtain $A_{32\mathbf{n}} = O_p(\frac{\phi_\mathbf{n}}{e_\mathbf{n}}) + O_p(\frac{\phi_\mathbf{n}^2}{e_\mathbf{n}^2}) + O_p(\frac{\phi_\mathbf{n}^3}{e_\mathbf{n}^3}) + O_p(\frac{\phi_\mathbf{n}^4}{e_\mathbf{n}^4})$. We can then conclude that $A_{3\mathbf{n}} = O_p(\frac{\phi_\mathbf{n}}{e_\mathbf{n}}) + O_p(\frac{\phi_\mathbf{n}^2}{e_\mathbf{n}^2}) + O_p(\frac{\phi_\mathbf{n}^3}{e_\mathbf{n}^3}) + O_p(\frac{\phi_\mathbf{n}^4}{e_\mathbf{n}^4})$. Corollary 3.1 is obtained from the fact that $\frac{\widehat{\mathbf{n}}^{1/2}}{e_\mathbf{n}}\phi_\mathbf{n} \sim \widehat{\mathbf{n}}^{-1/2+c_2+c_1}\sqrt{\log\widehat{\mathbf{n}}}$. Since $-1/2+c_2+c_1<0$, we obtain that $\frac{\widehat{\mathbf{n}}^{1/2}}{e_\mathbf{n}}\phi_\mathbf{n} \to 0$ as $\mathbf{n}\to+\infty$. Then, $\frac{\phi_\mathbf{n}}{e_\mathbf{n}} = O_p(\widehat{\mathbf{n}}^{-1/2}) = o_p(1)$ and, consequently, $\frac{\phi_\mathbf{n}^\ell}{e_\mathbf{n}^\ell} = O_p\left(\frac{\phi_\mathbf{n}}{e_\mathbf{n}}\right) = O_p(\widehat{\mathbf{n}}^{-1/2})$ for $\ell = 2,3,4$. For proving Corollary 3.2, we write $\widehat{\Sigma}_\mathbf{n} - \Sigma = \frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_\mathbf{n}}\mathscr{V}_\mathbf{i} - \left(\overline{X}-\mathbb{E}(X)\right)\overline{X}^T - \mathbb{E}(X)\left(\overline{X}-\mathbb{E}(X)\right)^T$, where $\mathscr{V}_\mathbf{i} := X_\mathbf{i}X_\mathbf{i}^T - \mathbb{E}(X)\mathbb{E}(X)^T$. Then using 6.3 of [9], we obtain $\frac{1}{\widehat{\mathbf{n}}}\sum_{\mathbf{i}\in\mathscr{I}_\mathbf{n}}\mathscr{V}_\mathbf{i} = O_p(1/\widehat{\mathbf{n}}) = o_p(1)$ and $\overline{X}-\mathbb{E}(X) = O_p(1/\widehat{\mathbf{n}}) = o_p(1)$. Therefore, $\widehat{\Sigma}_\mathbf{n} - \Sigma = O_p(1/\widehat{\mathbf{n}}) = o_p(1)$ and $\|\widehat{\Sigma}_\mathbf{n}^{-1/2} - \Sigma^{-1/2}\| = o_p(1)$. Applying Lemma 1 of [5] and Theorem 8 permits to obtain t $\|\widehat{\tau}_j - \tau_j\| = o_p(1)$ for $j = 1,2,\ldots,N$. Finally, from $\|\widehat{\beta}_j - \beta_j\| \le \|\widehat{\Sigma}_\mathbf{n}^{-1/2} - \Sigma^{-1/2}\|\,\|\widehat{\tau}_j\| + \|\Sigma^{-1/2}\|\,\|\widehat{\tau}_j - \tau_j\|$ we conclude that $\|\widehat{\beta}_j - \beta_j\| = o_p(1)$.

References

[1] M. K. Attouch, A. Gheriballah, A. Laksaci, "Robust nonparametric estimation for functional spatial regression", in *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, Physica-Verlag, 2011, p. 27-31.

[2] M. Carbon, C. Francq, L. T. Tran, "Kernel regression estimation for random fields", *J. Stat. Plann. Inference* **137** (2007), no. 3, p. 778-798.

[3] M. Carbon, L. T. Tran, B. Wu, "Kernel density estimation for random fields", *Stat. Probab. Lett.* **36** (1997), no. 2, p. 115-125.

[4] R. D. Cook, S. Weisberg, "Discussion of a paper by K. C. Li", *J. Am. Stat. Assoc.* **86** (1991), p. 328-332.

[5] L. Ferré, A.-F. Yao, "Functional Sliced Inverse Regression analysis", *Statistics* **37** (2003), no. 6, p. 475-488.

[6] M. Hallin, Z. Lu, L. T. Tran, "Local linear spatial regression", *Ann. Stat.* **32** (2004), no. 6, p. 2469-2500.

[7] J. Li, L. T. Tran, "Nonparametric estimation of conditional expectation", *J. Stat. Plann. Inference* **139** (2009), no. 2, p. 164-175.

[8] K.-C. Li, "Sliced inverse regression for dimension reduction", *J. Am. Stat. Assoc.* **86** (1991), no. 414, p. 316-342.

[9] J. M. Loubes, A.-F. Yao, "Kernel Inverse Regression for spatial random fields", *Int. J. Appl. Math. Stat.* **32** (2013), p. 1-26.

[10] Z. Lu, X. Chen, "Spatial kernel regression estimation : weak consistency", *Stat. Probab. Lett.* **68** (2004), no. 2, p. 125-136.

[11] R. Menezes, P. García-Soidán, C. Ferreira, "Nonparametric spatial prediction under stochastic sampling design", *J. Nonparametric Stat.* **22** (2010), no. 3, p. 363-377.

[12] L. T. Tran, "Kernel density estimation on random fields", *J. Multivariate Anal.* **34** (1990), no. 1, p. 37-53.

[13] L.-X. Zhu, K.-T. Fang, "Asymptotics for kernel estimate of sliced inverse regression", *Ann. Stat.* **24** (1996), no. 3, p. 1053-1068.

[14] L. Zhu, L.-X. Zhu, "On kernel method for sliced average variance estimation", *J. Multivariate Anal.* **98** (2007), no. 5, p. 970-991.