

logo-univ.png

3*

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

THÈSE

En vue de l'obtention du

**Doctorat 3ème cycle de l'Université Djillali
Liabès de Sidi Bel Abbès (S.B.A.), Algérie**

Discipline : Mathématiques Appliquées

Spécialité : Statistique

présentée et soutenue par

Nafissa Messabihi

Juillet 2017

sous le titre :

Sur la régression relative et ses applications

Devant le jury composé de :

Toufik Benchikh	Univ. Djillali Liabès, S.B.A. (Algérie)	Président
Abdelkader Gheriballah	Univ. Djillali Liabès, S.B.A. (Algérie)	Rapporteur
Saadia Rahmani	Univ. Saida	Rapporteur
Ali Laksaci	Univ. Djillali Liabès, S.B.A. (Algérie)	Directeur de thèse
Mohamed Attouch	Univ. Grenoble Alpes, Grenoble (France)	Co-Directeur de thèse

Table des matières

1	Introduction	5
1.1	L'estimation par la méthode du noyau	5
1.2	L'estimation par la régression relative	8
1.3	La régression non paramétrique fonctionnelle : Etat de l'art . .	10
1.4	Résultats existants	14
1.5	Le cas multi-varié	14
1.6	Le cas fonctionnel	16
1.7	Contribution de la thèse	18
1.8	Bibliographie	20
2	Cas vectoriel	29
2.1	Introduction	29
2.2	Construction of the spatial estimate	32
2.3	Main results	34
2.3.1	Consistency	34
2.3.2	The asymptotic normality	36
2.4	Discussion and Application	37
2.4.1	Comparison to existing results	37

TABLE DES MATIÈRES

2.4.2	Application to prediction	38
2.4.3	Conditional Confidence Interval	39
2.5	On the finite-sample performance of the method	40
2.6	Application on real data	43
2.7	Proofs	48
3	Cas fonctionnel	63
3.1	Construction of the spatial estimate	63
3.2	Main results	64
3.2.1	Consistency	65
3.3	Références	86

Chapitre 1

Introduction

1.1 L'estimation par la méthode du noyau

Historiquement l'estimateur à noyau a été considéré séparément à la première fois par Nadaraya et Watson en 1964. Les deux auteurs ont adopté ces techniques d'estimation à la fonction de régression $r(x) = E(Y/X = x)$. Cette approche de la méthode du noyau peut être vu comme généralisation de la méthode de régressogramme proposée par Tukey en 1961. Cependant la méthode du noyau donne des estimateurs très lisses et faciles à utiliser dans la pratique. Ce qui a motivé les auteurs d'accorder plus d'importance à cette méthode d'estimation comparativement aux autres procédures existantes. Les propriétés asymptotique de la forme actuelle de l'estimateur à noyau de la fonction de régression se date du 1981. Ils ont été développés par Collomb (1981), dont il a démontré la convergence presque sûre. Bosq en 1987 a étudié la convergence en moyenne quadratique. Il a donné la forme explicite du biais et de la variance de l'estimateur à noyau de la fonction

de régression. Il est largement connu que les propriétés asymptotiques de l'estimateur à noyau de la fonction de régression sont étroitement liées aux conditions des régularités de la fonction de régression. Nous signalons que les résultats asymptotiques précédemment cités sont obtenus sous des conditions de dérivabilité de type, classe C^k pour le modèle de régression. Cependant Mammen en 1991 a établi des résultats similaires sous des conditions alternatives. L'étude de la convergence en norme L_p de l'estimateur à noyau de la fonction de régression a été obtenu par Devroye en 1987 . La plus part des résultats sur cités ont été démontré pour le cas des observations indépendantes identiquement distribuées. Motivé par son importance en application en prévision l'étude des observations dépendantes est l'un des sujets privilégiés pour l'analyse de série chronologique. Les premiers résultats conséquents sur ce problème ont été obtenus par Collomb et al. en 1986. Ces auteurs ont abordé le cas où les observations sont ϕ -mélange. On trouvera dans Gyorfi et al. (1989) une collection très riche des propriétés asymptotiques couvrant plusieurs cas des observations dépendantes tels le ϕ -mélange, le α -mélange, le ρ -mélange ainsi que le cas ergodique. L'une des difficultés de l'implémentation de l'estimateur à noyau dans la pratique est la question du choix du paramètre de lissage. Cette question technique a été aussi largement considérée. On trouvera dans la littérature statistique une variété très large des méthodes pratiques permettant le choix optimal de ce dernier. Nous citons la méthode de validations croisées développée par Hardel et al. (1987), Jones et al. (1996), Vieu (1991) parmi d'autres. D'autres approches alternatives peuvent être cité, la première sur le rééchantillonnage voir Cao-Abad (1997), la deuxième basée sur la technique d'injection abordée par Herrmann (1997)

et une méthode moins connue qui se base sur les quantiles de régression introduite par Kozek et Shuster (1990). Pour une étude comparative des différentes méthodes utilisées pour le choix optimal du paramètre de lissage se fier à Loader (1990). Cette méthode d'estimation a été aussi adaptée pour d'autres modèles non paramétriques tels la fonction de répartition conditionnelle, la densité conditionnelle et la fonction du hasard conditionnelle. Il est évident que la fonction de répartition conditionnelle peut être vu comme une fonction de régression de la fonction indicatrice. En utilisant cette démarche Roussas en 1968 a construit un estimateur dans le cas où les observations sont de type Markovien. Depuis cette publication l'estimation par la méthode du noyau de la fonction de répartition conditionnelle a été largement développée en tant que modèle préliminaire pour l'estimation des quantiles conditionnels. D'autre part, la densité conditionnelle a été estimée par la méthode du noyau comme étant la dérivée de la fonction de répartition conditionnelle. L'initiateur de cette approche est aussi Roussas en 1969 toujours dans le cas des données markoviennes. En 1992 Youndji a étudié le problème du choix du paramètre de lissage pour l'estimation de la densité conditionnelle en considérant les deux cas(cas des observations indépendantes ainsi que le cas des observations fortement mélangeantes). Laksaci en collaboration avec Yousfat en 2002 ont obtenu la convergence en norme L_p de l'estimateur à noyau de la densité conditionnelle. D'autres auteurs se sont intéressés à l'estimation de la densité conditionnelle comme modèle préliminaire du mode conditionnel. On cite Collomb et al (1987), Vieu et Quentaladelrio (1997) , Louani et Ould -Said (1999), Berlinet et al. (1999). D'autre type des régressions ont été considéré et estimé par la méthode du noyau. A titre d'exemple la régres-

sion robuste par Bonet et Fermain en 1989, la régression local linéaire par Fan et al . 1996... . Dans ce dernier temps l'estimation de la régression non paramétrique a été élargi à d'autre type des données tels les observations censurées par Gaussoum et Ould-Said (2011) , le cas des données tronquées par Lemadani et Ould-said (2010) ou bien à d'autre type de corrélation tels les observations ergodiques par Laib et al (2009), les variables associées par Douschich (2013). Il est à noter que la plus part des travaux actuels s'intéressent au cas où les observations sont de nature fonctionnelle. Un paragraphe dédié a ce dernier cas est également présenté par la suite.

1.2 L'estimation par la régression relative

La question de prévision est l'un des problèmes le plus important en statistique non paramétrique. Cette question est interprétée comme une étude de la co-variabilité entre deux variables aléatoires. Plus précisément on suppose que le lien entre une variable explicative X et une autre variable réponse Y est exprimé par la relations suivantes

$$Y = r(x) + \text{bruit}.$$

Dans des conditions particulières, la fonction r est traitée comme étant l'espérance conditionnelle de Y sachant X . Typiquement cette considération est obtenue en minimisant le critère suivant

$$\min_r E[(Y - r)^2 | X = x]$$

Il est clair que ce critère ne tient pas en compte le poids de la variable réponse Y . Autrement dit, il traite toutes les variables réponses de la même manière.

Ceci est l'un de défauts de cette approche, notamment lorsque les données traitées contiennent des valeurs aberrantes. Afin de surmonter ce problème Jones (2008) a développé une autre approche basée sur le critère de l'erreur relative suivante

$$\min_r E\left[\frac{(Y - r)^2}{Y^2} | X = x\right].$$

Motivé par sa résistance aux présences des valeurs aberrantes, cette procédure d'estimation a été extensivement utilisée dans le cas paramétrique. Nous renvoyons à Narula et Wellington (1977), Shen et al (1985), Bernhard et Stahlecker (2003) et Yang et Ye (2013) pour ce type des modèles. Cependant très peu des littérateurs dans la statistique non paramétrique ont abordé cette question. Sans prétendre de l'exhaustivité, nous citons l'article de Jones et al. 2008 où il a traité la convergence en moyenne quadratique de deux estimateurs issus de la méthode du noyau. Le premier, il a été construit par la méthode du noyau classique tandis que le deuxième concerne la méthode locale linéaire. Ses résultats asymptotiques ont été obtenus dans le cas où les observations sont indépendantes identiquement distribuées. Très récemment Mechab et Laksaci (2016) ont établi la convergence presque complète et la normalité asymptotique de l'estimateur à noyau de la régression relative. Dans notre contribution (Attouch et al. (2016)), nous avons considéré le cas des données spatiales, dont nous avons obtenu la convergence presque complète et la normalité asymptotique. Nous renvoyons à Demengeot et al. Pour le cas des données fonctionnelles.

1.3 La régression non paramétrique fonctionnelle : Etat de l'art

La statistique fonctionnelle est une nouvelle branche de la statistique mathématiques qui traite les variables aléatoires observées d'une manière continue dans l'espace et /ou dans le temps. Ce phénomène est actuellement usuel grâce au développement technologique. La révolution numérique sur les outils de mesure et les appareils d'enregistrement a permis la surveillance permanentes de plusieurs objets tels les indices boursiers, la pollution, les courbes de consommation , le microclimat, l'imagerie satellitaire. Ainsi, la modernisation des outils statistiques, avec cette révolution technologique, est devenue plus que nécessaire. A cet objectif, il est connu que les modèles non paramétriques sont plus adaptés que les modèles paramétriques pour le cas des données fonctionnelles. En effet, les modèles non paramétriques nécessitent une prédétermination de la forme modèle en utilisant une résolution graphique. Cependant cet outil n'est pas disponible en statistique fonctionnelle. L'analyse non paramétrique des données fonctionnelles se date des années 2000. Le premier article sur le sujet a été publié par Ferraty et Vieu (2000) dont ils ont démontré la convergence presque complète de la version fonctionnelle de l'estimateur de Nadarya Watson. Ils ont considéré pour cela des observations indépendantes identiquement distribuées. L'une des applications la plus importante de la modélisation des données fonctionnelles est la prévision sur les processus de trajectoire continue. Cet aspect a été abordé par Ferraty et al. (2002). Ils ont généralisé leur résultat du cas iid (2000) au cas des observations dépendantes. Une autre application importante de

l'analyse statistique des observations fonctionnelles est la classification des courbes. Cette question a été développé par Ferraty et Vieu (2003). En 2003 Dabo-Niang et Rhomari ont établi la convergence en norme L_p de l'estimateur à noyau de la régression non paramétrique. La normalité asymptotique de cet estimateur a été obtenue par Masry (2005). Elle a été obtenue sous des conditions de mélange fort. Ferraty en collaboration avec André Mas et Philipe Vieu en 2007 ont déterminé l'erreur quadratique asymptotiquement exacte de l'estimateur à noyau de l'opérateur de régression. Delsol en (2007, 2009) a obtenu le terme dominant de l'erreur de la norme L_p de la fonction de régression dans les deux cas (cas iid et cas de mélange). La convergence en moyenne quadratique est un résultat préliminaire très important pour la détermination du paramètre de lissage optimal. Cette question du choix du paramètre de lissage a été abordée par Rachdi et Vieu en 2007. Ils ont élaboré un critère de sélection de la fenêtre de lissage et ils ont montré son optimalité asymptotique dans le cas des observations indépendantes identiquement distribuée. Leur procédure du choix a été aussi adapté pour deux types de choix a savoir le choix local et le choix global. Dans l'article de Benhenni et al (2007) on trouvera une comparaison entre les deux critères. Une méthode d'estimation alternative à la méthode du noyau a été introduite par Burba et al (2008, 2009). Ces derniers ont établi sur une approche basée sur le nombre des voisins le plus proches. Ils ont démontré la convergence presque complète de l'estimateur construit. Le cas iid a été obtenu en 2008, tandis que le cas de mélange a été publié en 2009. Ferraty, Laksaci et al (2010) ont abordé la question de la convergence uniforme en statistique fonctionnelle. Ils ont montré que la vitesse de convergence uniforme

est différente de la vitesse de convergence ponctuellement contrairement au cas vectoriel. Ils ont constaté aussi que la vitesse de convergence est étroitement liée à la structure topologique de l'espace fonctionnel des données. La convergence uniforme de l'estimateur k -plus proches voisins de l'opérateur de régression a été étudié par Kudraszow et Vieu en 2013. Certains auteurs se sont intéressés à d'autre type de variables fonctionnelles telles les données ergodiques par Laib et Louani (2011), le cas des variables associées par Douge (2010) , les variables spatialement dépendantes par Dabo-Niang et al (2012). D'autres modèles de régression non paramétrique ont été considéré dans le littérature de la statistique fonctionnelle. On cite la régression d'indice simple par Ait Saidi et al. (2008a, 2008b). La convergence presque complète est traitée dans le premier article tandis que le deuxième article est consacré au choix optimal de l'indice. La M - régression a été introduite par Azzedine et al (2008). Ils ont démontré la convergence presque complète d'un estimateur construit par des techniques robustes. Ce résultat a été obtenu dans le cas des observations indépendantes identiquement distribuées. Dans le même contexte Attouch et al . (2009) considèrent le problème de la normalité asymptotique de cet estimateur. Carmbes et al (2008) ont étudié la convergence en norme Lp de l'estimateur de Azzedine et al. (2008) .Dans cette publication les auteurs considèrent les deux types de corrélation (cas iid ainsi que le cas mixing). Considérant cette dernière corrélation Attouch et al.(2010) ont montré la normalité asymptotique de l'estimateur à noyau de la M - régression. La convergence presque complète du même estimateur sous des conditions de mélange fort a été obtenue par les mêmes auteurs en 2012. La généralisation de ces résultats a été obtenue en 2012 par Attouch et al.

1.3. LA RÉGRESSION NON PARAMÉTRIQUE FONCTIONNELLE : ETAT DE L'ART13

(2012a, 2012b). Gheriballah et al. (2013) ont étudié l'estimation robuste de la fonction de régression pour le cas des données ergodiques. Ils ont montré la convergence presque complète avec précision de la vitesse de convergence de cet estimateur. Un autre modèle de régression a été largement considéré dans ce dernier temps en statistique fonctionnelle. Il s'agit de la régression locale linéaire. Les premiers résultats sur le sujet ont été établis par Baillo et al (2009). Ils ont construit un estimateur pour la fonction de régression en utilisant la méthode locale linéaire. Comme résultats asymptotique ils ont prouvé la convergence en moyenne quadratique de leur estimateur. Il est à noter que le fonctionnel modèle considéré est à co-variable hilberthienne. Une version simplifiée qui s'applique même pour le cas des régresseurs à valeur dans un espace de Banach a été construit par Barrientos et al (2010). Ils ont montré la convergence presque complète de cet estimateur dans le cas des observations indépendantes identiquement distribuées. Berlinet et al. (2011) utilisent une autre version basée sur l'inverse local de l'opérateur de covariance. La détermination des termes dominants de l'erreur quadratique de cet estimateur est la contribution principale de ces auteurs. Certains auteurs se sont intéressés au cas où la variable réponse est aussi fonctionnelle. Nous citons Dabo-Niang et Rhomari (2009) pour la convergence en norme Lp dans le cas iid. Nous renvoyons a Ferraty, Laksaci (2011) et pour la convergence presque complète. La normalité asymptotique a été considérée par Ferraty et al (2013) . Tous ces derniers résultats ont été obtenus dans le cas des observations indépendantes. A notre connaissance le seul résultat pour le cas dépendant a été obtenu par Ferraty, Laksaci et al (2012). A la fin de cet état de l'art sur l'estimation non paramétrique de la régression dans le cas

des données fonctionnelles nous citons quelques contributions déterminantes et d'intérêt général pour la statistique fonctionnelle , telle Ramsay et Silverman (2005), Bosq (2000), Ferraty et Vieu (2005), Horváth et Kokoszka (2012), Zhang (2013), Bongiorno et al. (2014) et Hsing et Eubank (2015). Le lecteur pourra se référer à ces contributions pour étude bibliographique plus approfondie sur la statistique fonctionnelle et ses applications.

1.4 Résultats existants

1.5 Le cas multi-varié

Théorème 1.5.1 *Sous les conditions suivantes :*

(H1) *K est symétrique*

(H2) *$r_l(x) = E[Y^{-l}|X = x]$ existe pour $l = 1, \dots, 4$ et $r_1''(x), r_2''(x), r_3''(x)$ et $r_4''(x)$ sont continues en $x \in [0, 1]$.*

On obtient :

(a) *Pour $h \leq x \leq 1 - h$,*

$$E\hat{g}_0(x)/X_1, \dots, X_n \asymp g(x) + \frac{1}{2}h^2 b_2(K)\left(\frac{r_1''}{r_2} - \frac{r_1 r_2''}{r_2^2} + 2\frac{f'g'}{f}\right)(x)$$

$$V\hat{g}_0(x)/X_1, \dots, X_n \asymp \frac{R(K)V_g(x)}{f(x)nh},$$

$$MSRE-\hat{g}_0(x)/X_1, \dots, X_n \asymp \frac{1}{4}h^4 b_2^2(K)r_2(x)\left(\frac{r_1''}{r_2} - \frac{r_1 r_2''}{r_2^2} + 2\frac{f'g'}{f}\right)^2(x) + \frac{r_2(x)R(K)V_g(x)}{f(x)nh}$$

(b) *Pour $h \leq x \leq 1 - h$,*

$$E\hat{g}_1(x)/X_1, \dots, X_n \asymp g(x) + \frac{1}{2}h^2 b_2(K)g''(x)$$

$$V\hat{g}_1(x)/X_1, \dots, X_n \asymp \frac{R(K)V_g(x)}{f(x)nh}$$

et

$$MSRE - \hat{g}_0(x)/X_1, \dots, X_n \asymp \frac{1}{4}h^4b_2^2(K)r_2(x)g''(x)^2 + \frac{r_2(x)R(K)V_g(x)}{f(x)nh}$$

$$V_g(x) = \frac{(r_2^3 - 2r_1r_2r_3 + r_1^2r_4(x))}{r_2^4(x)}$$

Ce résultat a été obtenu par Jones et al en 2008. Voir ce dernier pour plus des détails sur les conditions et sur la démonstration.

Théorème 1.5.2 *Sous les conditions suivantes :*

- (i) *Les fonctions $r_l(x) = E[Y^{-l}/X = .]$; ($l = 1, 2$) sont de classe C^2 dans un compact S .*
- (ii) *Les variables $(X_i, Y_i), i \in \mathbb{N}$ sont quasi-associées avec coefficient de covariance $\lambda_k, k \in \mathbb{N}$*

$$\exists a > 0, \exists C > 0, \text{tel que } \lambda_k \leq C e^{-ak}$$

avec

$$\lambda_k = \sup_{s \geq k} \sum_{|i-j| \geq s} \lambda_{i,j}$$

où

$$\lambda_{i,j} = \sum_{k=1}^d \sum_{l=1}^d |Cov(X_i^k, X_j^k)| + \sum_{k=1}^d |Cov(X_i^k, Y_j)| \sum_{l=1}^d |Cov(Y_i, X_j^l)| + |Cov(Y_i, Y_j)|$$

- (iii) *Il existe $\gamma \in (0, 1)$ et $\xi_1, \xi_2 > 0$ tel que*

$$\frac{1}{n^{2(1-\gamma/9-\xi_2)/(d+2)}} \leq h \leq \frac{C}{n^{(1+\xi_1)/(d+4)}}$$

On obtient ;

$$\sup |\tilde{r}(x) - r(x)| = O(h^2) + O_{a.co.}(\sqrt{\frac{\log n}{n^{1-\gamma} h^d}})$$

Ce résultat a été obtenu par Mechab et Laksaci en 2015 dans le cas où les variables sont quasi-associées.

1.6 Le cas fonctionnel

Théorème 1.6.1 *Sous les conditions suivantes :*

(H1) $\mathbb{P}(X \in B(x, s)) =: \phi_x(s) > 0$ pour tout $s > 0$ et $\lim_{s \rightarrow 0} \phi_x(s) = 0$.

(H2) Pour tout $(x_1, x_2) \in \mathfrak{N}_x^2$, on a :

$$|g_\gamma(x_1) - g_\gamma(x_2)| \leq C d^{k_\gamma}(x_1, x_2) \text{ pour } k_\gamma > 0$$

(H3) Les moments de la variable réponse vérifient :

$$\text{pour tout } m \geq 2, \mathbb{E}[Y^{-m}|X = x] < C < \infty$$

On obtient

$$|\tilde{r}(x) - r(x)| = O(h^{k_1}) + O(h^{k_2}) + O_{a.co.}(\sqrt{\frac{\log n}{n\phi_x(h)}})$$

Ce résultat a été obtenu par Demongeot et al (2016).

Théorème 1.6.2 *Sous les conditions suivantes :*

(H1) Pour tout $x \in S_{\mathcal{F}}$ et $s > 0$:

$$0 < C\phi_x(s) \leq \mathbb{P}(X \in B(x, s)) \leq C'\phi_x(s) < \infty$$

(H2) Il existe $\eta > 0$, tel que :

$$\text{pour tout } x, x' \in S_{\mathcal{F}}^{\eta}, |g_{\gamma}(x) - g_{\gamma}(x')| \leq C d^{k_{\gamma}}(x, x'),$$

$$\text{où } S_{\mathcal{F}}^{\eta} = \{x \in \mathcal{F} : \text{il existe } x' \in S_{\mathcal{F}} \text{ tel que } d(x, x') \leq \eta\}$$

On obtient

$$\sup_{x \in S_{\mathcal{F}}} |\tilde{r}(x) - r(x)| = O(h^{k_1}) + O(h^{k_2}) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\frac{\log n}{n})}{n\phi_x(h)}}\right)$$

Théorème 1.6.3 Sous les conditions suivantes :

(H1) Il existe une fonction $\chi_x(\cdot)$ tel que :

$$\text{pour tous } s \in [0, 1], \lim_{r \rightarrow 0} \frac{\phi_x(sr)}{\phi_x(r)} = \chi_x(s)$$

(H2) Pour $\gamma \in \{1, 2\}$, les fonctions $\Psi_{\gamma}(\cdot) = \mathbb{E}[g_{\gamma}(X) - g_{\gamma}(x)|d(x, X) = \cdot]$

sont dérivables en 0.

On a :

$$\mathbb{E}[\tilde{r}(x) - r(x)]^2 = B_n^2(x)h^2 + \frac{\sigma^2(x)}{n\phi_x(h)} + o(h) + o\left(\frac{1}{n\phi_x(h)}\right)$$

où

$$B_n(x) = \frac{(\Psi'_1(0) - r(x)\Psi'_2(0))\beta_0}{\beta_1 g_2(x)}$$

et

$$\sigma^2 = \frac{(g_2(x) - 2r(x)\mathbb{E}[Y^{-3}|X = x] + r^2(x)\mathbb{E}[Y^{-4}|X = x])\beta_2}{g_2^2(x)\beta_1^2}$$

avec

$$\beta_0 = K(1) - \int_0^1 (sK(s))' \chi_x(s) ds$$

et

$$\beta_j = K(1)^j - \int_0^1 (K^j)'(s) \chi_x(s) ds \text{ pour } j = 1, 2.$$

1.7 Contribution de la thèse

L'objectif de la thèse est la prévision par la régression non paramétrique en utilisant le critère de l'erreur relative. Nous avons contré sur le cas où les variables sont spatialement dépendantes. Il est intéressant de noter qu'avec le développement de GPS la modélisation des données spatiales joue un rôle important dans l'analyse des champs aléatoires. Précisément, on distingue trois types d'applications sur les données spatialement dépendantes : L'analyse géostatistique qui consiste à traiter les variables à trajectoire continue dans un espace de dimension supérieur à deux comme le cas des données localisées par les coordonnées géographiques . On peut trouver Ce type d'analyse dans des divers domaines tels l'environnement, l'économie, la finance , la circulation routière,.... Le deuxième type est les données spatiales et la modélisation des variables spatio-temporelles . Il s'agit du cas où les observations sont variées sur l'espace ainsi que sur le temps. De même on peut trouver plusieurs exemples de ce type dans des domaines variés. Le dernier cas et les données d'imagerie où les observations sont spatiales mais prennent des valeurs discrètes. Afin de couvrir toutes ces situations nous avons traité notre modèle dans deux cas différents.

Le premier est le cas où les observations sont de dimension finie , tandis que le deuxième cas traite les variables fonctionnelles. Dans le cas vectoriel nous avons construit un estimateur pour notre modèle. Nous avons étudié ses propriétés asymptotiques. Plus Précisément, nous avons obtenu la convergence presque complète et la normalité asymptotique. Ces résultats ont été obtenu sous des conditions standards qui permet d'exploiter la

corrélation spatiale des observations ainsi que l'aspect non paramétrique du modèle. Dans cette partie nous avons aussi appliqué notre méthode des données réelles.

Dans la deuxième partie nous avons généralisé les résultats de la première partie au cas des données fonctionnelles. Nous avons adopté la version fonctionnelle de notre estimateur et nous avons étudié ses propriétés asymptotiques. De même que le premier cas nous obtenons la vitesse de convergence pour le cas de la convergence presque complète et nous donnons l'expression de la variance asymptotique pour la normalité asymptotique. La dimensionnalité des observations est exprimée dans les deux vitesses de convergence à travers la fonction des petites boules que contrôle la propriété de concentration de la variable explicative.

L'essentiel de la thèse est présenté en quatre chapitres : Après ce chapitre introductif, dont on a présenté la bibliographie de la thématique traitée dans cette thèse. Le deuxième chapitre est consacré au cas vectoriel. Nous traitons le cas fonctionnel au troisième chapitre. Nous achèverons cette thèse par une conclusion dont on met en évidence la supériorité de la régression relative sur la régression classique dans le cas où les observations ne sont pas uniformes et contiennent des valeurs aberrantes.

1.8 Bibliographie

Ait-Saïdi, Ahmed ; Ferraty, Frédéric ; Kassa, Rabah ; Vieu, Philippe. Choix optimal du paramètre fonctionnel dans le modèle à indice fonctionnel simple. (French) [Optimal functional parameter in the single functional index model] C. R. Math. Acad. Sci. Paris 346 (2008), no. 3-4, 217–220.

Arnold, Bernhard F. ; Stahlecker, Peter. Relative squared error prediction in the generalized linear regression model. Statist. Papers 44 (2003), no. 1, 107–115.

Attouch, Mohammed Kadi ; Benchikh, Tawfik. Asymptotic distribution of robust k-nearest neighbour estimator for functional nonparametric models. Mat. Vesnik 64 (2012), no. 4, 275–285.

Attouch, Mohammed ; Laksaci, Ali ; Ould-Saïd, Elias. Asymptotic distribution of robust estimator for functional nonparametric models. Comm. Statist. Theory Methods 38 (2009), no. 8-10, 1317–1335.

Attouch, Mohammed ; Laksaci, Ali ; Ould Saïd, Elias. Asymptotic normality of a robust estimator of the regression function for functional time series data. J. Korean Statist. Soc. 39 (2010), no. 4, 489–500.

Azzedine, Nadjia ; Laksaci, Ali ; Ould-Saïd, Elias. On robust nonparametric regression estimation for a functional regressor. Statist. Probab. Lett. 78 (2008), no. 18, 3216–3221.

Baća, Martin ; Barrientos, Christian. Graceful and edge-antimagic labelings. Ars Combin. 96 (2010), 505–513.

- Baïllo, Amparo. A note on functional linear regression. *J. Stat. Comput. Simul.* 79 (2009), no. 5-6, 657–669.
- Benhenni, K. ; Ferraty, F. ; Rachdi, M. ; Vieu, P. Local smoothing regression with functional data. *Comput. Statist.* 22 (2007), no. 3, 353–369.
- Berlinet, Alain F. ; Servien, Rémi. Necessary and sufficient condition for the existence of a limit distribution of the nearest-neighbour density estimator. *J. Nonparametr. Stat.* 23 (2011), no. 3, 633–643.
- Beirlant, J. ; Berlinet, A. ; Györfi, L. On piecewise linear density estimators. *Statist. Neerlandica* 53 (1999), no. 3, 287–308.
- Blanke, D. ; Bosq, D. A family of minimax rates for density estimators in continuous time. *Stochastic Anal. Appl.* 18 (2000), no. 6, 871–900.
- Bosq, Denis. La statistique non paramétrique des processus. (French) [The nonparametric statistics of processes] *Rend. Sem. Mat. Univ. Politec. Torino* 45 (1987), no. 2, 1–24 (1988).
- Burba, Florent ; Ferraty, Frédéric ; Vieu, Philippe. Convergence de l'estimateur à noyau des k plus proches voisins en régression fonctionnelle non-paramétrique. (French) [Convergence of the k nearest neighbor kernel estimator in nonparametric functional regression] *C. R. Math. Acad. Sci. Paris* 346 (2008), no. 5-6, 339–342.
- Burba, Florent ; Ferraty, Frédéric ; Vieu, Philippe. k -nearest neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.* 21 (2009), no. 4, 453–469.
- Collomb, G. ; Härdle, W. ; Hassani, S. A note on prediction via estimation of

the conditional mode function. *J. Statist. Plann. Inference* 15 (1987), no. 2, 227–236.

Collomb, Gérard. Analyse d'une série temporelle et prédition non paramétriques : méthodes de la fenêtre mobile et des k points les plus proches en estimation de régression pour des observations dépendantes. (French) [Non-parametric time series analysis and prediction : the window and nearest- k -points methods in regression estimation for dependent observations] Asymptotic theory for non i.i.d. processes (Luminy-Marseille, 1984), 37–75, Travaux Rech., 8, Publ. Fac. Univ. Saint-Louis, Brussels, 1986.

Collomb, Gérard. Estimation non-paramétrique de la régression : revue bibliographique. (French) *Internat. Statist. Rev.* 49 (1981), no. 1, 75–93.

Crambes, Christophe ; Delsol, Laurent ; Laksaci, Ali. Robust nonparametric estimation for functional data. *Functional and operatorial statistics*, 109–116, Contrib. Statist., Physica-Verlag/Springer, Heidelberg, 2008.

Dahmani, Abdelnasser ; Ait Saidi, Ahmed ; Bouhmila, Fatah ; Aissani, Mouloud. Consistency of the Tikhonov's regularization in an ill-posed problem with random data. *Statist. Probab. Lett.* 79 (2009), no. 6, 722–727.

Dabo-Niang, Sophie ; Kaid, Zoulikha ; Laksaci, Ali. Spatial conditional quantile regression : weak consistency of a kernel estimate. *Rev. Roumaine Math. Pures Appl.* 57 (2012), no. 4, 311–339.

Dabo-Niang, Sophie ; Rhomari, Noureddine. Estimation non paramétrique de la régression avec variable explicative dans un espace métrique. (French) [Kernel regression estimation when the regressor takes values in metric space]

- C. R. Math. Acad. Sci. Paris 336 (2003), no. 1, 75–80.
- Dabo-Niang, Sophie ; Rhomari, Noureddine. Kernel regression estimation in a Banach space. J. Statist. Plann. Inference 139 (2009), no. 4, 1421–1434.
- Delsol, Laurent. CLT and \mathbb{L}^q errors in nonparametric functional regression. C. R. Math. Acad. Sci. Paris 345 (2007), no. 7, 411–414.
- Delsol, Laurent. Advances on asymptotic normality in non-parametric functional time series analysis. Statistics 43 (2009), no. 1, 13–33.
- Desu, M. M. ; Narula, Subhash C. Reliability estimation under competing causes of failure. The theory and application of reliability, with emphasis on Bayesian and nonparametric methods (Conf., Univ. South Florida, Tampa, Fla., 1975), Vol. II, pp. 471–481. Academic Press, New York, 1977.
- Devroye, Luc. An application of the Efron-Stein inequality in density estimation. Ann. Statist. 15 (1987), no. 3, 1317–1320.
- Douge, Lahcen. Théorèmes limites pour des variables quasi-associées hilbertiennes. (French) [Limit theorems for Hilbertian quasi-associated variables] Ann. I.S.U.P. 54 (2010), no. 1-2, 51–60.
- Fan, Xian-ling ; Wang, Bi-xiang. Error analysis of nonlinear Galerkin methods for Kuramoto-Sivashinsky equations. Numer. Math. J. Chinese Univ. (English Ser.) 5 (1996), no. 1, 49–61.
- Ferraty, F. ; Vieu, Ph. Functional nonparametric statistics : a double infinite dimensional framework. Recent advances and trends in nonparametric statistics, 61–76, Elsevier B. V., Amsterdam, 2003.

Ferraty, Frédéric ; Laksaci, Ali ; Tadj, Amel ; Vieu, Philippe. Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference* 140 (2010), no. 2, 335–352.

Ferraty, Frédéric ; Laksaci, Ali ; Tadj, Amel ; Vieu, Philippe. Kernel regression with functional response. *Electron. J. Stat.* 5 (2011), 159–171.

Ferraty, Frédéric ; Laksaci, Ali ; Tadj, Amel ; Vieu, Philippe. Estimation de la fonction de régression pour variable explicative et réponse fonctionnelles dépendantes. [Nonparametric regression for functional response and functional regressor under dependence] *C. R. Math. Acad. Sci. Paris* 350 (2012), no. 13-14, 717–720.

Ferraty, Frédéric ; Laksaci, Ali ; Vieu, Philippe. Functional time series prediction via conditional mode estimation. *C. R. Math. Acad. Sci. Paris* 340 (2005), no. 5, 389–392.

Ferraty, Frédéric ; Mas, André ; Vieu, Philippe. Nonparametric regression on functional data : inference and practical aspects. *Aust. N. Z. J. Stat.* 49 (2007), no. 3, 267–286.

Ferraty, Frédéric ; Sued, Mariela ; Vieu, Philippe. Mean estimation with data missing at random for functional covariates. *Statistics* 47 (2013), no. 4, 688–706.

Ferraty, Frédéric ; Vieu, Philippe. Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. (French) [Fractal dimensionality and regression estimation in seminormed vector spaces] *C. R. Acad. Sci. Paris Sér. I Math.* 330 (2000), no. 2, 139–142.

Gheriballah, Abdelkader ; Laksaci, Ali ; Sekkal, Soumeya. Nonparametric M -regression for functional ergodic data. *Statist. Probab. Lett.* 83 (2013), no. 3, 902–908.

Györfi, László ; Härdle, Wolfgang ; Sarda, Pascal ; Vieu, Philippe. Nonparametric curve estimation from time series. Lecture Notes in Statistics, 60. Springer-Verlag, Berlin, 1989. viii+153 pp. ISBN : 3-540-97174-2.

Horváth, Lajos ; Kokoszka, Piotr. Inference for functional data with applications. Springer Series in Statistics. Springer, New York, 2012. xiv+422 pp. ISBN : 978-1-4614-3654-6

Hsing, Tailen ; Eubank, Randall. Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley Series in Probability and Statistics. John Wiley and Sons, Ltd., Chichester, 2015. xiv+334 pp. ISBN : 978-0-470-01691-6

Jones, M. C. ; Lunn, A. D. Transformations and random variate generation : generalised ratio-of-uniforms methods. *J. Statist. Comput. Simulation* 55 (1996), no. 1-2, 49–55.

Jones, Trevor Healy. The heat kernel on noncompact Riemann surfaces. Thesis (Ph.D.)-University of New Brunswick (Canada). ProQuest LLC, Ann Arbor, MI, 2008. 82 pp. ISBN : 978-0494-63685-5

Khardani, Salah ; Lemdani, Mohamed ; Ould Saïd, Elias. Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship. *J. Korean Statist. Soc.* 39 (2010), no. 4, 455–469.

Kudraszow, Nadia L. ; Vieu, Philippe. Uniform consistency of k NN regressors

for functional variables. *Statist. Probab. Lett.* 83 (2013), no. 8, 1863–1870.

Laib, Naâmane ; Louani, Djamel. Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *J. Statist. Plann. Inference* 141 (2011), no. 1, 359–372.

Laksaci, Ali ; Yousfate, Abderrahmane. Estimation fonctionnelle de la densité de l'opérateur de transition d'un processus de Markov à temps discret. (French) [Functional estimate of Markov transition operator density : discrete time case] *C. R. Math. Acad. Sci. Paris* 334 (2002), no. 11, 1035–1038.

Loader, Clive Roland. Change point problems for Poisson processes. Thesis (Ph.D.)-Stanford University. ProQuest LLC, Ann Arbor, MI, 1990. 236 pp.

Louani, Djamel ; Ould-Saïd, Elias. Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *J. Nonparametr. Statist.* 11 (1999), no. 4, 413–442.

Masry, Elias. Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stochastic Process. Appl.* 115 (2005), no. 1, 155–177.

Mammen, E. Nonparametric curve estimation and simple curve characteristics. Nonparametric functional estimation and related topics (Spetses, 1990), 133–139, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 335, Kluwer Acad. Publ., Dordrecht, 1991.

Mechab, Wardia ; Laksaci, Ali. Nonparametric relative regression for associated random variables. *Metron* 74 (2016), no. 1, 75–97.

Narula S.C. and Wellington, J.F. (1977) Prediction, linear regression and the

minimum sum of relative errors. *Technometrics* 19, 185U190.

Quintela-Del-Río, A. ; Vieu, Ph. A nonparametric conditional mode estimate. *J. Nonparametr. Statist.* 8 (1997), no. 3, 253–266.

Ramsay, J. O. ; Silverman, B. W. Functional data analysis. Second edition. Springer Series in Statistics. Springer, New York, 2005. xx+426 pp. ISBN : 978-0387-40080-8 ; 0-387-40080-X

Roussas, G. G. Asymptotic normality of the maximum likelihood estimate in Markov processes. *Metrika* 14 1968 62–70.

Roussas, George G. Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.* 21 1969 73–87.

Tukey, John W. Curves as parameters, and touch estimation. 1961 Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I pp. 681–694 Univ. California Press, Berkeley, Calif.

Vieu, Philippe. Quadratic errors for nonparametric estimates under dependence. *J. Multivariate Anal.* 39 (1991), no. 2, 324–347.

Watson, Geoffrey S. Smooth regression analysis. *Sankhyā ? Ser. A* 26 1964 359–372.

Yang, Hui Min ; Ye, Guo Ju ; Zhou, Xue Yuan ; Zhou, Hao ; Wang, Ou. An existence-uniqueness theorem for the solution to an equation involving the distributional derivative. (Chinese) *J. Jilin Univ. Sci.* 51 (2013), no. 5, 749–754.

Zhang, Shu Yi. Convergence of multi-step parallel iterative algorithms for

uniformly Lipschitz asymptotically φ_i -type quasi-pseudo-contractive mappings. (Chinese) *J. Systems Sci. Math. Sci.* 33 (2013), no. 10, 1233–1242.

Chapitre 2

Cas vectoriel

2.1 Introduction

Let $Z_i = (X_i, Y_i)$, $i \in \mathbb{N}^N$ be a $\mathbb{R}^d \times \mathbb{R}$ -valued measurable strictly stationary spatial process, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Our main aim in this work is to study the spatial co-variation between both variables (X_i and Y_i). Recall that, this links between $(X_i$ and $Y_i)$ is usually modeled through the conditional expectation $\mathbb{E}[Y_i|X_i]$. The latter is based on some restrictive condition that is the variance of the residual is the same for all observations. This consideration gives the same weight for all observations, which is inadequate when the data contains some outliers. In this paper, we consider an alternative approach allow to construct an efficient spatial predictor even if the data is affected by the presence of outliers. Our approach is based on the minimization of the mean squared relative error as prediction loss function. Formally, our nonparametric predictor is obtained by minimizing, with

respect to (w.r.t.) t , the following criterium

$$\text{For } Y > 0 \quad E \left[\left(\frac{Y - t}{Y} \right)^2 | X \right]. \quad (1)$$

Clearly, this loss function is more meaningful measure of prediction performance in the presence of outliers, because it takes into consideration the range of predicted values.

Notice that, this kind of model so-called relative error regression has been widely studied in parametric regression analysis (linear or multiplicative models). See, for instance, Narula and Wellington (1977), Shen et al. (1985) or Bernhard and Stahlecker (2003) for previous study and Yang and Ye (2013) for recent advanced and references). The nonparametric analysis of these models has not yet been fully explored. To the best of our knowledge, only the paper by Jones et al (2008) has paid attention to study the nonparametric prediction via relative error regression. They studied the asymptotic properties of an estimator minimizing the sum of the squared relative errors by considering both (kernel method and local linear approach). On the other hand, spatial statistics has become a major topic of research in applied statistics, mainly due to the interaction with other applied fields (see, Volker (2014) and the references therein). This area concerns the modelization of spatially varying random variables, which appears in applied sciences such as economics, soil science, epidemiology or environmental science, among others. Spatially referred data can not be usually treated as independent, so, suitable statistical methodologies for estimation and prediction within the spatial context are necessary (see Cressie (1993) or Diggle and Ribeiro (2007), among others). One of the main difficulties that arise in the analysis

of spatial data comes from the fact that points in the N -dimensional space do not have a linear order, as it happens for time series data.

Nonparametric methods have been also considered for spatial dependence. The first results have been obtained by Tran (1990). He obtained the asymptotic normality for the density kernel estimator, whereas the nonparametric spatial regression problem has been studied by Lu and Chen (2004) and Biau and Cadre (2004), who used the Nadaraya-Watson weights to obtain a kernel estimator, establishing the weak convergence and asymptotic distribution. The nonparametric auto-regression model in a prediction context on random fields has been studied by Carbon *et al.* (2007). Li and Tran (2009) use the k -nearest neighbor (kNN) technique to estimate the spatial nonparametric regression. They showed the asymptotic normality of the construct estimate. We return to El-Machkouri and Stoica (2010), Robinson (2011) and Dabo-Niang et al. (2014) for recent advances and references in nonparametric spatial data analysis.

The exploration and analysis of data in presence of outliers is a great challenge in statistics. In particular, in spatial statistics several robust models resistant to these anomalies have been studied. Concerning the nonparametric modeling case, the first approach is given by Xu and Wang (2008). The latter consider a local linear estimate of the regression function based on the least absolute deviation. Hallin, et al. (2009) establish the consistency and asymptotic normality of a spatial version of the local linear estimate of the conditional quantiles. The spatial version of the M-estimation of the regression function has introduced by Gheriballah et al. (2010). They obtained the

almost complete convergence and the asymptotic normality of this estimate. Dabo-Niang and Thiam (2010) have paid attention to study nonparametric quantile regression by the L_1 method. They stated the weak consistency and the asymptotic normality of the constructed estimator.

In this paper, we are interested in the spatial prediction problem via the nonparametric relative regression. More precisely, we construct a nonparametric spatial predictor by using a kernel method. Under some general mixing assumptions, we state the almost complete convergence (with rate) and the asymptotic normality of this estimate. Noting that, this work constitutes a generalization of the study of Jones et al (2008) in time series analysis to the spatially dependence case.

The paper is organized as follows : We present our model in Section 2. Section 3 is dedicated to the almost complete convergence. Section 4 is devoted to the asymptotic normality results of our kernel estimate. In Section 4, we discuss the impact of our asymptotic result in some statistical problems such as the choice of the smoothing parameters, the determination of confidence intervals and in prediction problems. In Section 6, we evaluate the performances of our estimator with a Monte Carlo study and a real data example. The proofs of the auxiliary results are relegated to the Appendix.

2.2 Construction of the spatial estimate

Throughout the paper, we suppose that (Z_i) is observed over a rectangular domain $I_{\mathbf{n}} = \{\mathbf{i} = (i_1, \dots, i_N) \in \mathbb{Z}^N, 1 \leq i_k \leq n_k, k = 1, \dots, N\}$, $\mathbf{n} = (n_1, \dots, n_N) \in \mathbb{Z}^N$. A point \mathbf{i} will be referred to as a *site*. We will write $\mathbf{n} \rightarrow \infty$

if $\min\{n_k\} \rightarrow \infty$ and $|\frac{n_j}{n_k}| < C$ for a constant C such that $0 < C < \infty$ for all j, k such that $1 \leq j, k \leq N$. For $\mathbf{n} = (n_1, \dots, n_N) \in \mathbb{Z}^N$, we set $\widehat{\mathbf{n}} = n_1 \times \dots \times n_N$ and $\mathbf{1} = (1, \dots, 1) \in \mathbb{Z}^N$. Clearly, the solution of (1) can be explicitly expressed by

$$\theta(x) = \frac{\mathbb{E}[Y^{-1}|X=x]}{\mathbb{E}[Y^{-2}|X=x]}.$$

Thus, a natural estimate of θ is

$$\widetilde{\theta}(x) = \frac{\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-1} K(h^{-1}(x - X_{\mathbf{i}}))}{\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-2} K(h^{-1}(x - X_{\mathbf{i}}))} \quad (2)$$

where K is a kernel and $h = h_{K,\mathbf{n}}$ is a sequence of positive real numbers. Our main goal is to study the asymptotic proprieties (consistency and asymptotic normality) of the nonparametric estimate $\widetilde{\theta}$ of θ when the random field $(Z_{\mathbf{i}}, \mathbf{i} \in \mathbb{N}^N)$ satisfies the following mixing condition :

$$\left\{ \begin{array}{l} \text{There exists a function } \varphi(t) \downarrow 0 \text{ as } t \rightarrow \infty, \text{ such that} \\ \forall E, E' \text{ subsets of } \mathbb{N}^N \text{ with finite cardinals} \\ \alpha(\mathcal{B}(E), \mathcal{B}(E')) = \sup_{B \in \mathcal{B}(E), C \in \mathcal{B}(E')} |\mathbb{P}(B \cap C) - \mathbb{P}(B)\mathbb{P}(C)| \\ \leq s(\text{Card}(E), \text{Card}(E')) \varphi(\text{dist}(E, E')), \end{array} \right. \quad (3)$$

where $\mathcal{B}(E)$ (*resp.* $\mathcal{B}(E')$) denotes the Borel σ -field generated by $(Z_{\mathbf{i}}, \mathbf{i} \in E)$ (*resp.* $(Z_{\mathbf{i}}, \mathbf{i} \in E')$), $\text{Card}(E)$ (*resp.* $\text{Card}(E')$) the cardinality of E (*resp.* E'), $\text{dist}(E, E')$ the Euclidean distance between E and E' and $s : \mathbb{N}^2 \rightarrow \mathbb{R}^+$ is a symmetric positive function nondecreasing in each variable such that :

$$s(n, m) \leq C \min(n, m), \quad \forall n, m \in \mathbb{N}. \quad (4)$$

We point out that condition (3) is used in Tran (1990) and is satisfied by a large class of spatial models (see Guyon (1987) for some examples). Noting

that if $s \equiv 1$, then Z_i is called strongly mixing (see Doukhan *et al.* (1994) for discussion on mixing and examples).

2.3 Main results

Throughout the paper, we fix a compact subset S in \mathbb{R}^d and when no confusion will be possible, we will denote by C or/and C' some strictly positive generic constants. Furthermore, for (X, Y) a couple of random variables with same distribution as (X_i, Y_i) , we denote by $r_l(\cdot) = E[Y^{-l}|X = \cdot]$ the conditional l -inverse moments of Y given X ; $l = 1, 2, 3, 4$ and by $g_l(\cdot) = r_l(\cdot)f(\cdot)$, with f is the density of X .

2.3.1 Consistency

Our first goal in this section, is to establish the uniform almost complete convergence¹ (a.co.) of $\tilde{\theta}(x)$ to $\theta(x)$ over compact S . To do that, we consider the following assumptions :

- (H1) The density f is a positive function, of class C^2 in S .
- (H2) The functions $r_l(\cdot)$; ($l = 1, 2$) is of class C^2 in S
- (H3) The joint probability density $f_{i,j}$ of X_i and X_j exists and satisfies

$$|f_{i,j}(u, v) - f(u)f(v)| \leq C \text{ for some constant } C \text{ and for all } u, v, i \text{ and } j.$$

1. Let $(z_n)_{n \in \mathbb{N}}$ be a sequence of real r.v.'s. We say that z_n converges almost completely (a.co.) toward zero if, and only if, $\forall \epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon) < \infty$. Moreover, we say that the rate of the almost complete convergence of z_n to zero is of order u_n (with $u_n \rightarrow 0$) and we write $z_n = O_{a.co.}(u_n)$ if, and only if, $\exists \epsilon > 0$ such that $\sum_{n=1}^{\infty} P(|z_n| > \epsilon u_n) < \infty$. This kind of convergence implies both almost sure convergence and convergence in probability.

(H4) The mixing coefficient defined in (2) satisfies

$$\sum_{i=1}^{\infty} i^{N-1} (\varphi(i))^a < \infty, \quad \text{for some } 0 < a < \frac{1}{2}.$$

(H5) K is continuous lipschitz function, symmetric , with compact support.

(H6) There exists $\gamma \in (0, 1)$ such that

$$\left\{ \begin{array}{l} \sum_{\mathbf{n}} \widehat{\mathbf{n}} (\log \widehat{\mathbf{n}})^{-1} \varphi(p_{\mathbf{n}}) < \infty \text{ for } p_{\mathbf{n}} = O \left(\frac{\widehat{\mathbf{n}}^{1-\gamma} h^d}{\log \widehat{\mathbf{n}}} \right)^{1/2N} \\ \text{and} \\ \widehat{\mathbf{n}}^\delta h \rightarrow \infty \quad \text{for certain } \delta > 0. \end{array} \right.$$

(H7) The inverse moments d'order $l = 1, 2$ of the response variable such that,

$$E \left(\exp(|Y^{-l}|) \right) \leq C \quad \text{and } \forall \mathbf{i}, \mathbf{j} \quad E \left(|Y_{\mathbf{i}}^{-l} Y_{\mathbf{j}}^{-l}| \mid X_{\mathbf{i}}, X_{\mathbf{j}} \right) \leq C'.$$

We precise that our conditions are very usual in this context of nonparametric spatial analysis (see Tran(1990) or Li and Tran (2009)). Furthermore, Assumptions (H1) and (H2) are regularity condition permit to evaluate the bias term of the estimator (1). Condition (H3) is automatically fulfilled when the densities $f_{\mathbf{i}, \mathbf{j}}$ and f are bounded. While assumptions (H4)-(H7) are technical conditions imposed in order to attaint the brevity of proofs and to obtain a convergence rate comparable to the i.i.d case.

Theorem 2.3.1 *Under assumptions (H1)-(H7) and, if $\inf_{x \in S} g_2(x) > 0$, we have :*

$$\sup_{x \in S} |\tilde{\theta}(x) - \theta(x)| = O(h^2) + O_{a.co.} \left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} h^d}} \right) \quad a.co. \quad (5)$$

2.3.2 The asymptotic normality

This section is devoted to the establishment of the asymptotic normality of $\tilde{\theta}(x)$. For this, we keep the conditions of the previous section and we replace Assumption (H6) by

(H6') There exists $\xi \in (0, 1)$ such that,

$$\left\{ \begin{array}{l} \widehat{\mathbf{n}} \sum_{i=1}^{\infty} i^{N-1} \varphi(iq_{\mathbf{n}}) \rightarrow 0, \quad \text{for } q_{\mathbf{n}} = o\left(\left[\widehat{\mathbf{n}}^{1-\gamma} h^{d(1+2(1-\xi)N)}\right]^{1/(2N)}\right), \\ h^{-d(1-\xi)} \sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1} \varphi^{1-\xi}(i) \rightarrow 0, \\ \text{and} \\ \widehat{\mathbf{n}} h^{d+4} \rightarrow 0. \end{array} \right.$$

We obtain the following Theorem :

Theorem 2.3.2 Assume that (H1)-(H5), (H6') and (H7) hold, then we have, for any $x \in \mathcal{A}$,

$$\left(\frac{\widehat{\mathbf{n}} h^d}{\sigma^2(x)} \right)^{1/2} \left(\tilde{\theta}(x) - \theta(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty.$$

where

$$\sigma^2(x) = \frac{(g_2(x) - 2\theta(x)r_3(x) + \theta^2(x)r_4(x))}{g_2^2(x)} \int_{R^d} K^2(z) dz,$$

$$\mathcal{A} = \left\{ x \in S, (g_2(x) - 2\theta(x)r_3(x) + \theta^2(x)r_4(x)) g_2^2(x) \neq 0 \right\}$$

and $\xrightarrow{\mathcal{D}}$ means the convergence in distribution.

2.4 Discussion and Application

2.4.1 Comparison to existing results

As indicated in the introduction, the general framework of this paper is the nonparametric regression modeling of spatial data. We have proposed an estimation procedure based on minimization of the mean squared relative error. This method is an alternative to the classical conditional mean regression. As asymptotic results, we have established the almost completely consistency and the asymptotic normality of the constructed estimator. From theoretical point view, these results provide an important advance in this topic. Indeed, while the contribution of Jones et al. (2008) treat only the i.d.d. case we study, here, a more general case that is the spatial dependency. In the other hand, our results complete the contribution of Jones and al. (2008) because we have studied another modes of convergence that are the almost completely and the asymptotic normality. Moreover, the convergence rate stated here is the same as obtained by Biau and Cadre (2004) for the classical regression, even if these authors consider only the weak consistency. In addition, from practical point of view, it is well known that the relative error regression has more advantages than the classical one by its outlier-resistance properties. It should be noted that, in contrast to traditional outliers, spatial outliers are more difficult to detect them, because, they have different form (global-outliers and local outliers). Thus, it is very important to develop a model insensitive to the presence of outliers. In conclusion, we can say that the importance of the present contribution is that under a similar conditions to Biau and Cadre (2004) on the classical regression we prove a strong consis-

tency with rate of an alternative regression model which is more adapted to the data affected by the presence of spatial outliers.

2.4.2 Application to prediction

Let us now discuss the applicability of our results to prediction problem. For this, we consider $(Z_t)_{t \in \mathbb{R}^N}$ be a \mathbb{R} -valued strictly stationary random spatial processus assumed to be bounded and observable over some subset $T \subset \mathbb{R}^N$. Clearly, in practice, the spatial continuously indexed process is observed at a finite number of appropriately chosen points. As noticed by Dabo-Niang and Yao (2007), the most used sampling designs, in practice, are deterministic design (points are chosen according to a deterministic rule, for example, periodic sampling) and random design (points are chosen randomly, for example, Poisson sampling). For the sake of shortness we consider the first one for which the process (Z_t) is observed on a discrete grid $I_n = \{\delta_n, 2\delta_n, \dots, n\delta_n\}^N$ with $\delta_n = o(1)$. Furthermore, our approach can be used to predict the value Z_{s_0} at an unobserved location $s_0 \notin I_n$. Indeed, we suppose that, the value of Z_{s_0} depends only on the values of the process (Z_t) in a bounded neighborhood $\mathcal{V}_{s_0} \subset I_n$ of s_0 . In addition, we suppose that \mathcal{V}_{s_0} such that $\mathcal{V}_{s_0} = s_0 + \mathcal{V}$, where \mathcal{V} is a fixed bounded set that does not contain 0. Similarly to Biau and Cadre (2004), $\tilde{\theta}(\tilde{Z})$ is the best approximation of Z_{s_0} given $\tilde{Z} = (Z_t, t \in \mathcal{V}_{s_0})$. The quantities $\tilde{\theta}(\tilde{Z})$, is computed by using the n^N pairs of r.v (Y_i, X_i) with

$$\forall i \in I_n, Y_i = Z_i, \quad \text{and} \quad X_i = (Z_t, t \in \mathcal{V} + i).$$

Obviously, the uniform consistency of Theorem 3.2.1 allows to deduce the following Corollary.

Corollary 1 Under the hypotheses of Theorem 3.2.1, we have

$$\tilde{\theta}(\tilde{Z}) - \theta(\tilde{Z}) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Remark 1 The same things can be concluded in the discrete spatial processus $(Z_t)_{t \in \mathbb{N}^*_N}$ (See, Biau and Cadre (2004) for more discussion in this case)

2.4.3 Conditional Confidence Interval

One of the most application of the asymptotic normality is the construction of the confidence interval for the true value. The determination of this interval requires the estimation of the unknown quantity $\sigma(x)$. A plug-in estimate for the asymptotic standard deviation $\sigma(x)$ can be obtained by estimating $r_l(x) = \mathbb{E}[Y^{-l}|X = x]; l = 1, 2, 3, 4$ by

$$\tilde{r}_l(x) = \frac{\sum_{i \in \mathcal{I}_n} Y_i^{-l} K(h^{-1}(x - X_i))}{\sum_{i \in \mathcal{I}_n} K(h^{-1}(x - X_i))}.$$

Furthermore, We get

$$\hat{\sigma}(x) := \left(\frac{\left(\tilde{g}_2(x) - 2\tilde{\theta}(x)\tilde{r}_3(x) + \tilde{\theta}^2(x)\tilde{r}_4(x) \right)}{\tilde{g}_2^2(x)} \int_{R^d} K^2(z) dz \right)^{1/2}.$$

Such estimator may be easily calculated in practice and we have the following approximate $(1 - \beta)$ confidence interval for $\theta(x)$

$$\tilde{\theta}(x) \pm t_{1-\zeta/2} \times \left(\frac{\hat{\sigma}^2(x)}{\hat{\mathbf{n}} h^d} \right)^{1/2}$$

where $t_{1-\zeta/2}$ denotes the $1 - \zeta/2$ quantile of the standard normal distribution.

2.5 On the finite-sample performance of the method

In this section we compare the finite-sample performance of the relative-error regression (R.E.R.) to the classical one (C.R.) via a short Monte Carlo study. More precisely, we compare the finite-sample efficiency of both regression functions as spatial prediction tools. In order to highlight the main feature of our procedure, we compare their sensitivity to the presence of outliers. For this purpose, we consider the following model

$$Y_{\mathbf{i}} = 5 \log((4 - X_{\mathbf{i}})^2 + 2) + \epsilon_{\mathbf{i}}, \quad \mathbf{i} = (i_1, i_2)$$

where $\epsilon_{(i_1, i_2)}$ follows a normal distribution with mean 0 and variance 0.25. For the sake of simplicity, we consider the same univariate spatial process $X_{\mathbf{i}}$ used by Xu and Wang (2008) defined by

$$X_{\mathbf{i}} = (\sqrt{2/m}) \sum_{k=1}^m (\cos(i_1 w_{k,1} + i_2 w_{k,2} + v_k))$$

where $w_{k,s}$, $s = 1, 2$ $k = 1, 2 \dots 500$ are independently, identically distributed with the standard normal distribution and are independent of v_k , $k = 1, \dots 500$ which are independently and identically distributed with the uniform distribution on $[-\pi, \pi]$. Recall that as $m \rightarrow \infty$, $X_{\mathbf{i}}$ is a Gaussian spatial ergodic process (see, Gressie 1993) which is an example of α -mixing spatial process.

We generate the random field $(X_{\mathbf{i}}, Y_{\mathbf{i}})$ at $\hat{\mathbf{n}} = 30 \times 30$ sites. Now, for our comparison study, we suppose that the response variable on a lattice point \mathbf{i}_0 is missing and we predict this missing observation by using the two regression

models

$$\tilde{\theta}(X_{i_0}) = \frac{\sum_{i \in \mathcal{I}_n} Y_i^{-1} K(h^{-1}(X_{i_0} - X_i))}{\sum_{i \in \mathcal{I}_n} Y_i^{-2} K(h^{-1}(X_{i_0} - X_i))}$$

et

$$\hat{\theta}(X_{i_0}) = \frac{\sum_{i \in \mathcal{I}_n} Y_i K(h^{-1}(X_{i_0} - X_i))}{\sum_{i \in \mathcal{I}_n} K(h^{-1}(X_{i_0} - X_i))}.$$

The performance of both estimators is described by absolute error

$$AE(\hat{r}) = |Y_{i_0} - \hat{r}(X_{i_0})|$$

where \hat{r} means both regression models $\tilde{\theta}$ and $\hat{\theta}$.

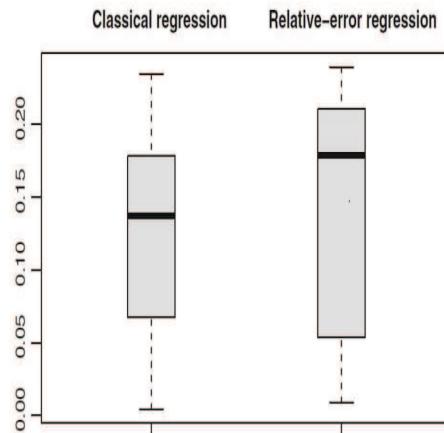


FIGURE 2.1 – The AE-errors of both models .

As in all smoothing methods, the choice of the smoothing parameter h has a crucial role in the computational issues. However, there is no data-driven

rule which permits to select automatically and optimally the bandwidth parameter in nonparametric spatial regression even in the classical case. In this illustration, we use the cross-validation procedure proposed by Xu and Wang (2008) for which the bandwidth h is chosen via the following rule :

$$h_{opt} = \arg \min_h CV(h) = \arg \min_h \sum_{\mathbf{j} \in \mathcal{I}_n} |Y_{\mathbf{j}} - \hat{r}^{(-\mathbf{j})}(X_{\mathbf{j}})|$$

where $\hat{r}^{(-\mathbf{j})}$ is the leave-one-out-curve estimator of $\tilde{\theta}$ and $\hat{\theta}$.

The box-plot of AE-error of both models in various values of \mathbf{i}_0 is given in Figure 1.

We observe that there is no meaningful difference between this spatial predictors. The two predictors are basically equivalent and both show the good behavior.

Now, in order to investigate the features of our approach, we introduced some artificial outliers by multiplying 15% values of Y by 10. We box-plot AE-errors of both models

Further we see from Figure 2 that the relative regression error is much more better than the classical one in this case. Moreover, looking at both figures, it appears clearly the AE-error of the C.R model has dramatically changed compared to the R.E.R case. This statement confirms that the relative error regression is more robust than the classical regression.

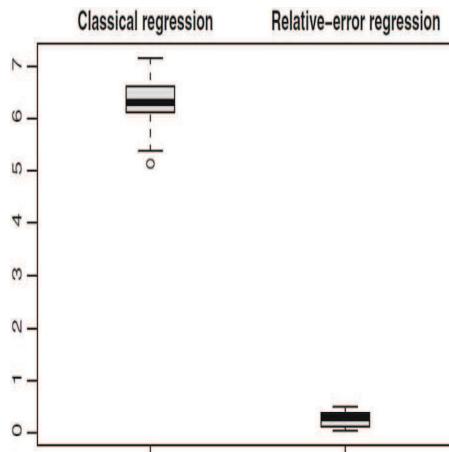


FIGURE 2.2 – The AE-errors in presence of outliers.

2.6 Application on real data

Our aim in this section is to show how we can implement easily and rapidly our approach in practice. To do that, we consider the *ozone2* spatial data in *fields* R-package. It concerns the average daily of ozone concentration for 153 sites in the midwestern US at 89 days. This data is measured in separated stations and is usually used as lattice data in *LatticeKrig* R-package. In general the spatial modeling of air quality data has been actively studied in the last years (see, Omidi and Mohammadzadeh (2015) for a list of references). Such approach is essentially used to produce maps of air pollutants. In the same context, we focus, in this illustration, on the prediction of the concentration of ozone in an unobserved stations knowing the values of this quantity on some fixed number of nearest stations. Precisely, we fix a day t and we predict $Y_{\mathbf{s}_0}$: the concentration ozone in station \mathbf{s}_0 at t , given the regressors $X_{\mathbf{s}_0} = (Y_{\mathbf{s}_0^1}, \dots, Y_{\mathbf{s}_0^k})$: vector of ozone concentrations at k -nearest

stations to s_0 at the same day t .

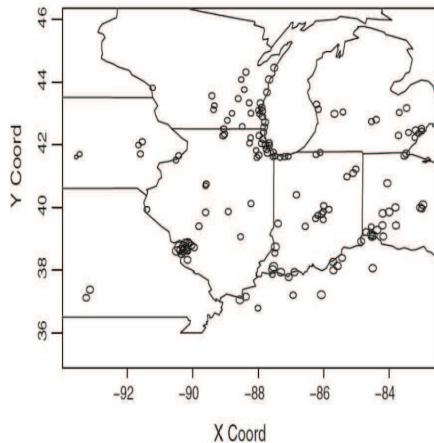


FIGURE 2.3 – The locations of stations in midwestern US.

Noting that the emission of volatile organic compounds is the principal source of ozone in urban areas and the formation of the latter is closely related to meteorological conditions, road traffic conditions and industrial activities. Furthermore, the studied geographic zone is very vast region that contains several states (see, Figure 3) with different meteorological conditions, different road traffic conditions and different industrial activities. Therefore, the studied data are not homogeneous and exhibits some heterogeneity. It is worth to noting that the outlier detection of air quality data has been considered by many authors in the last years (see Martínez et al. (2014) or Bobbia et al.(2015)). We refer to Filzmoser P. et al.(2014) or Liu et al (2010) for some algorithms spatial outlier detection. In this contribution we adopt the procedure of Liu et al (2010) to determine the outlying observations in our data-set. Precisely, we use the routine *RWBP* in *RWBP* R-package to detect

them. The latter shows various outlierness scores in this data. In particular, the lowest score are given in 57th day and 64th while the the strongest score are given in 67th and 86th day. Therefore, 57th and 64th days are more affected by the presence of outliers than the 67th and 86th days. In order to illustrate the main feature of our approach, we compare both spatial predictors (the classical regression (C.R.) and the relative error regression (R.E.R.)) in these four days.

Concerning the spatial correlation of the data, we recall that, the spatial dependency is, often, modeled through the covariogram or the variogram function (see Cressie, 1993). However, in our theoretical study we have exploited the correlation structure by means of Conditions (2), (3) and (H4). So, the main challenge is how we can control this aspect in practice. As suggested by Dabo-Niang *et al.* (2009) a simple manner to control the spatial correlation in practice is to integrate it in the computation of the estimator by taking $\theta(X_k)$, for a new observation X_k in new site $k \notin I_n$, as

$$\tilde{\theta}(X_k) = \frac{\sum_{i \in I_n} K(h_K^{-1}d(X_k, X_i))Y_i^{-1}\mathbb{I}_{W_k}(i)}{\sum_{i \in I_n} K(h_K^{-1}d(X_k, X_i))Y_i^{-2}}$$

$$\text{and } \hat{\theta}(X_k) = \frac{\sum_{i \in I_n} K(h_K^{-1}d(X_k, X_i))Y_i\mathbb{I}_{W_k}(i)}{\sum_{i \in I_n} K(h_K^{-1}d(X_k, X_i))}$$

where W_k is a vicinity set of the fixed site k . Based on the isotropic condition, we proceed with the following vicinity set

$$W_k = \{i, \text{ such that } dis(i, k) \leq \nu_n\}$$

where ν_n is a appropriate sequence of positive real numbers and dis is the distance between the sites. Furthermore, in order to exploit the spatial correlation of this data-set and to show the influence of the local outliers as well as

the global one we examine the performance of our methodology with respect the values of ν_n and the fixed number of nearest stations k . More precisely, we select ν_n among the quantile of order q of the estimated vector of the distance between the sites given by the code *geodistance* in *McSpatial* package and we calculate following mean squared prediction errors :

$$\text{MAE}(\hat{r}) = \frac{1}{\hat{n}} \sum_{i \in \mathcal{I}_n} |Y_i - \hat{r}(X_i)|$$

where \hat{r} means both regression models $\tilde{\theta}$ and $\hat{\theta}$. Therefore, these errors are computed as function of q and k . We point out that we have used the same smoothing selection method and the multidimensional gaussian kernel K . We calculate the above errors for the day 57th, 64th, 67th and 86th. We report the obtained result in the following results

Days	q	k	MAE(R.E.R.)	MAE(R.C.)
57	0.25	1	5.25	10.10
	0.75	1	5.78	10.87
	0.5	3	6.67	12.07
	0.5	5	7.45	12.79
64	0.25	1	4.56	8.90
	0.75	1	5.92	10.08
	0.5	3	6.87	10.82
	0.5	5	6.94	11.19
67	0.25	1	3.11	2.10
	0.75	1	3.93	2.87
	0.5	3	4.42	4.07
	0.5	5	5.21	4.68
86	0.25	1	3.09	2.72
	0.75	1	3.29	3.91
	0.5	3	4.67	4.26
	0.5	5	4.03	5.84

Table 1 Comparison results.

Undoubtedly, the performance of both spatial prediction models is closely linked to the structure of date. But, it seem that the relative error regression is more resistant to all deviation in the structure of the data such as the outlying observation, dimension of regressor k , or the measure of spatial dependency q . This fact is confirmed by the variability of MAE-error for both models. It varied between (3.09, 7.45) for the R.E.R, while is between

(2.10,12.79) for the C.R model. In conclusion, we can say that the R.E.R. model is more robust than the classical regression.

2.7 Proofs

Proof of Theorem 3.2.1 Let

$$\tilde{\theta}(x) = \frac{\tilde{g}_1(x)}{\tilde{g}_2(x)}$$

with

$$\tilde{g}_l(x) = \frac{1}{\widehat{\mathbf{n}}h^d} \sum_{\mathbf{i} \in \mathcal{I}_n} Y_{\mathbf{i}}^{-l} K(h^{-1}(x - X_{\mathbf{i}})) \quad \text{for } l = 1, 2.$$

Next, we use the following decomposition :

$$\tilde{\theta}(x) - \theta(x) = \frac{1}{\tilde{g}_2(x)} [\tilde{g}_1(x) - g_1(x)] + [g_2(x) - \tilde{g}_2(x)] \frac{\theta(x)}{\tilde{g}_2(x)} \quad (6)$$

Thus, Theorem 3.2.1 is a consequence of the following intermediate results

LEMMA 1 *Under hypotheses (H1), (H2), and (H5), we have, for $l = 1, 2$, that :*

$$|E\tilde{g}_l(x) - g_l(x)| = O(h^2).$$

Proof of Lemma 1. By a change of variables, we get, for $l = 1, 2$,

$$\begin{aligned} E[\tilde{g}_l(x)] &= \frac{1}{h^d} \int_{\mathbb{R}^d} E[Y^{-l}|X = u)] K\left(\frac{x-u}{h}\right) f(u) du \\ &= \int_{\mathbb{R}^d} g_l(x - hz) K(z) dz. \end{aligned}$$

Since both functions f and r_l are of class \mathcal{C}^2 , we use a Taylor expansion of $g_l(\cdot)$ to write, under (H4)

$$|E[\tilde{g}_l(x)] - g_l(x)| \leq Ch^2.$$

The last result complete the proof of lemma. ■

LEMMA 2 Under hypotheses (H3)-(H7), we have, for $l = 1, 2$, that :

$$\sup_{x \in S} |\tilde{g}_l(x) - E\tilde{g}_l(x)| = O_{a.co.} \left(\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h^d}} \right).$$

Proof of Lemma 2. Consider

$$\tilde{g}_l^*(x) = \frac{1}{\hat{\mathbf{n}} h^d} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K\left(h^{-1}(x - X_i)\right) Y_i^{-l} \mathbb{1}_{|Y_i^{-1}| < \mu_{\mathbf{n}}} \text{ with } \mu_{\mathbf{n}} = \hat{\mathbf{n}}^{\gamma/2}.$$

Therefore, it suffices to prove the following intermediates results

$$\sup_{x \in S} |E[\tilde{g}_l^*(x)] - E[\tilde{g}_l(x)]| = O \left(\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h^d}} \right), \quad (7)$$

$$\sup_{x \in S} |\tilde{g}_l^*(x) - \tilde{g}_l(x)| = O_{a.co.} \left(\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h^d}} \right) \quad (8)$$

and

$$\sup_{x \in S} |\tilde{g}_l^*(x) - E[\tilde{g}_l^*(x)]| = O_{a.co.} \left(\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h^d}} \right). \quad (9)$$

— Firstly, for (9), we use the compactness of S to write

$$S \subset \bigcup_{j=1}^{d_{\mathbf{n}}} B(x_k, \tau_{\mathbf{n}}),$$

with $d_{\mathbf{n}} \leq \hat{\mathbf{n}}^\beta$ and $\tau_{\mathbf{n}} \leq d_{\mathbf{n}}^{-1}$ where $\beta = \frac{\delta(d+2)}{2} + \frac{1}{2} + \frac{\gamma}{2}$. So, for all $x \in S$, we pose

$$k(x) = \arg \min_{k \in \{1, \dots, d_{\mathbf{n}}\}} \|x - x_k\|.$$

Thus, for $l = 1, 2$,

$$\begin{aligned} \sup_{x \in S} \left| \tilde{g}_l^*(x) - E[\tilde{g}_l^*(x)] \right| &\leq \underbrace{\sup_{x \in S} \left| \tilde{g}_l^*(x) - \tilde{g}_l^*(x_{k(x)}) \right|}_{T_1} \\ &+ \underbrace{\sup_{x \in S} \left| \tilde{g}_l^*(x_{k(x)}) - E[\tilde{g}_l^*(x_{k(x)})] \right|}_{T_2} \\ &+ \underbrace{\sup_{x \in S} \left| E[\tilde{g}_l^*(x_{k(x)})] - E[\tilde{g}_l^*(x)] \right|}_{T_3}. \end{aligned}$$

- Furthermore, for T_2 , for both $l = 1, 2$, we have

$$\sup_{x \in S} \left| \tilde{g}_l^*(x_{k(x)}) - E[\tilde{g}_l^*(x_{k(x)})] \right| = \max_{k=1, \dots, d_n} \left| \tilde{g}_l^*(x_k) - E[\tilde{g}_l^*(x_k)] \right|$$

Thus it suffices to evaluate almost completely

$$\max_{k=1, \dots, d_n} \left| \tilde{g}_l^*(x_k) - E[\tilde{g}_l^*(x_k)] \right|.$$

To do that, we write :

$$\tilde{g}_l^*(x_k) - E[\tilde{g}_l^*(x_k)] = \frac{1}{\widehat{\mathbf{n}} h^d} \sum_{\mathbf{i} \in \mathcal{I}_n} \Delta_{\mathbf{i}}$$

where

$$\Delta_{\mathbf{i}} = Y_{\mathbf{i}}^{-l} K(h^{-1}(x_k - X_{\mathbf{i}})) \mathbb{1}_{(|Y_{\mathbf{i}}^{-1}| < \mu_n)} - E[Y^{-l} K(h^{-1}(x_k - X)) \mathbb{1}_{(|Y^{-1}| < \mu_n)}].$$

Now, similarly to (Tran (1990)), we use the classical spatial block decomposition for the sum $\sum_{\mathbf{i} \in \mathcal{I}_n} \Delta_{\mathbf{i}}$ as follows

$$U(1, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k=2j_k p_n + 1 \\ k=1, \dots, N}}^{2j_k p_n + p_n} \Delta_{\mathbf{i}},$$

$$\begin{aligned}
U(2, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+1 \\ k=1,\dots,N-1}}^{2j_k p_{\mathbf{n}}+p_{\mathbf{n}}} \sum_{i_N=2j_N p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{(j_N+1)p_{\mathbf{n}}} \Delta_i, \\
U(3, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+1 \\ k=1,\dots,N-2}}^{2j_k p_{\mathbf{n}}+p_{\mathbf{n}}} \sum_{i_{N-1}=2j_{N-1} p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{2(j_{N-1}+1)p_{\mathbf{n}}} \sum_{i_N=2j_N p_{\mathbf{n}}+1}^{2j_N p_{\mathbf{n}}+p_{\mathbf{n}}} \Delta_i, \\
U(4, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+1 \\ k=1,\dots,N-2}}^{2j_k p_{\mathbf{n}}} \sum_{i_{N-1}=2j_{N-1} p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{2(j_{N-1}+1)p_{\mathbf{n}}} \sum_{i_N=2j_N p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{2(j_N+1)p_{\mathbf{n}}} \Delta_i,
\end{aligned}$$

and so on. Finally

$$\begin{aligned}
U(2^{N-1}, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+p_{\mathbf{n}}+1 \\ k=1,\dots,N-1}}^{2(j_k+1)p_{\mathbf{n}}} \sum_{i_N=2j_N p_{\mathbf{n}}+1}^{2j_N p_{\mathbf{n}}+p_{\mathbf{n}}} \Delta_i, \\
U(2^N, \mathbf{n}, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+p_{\mathbf{n}}+1 \\ k=1,\dots,N}}^{2(j_k+1)p_{\mathbf{n}}} \Delta_i
\end{aligned} \tag{10}$$

with $p_{\mathbf{n}}$ is a real sequence will be specified later.

Now, we put for all $i = 1, \dots, 2^N$,

$$T(\mathbf{n}, i) = \sum_{\mathbf{j} \in \mathcal{J}} U(i, \mathbf{n}, \mathbf{j}). \tag{11}$$

with $\mathcal{J} = \{0, \dots, r_1 - 1\} \times \dots \times \{0, \dots, r_N - 1\}$ and $r_l = 2n_l p_{\mathbf{n}}^{-1}$; $l = 1, \dots, N$.

Then,

$$|\tilde{g}_l^*(x_k) - E[\tilde{g}_l^*(x_k)]| = \frac{1}{\widehat{\mathbf{n}} h^d} \sum_{i=1}^{2^N} T(\mathbf{n}, i).$$

Thus, all it remains to compute

$$\mathbb{P}(T(\mathbf{n}, i) \geq \eta \widehat{\mathbf{n}} h^d), \quad \text{for all } i = 1, \dots, 2^N. \tag{12}$$

Without loss of generality, we will only consider the case $i = 1$. For this, we enumerate the $M = \prod_{k=1}^N r_k = 2^{-N} \hat{\mathbf{n}} p_{\mathbf{n}}^{-N} \leq \hat{\mathbf{n}} p_{\mathbf{n}}^{-N}$ random variables $U(1, \mathbf{n}, \mathbf{j}) ; \mathbf{j} \in \mathcal{J}$ in the arbitrary way Z_1, \dots, Z_M .

The rest of the proof is very similar to Biau and Cader (2004) which is based on Lemma 4.5 of Carbon *et al* (1997). According this Lemma we can find independent random variables Z_1^*, \dots, Z_M^* has the same law as $Z_{j=1, \dots, M}$ and such that

$$\sum_{j=1}^r E|Z_j - Z_j^*| \leq 2C\mu_{\mathbf{n}} M p_{\mathbf{n}}^N s(M-1)p_{\mathbf{n}}^N, p_{\mathbf{n}}^N) \varphi(p_{\mathbf{n}}).$$

It follows that

$$\mathbb{P}(T(\mathbf{n}, 1) \geq \eta \hat{\mathbf{n}} h^d) \leq \mathbb{P}\left(\left|\sum_{j=1}^M Z_j^*\right| \geq \frac{\eta \hat{\mathbf{n}} h^d}{2}\right) + \mathbb{P}\left(\sum_{j=1}^M |Z_j - Z_j^*| \geq \frac{\eta \hat{\mathbf{n}} h^d}{2}\right).$$

Thus, from the Bernstein and Markov inequalities we deduce that

$$B_1 := \mathbb{P}\left(\left|\sum_{j=1}^M Z_j^*\right| \geq \frac{M \eta \hat{\mathbf{n}} h^d}{2M}\right) \leq 2 \exp\left(-\frac{(\eta \hat{\mathbf{n}} h^d)^2}{MVar(Z_1^*) + C p_{\mathbf{n}}^N \eta \hat{\mathbf{n}} h^d}\right)$$

and

$$B_2 := \mathbb{P}\left(\sum_{j=1}^M |Z_j - Z_j^*| \geq \frac{\eta \hat{\mathbf{n}} h^d}{2}\right) \leq \frac{2}{\eta \hat{\mathbf{n}} h^d} \sum_{j=1}^M E|Z_j - Z_j^*|.$$

By using Lemma 4.5 of Carbon et al. (2007), the fact that $\hat{\mathbf{n}} = 2^N M p_{\mathbf{n}}^N$ and $s((M-1)p_{\mathbf{n}}^N, p_{\mathbf{n}}^N) \leq p_{\mathbf{n}}^N$ we get for $\eta = \eta_0 \sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h^d}}$

$$B_2 \leq \mu_{\mathbf{n}} \hat{\mathbf{n}} p_{\mathbf{n}}^N (\log \hat{\mathbf{n}})^{-1/2} (\hat{\mathbf{n}} h^d)^{-1/2} \varphi(p_{\mathbf{n}}).$$

Since $p_{\mathbf{n}} = C \left(\frac{\hat{\mathbf{n}} h^d}{\log \hat{\mathbf{n}} \mu_{\mathbf{n}}^2}\right)^{1/2N}$, then

$$B_2 \leq \hat{\mathbf{n}} (\log \hat{\mathbf{n}})^{-1} \varphi(p_{\mathbf{n}}).$$

Concerning B_1 term, by a standard arguments we obtain

$$\text{Var}[Z_1^*] = O(p_{\mathbf{n}}^N h^d).$$

Using this last result, together with the definitions of $p_{\mathbf{n}}$, M and η , we get

$$B_1 \leq \exp(-C(\eta_0) \log \hat{\mathbf{n}})$$

Consequently, from (H6), we have

$$\exists \eta_0 \text{ such that } d_{\mathbf{n}} \sum_{\mathbf{n}} (B_1 + B_2) < \infty.$$

which complete the first result of this lemma.

- Now, we evaluate terms T_1 and T_3 : To do that, we use the Lipschitz's condition of the kernel K in (H4) allows to write directly,

$$\begin{aligned} |\tilde{g}_l^*(x) - \tilde{g}_l^*(x_{k(x)})| &= \frac{1}{\hat{\mathbf{n}} h^d} \left| \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-l} K_{\mathbf{i}}(x) - \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-l} K_{\mathbf{i}}(x_{k(x)}) \right| \\ &\leq \frac{C}{\hat{\mathbf{n}} h^{d+1}} \|x - x_{k(x)}\| \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-l} \\ &\leq \frac{C \tau_{\mathbf{n}}}{\mu_{\mathbf{n}}^l \hat{\mathbf{n}} h^{d+1}} \leq \frac{C \tau_{\mathbf{n}}}{\mu_{\mathbf{n}} \hat{\mathbf{n}} h^{d+1}}. \end{aligned}$$

By the definition of $\tau_{\mathbf{n}}$ we obtain

$$\sup_{x \in S} |\tilde{g}_l^*(x) - \tilde{g}_l^*(x_{k(x)})| = O\left(\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h^d}}\right) \quad (13)$$

and

$$\sup_{x \in S} \left| E[\tilde{g}_l^*(x)] - E[\tilde{g}_l^*(x_{k(x)})] \right| = O\left(\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} h^d}}\right) \quad (14)$$

— Secondly, we proof (7). Indeed, we have

$$\begin{aligned} \sup_{x \in S} |E[\tilde{g}_l(x)] - E[\tilde{g}_l^*(x)]| &= \frac{1}{\hat{\mathbf{n}}h^d} \left| E \left[\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-l} \mathbb{1}_{\{|Y_{\mathbf{i}}^{-1}| > \mu_{\mathbf{n}}\}} K_{\mathbf{i}}(x) \right] \right| \\ &\leq h^{-d} E \left[|Y_1^{-l}| \mathbb{1}_{\{|Y_1^{-1}| > \mu_n\}} K_1(x) \right] \\ &\leq h^{-d} E \left[\exp(|Y_1^{-l}|/4) \mathbb{1}_{\{|Y_1^{-1}| > \mu_n\}} K_1(x) \right]. \end{aligned}$$

Furthermore, using The Holder's inequality to show that,

$$\begin{aligned} \sup_{x \in S} |E[\tilde{g}_l(x)] - E[\tilde{g}_l^*(x)]| &\leq h^{-d} \left(E \left[\exp(|Y_1^{-l}|/2) \mathbb{1}_{\{|Y_1^{-1}| > \mu_n\}} \right] \right)^{\frac{1}{2}} (E(K_1^2(x)))^{\frac{1}{2}} \\ &\leq h^{-d} \exp(-\mu_{\mathbf{n}}^l/4) (E[\exp(|Y^{-l}|)])^{\frac{1}{2}} (E(K_1^2(x)))^{\frac{1}{2}} \\ &\leq Ch^{\frac{-d}{2}} \exp(-\mu_{\mathbf{n}}^l/4). \end{aligned}$$

Since $\mu_{\mathbf{n}} = \hat{\mathbf{n}}^{\gamma/2}$ then, we can write

$$\sup_{x \in S} |E[\tilde{g}_l(x_k)] - E[\tilde{g}_l^*(x_k)]| = o \left(\left(\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}}h^d} \right)^{1/2} \right).$$

— Thirdly, the proof of the last claimed result (8) is based on the Markov's inequality. Indeed, observe that, for all $\epsilon > 0$

$$\begin{aligned} \mathbb{P} \left[\sup_{x \in S} |\tilde{g}_l(x) - \tilde{g}_l^*(x)| > \epsilon \right] &= \mathbb{P} \left(\frac{1}{\hat{\mathbf{n}}h^d} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-l} \mathbb{1}_{\{|Y_{\mathbf{i}}^{-1}| > \mu_{\mathbf{n}}^l\}} K_i | > \epsilon \right) \\ &\leq \hat{\mathbf{n}} \mathbb{P} (|Y^{-l}| > \mu_{\mathbf{n}}) \\ &\leq \hat{\mathbf{n}} \exp(-\mu_{\mathbf{n}}^l) E(\exp(|Y^{-1}|)) \\ &\leq C \hat{\mathbf{n}} \exp(-\mu_{\mathbf{n}}^l). \end{aligned}$$

So,

$$\sum_{\mathbf{n}} \mathbb{P} \left(\sup_{x \in S} |\tilde{g}_l(x) - \tilde{g}_l^*(x)| > \epsilon_0 \left(\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}}h^d}} \right) \right) \leq C \sum_{\mathbf{n}} \hat{\mathbf{n}} \exp(-\mu_{\mathbf{n}}). \quad (15)$$

The use of the definition $\mu_{\mathbf{n}}$ complete the proof of Lemma.

■

Corollary 2 Under the hypotheses of Theorem 3.2.1, we obtain :

$$\sum_{\mathbf{n}} \mathbb{P} \left(\inf_{x \in S} |\tilde{g}_2(x)| \leq \frac{\inf_{x \in S} g_2(x)}{2} \right) < \infty.$$

Proof of Corollary 2. It is clear that

$$\inf_{x \in S} \tilde{g}_2(x) < \frac{\inf_{x \in S} g_2(x)}{2} \Rightarrow \sup_{x \in S} |g_2(x) - \tilde{g}_2(x)| > \frac{\inf_{x \in S} g_2(x)}{2}.$$

Thus,

$$\mathbb{P} \left(\inf_{x \in S} |\tilde{g}_2(x)| \leq \frac{\inf_{x \in S} g_2(x)}{2} \right) \leq \mathbb{P} \left(\sup_{x \in S} |g_2(x) - \tilde{g}_2(x)| \geq \frac{\inf_{x \in S} g_2(x)}{2} \right).$$

The use of the results of Lemma 1 and Lemma 2 complete the proof of the corollary. ■

Proof of Theorem 2.3.2. We write :

$$\tilde{\theta}(x) - \theta(x) = \frac{1}{\tilde{g}_2(x)} [B_{\mathbf{n}} + V_{\mathbf{n}} (\tilde{g}_2(x) - E\tilde{g}_2(x))] + V_{\mathbf{n}}$$

where

$$V_{\mathbf{n}} = \frac{1}{E\tilde{g}_2(x)g_2(x)} \left[[E\tilde{g}_1(x)]g_2(x) - [E\tilde{g}_2(x)]g_1(x) \right]$$

and

$$B_{\mathbf{n}} = \frac{1}{g_2(x)} \left[[\tilde{g}_1(x) - E\tilde{g}_1(x)]g_2(x) + [E\tilde{g}_2(x) - \tilde{g}_2(x)]g_1(x) \right].$$

Therefore,

$$\tilde{\theta}(x) - \theta(x) - V_{\mathbf{n}} = \frac{1}{\tilde{g}_2(x)} [B_{\mathbf{n}} + V_{\mathbf{n}} (\tilde{g}_2(x) - E\tilde{g}_2(x))] \quad (16)$$

Finally, Theorem 2.3.2 is a consequence of the following results .

LEMMA 3 *Under the hypotheses of Theorem 2.3.2, we obtain :*

$$\left(\frac{\widehat{\mathbf{n}}h^d}{g_2^2(x)\sigma^2(x)} \right)^{1/2} (B_{\mathbf{n}} - \mathbb{E}[B_{\mathbf{n}}]) \rightarrow N(0, 1).$$

Proof of Lemma 3. Considering the same notations of Lemma 1 and write,

$$B_{\mathbf{n}} = B_{\mathbf{n}} - B_{\mathbf{n}}^* + B_{\mathbf{n}}^*$$

where

$$B_{\mathbf{n}}^* = \frac{1}{g_2(x)} \left[\left[\widetilde{g}_1^*(x) - E\widetilde{g}_1^*(x) \right] g_2(x) + \left[E\widetilde{g}_2^*(x) - \widetilde{g}_2^*(x) \right] g_1(x) \right].$$

Similarly to Lemma 1 we get, for fixed $x \in \mathbb{R}^d$

$$\left(\frac{\widehat{\mathbf{n}}h^d}{(g_2(x))^2 \sigma^2(x)} \right)^{1/2} |\widehat{g}_l(x) - \widehat{g}_l^*(x)| = o_p(1).$$

As $\mathbb{E}[B_{\mathbf{n}}] = \mathbb{E}[B_{\mathbf{n}}^*] = 0$, then it suffices to show the asymptotic normality of

$$\left(\frac{\widehat{\mathbf{n}}h^d}{(g_2(x))^2 \sigma^2(x)} \right)^{1/2} |B_{\mathbf{n}}^* - E[B_{\mathbf{n}}^*]|.$$

For this, we put, for $\mathbf{i} = \mathbf{1} \in I_{\mathbf{n}}$,

$$\Lambda_{\mathbf{i}} := \frac{1}{\sqrt{h^d}} \left(K_i(Y_{\mathbf{i}}^{-1}g_2(x) - Y_{\mathbf{i}}^{-2}g_1(x)) \mathbb{1}_{|Y_{\mathbf{i}}^{-1}| < \mu_{\mathbf{n}}} - E \left[K_i(Y_{\mathbf{i}}^{-1}g_2(x) - Y_{\mathbf{i}}^{-2}g_1(x)) \mathbb{1}_{|Y_{\mathbf{i}}^{-1}| < \mu_{\mathbf{n}}} \right] \right).$$

So,

$$\sqrt{\widehat{\mathbf{n}}h^d} [\sigma_1(x)]^{-1} \left(B_{\mathbf{n}}^* - E[B_{\mathbf{n}}^*] \right) = (\widehat{\mathbf{n}}\sigma_1^2(x))^{-1/2} S_{\mathbf{n}}$$

where $S_{\mathbf{n}} = \sum_{\mathbf{i} \in I_{\mathbf{n}}} \Lambda_{\mathbf{i}}$. Thus, our claimed result is, now

$$(\widehat{\mathbf{n}}\sigma_1^2(x))^{-1/2} S_{\mathbf{n}} \rightarrow \mathcal{N}(0, 1). \quad (17)$$

where $\sigma_1^2(x) = (g_2(x))^2 \sigma^2(x)$. The proof of (11) follows the same lines of Lemma 3.2 in Tarn (1990). It is based on spatial blocking technique for $S_{\mathbf{n}} = \sum_{\mathbf{i} \in I_{\mathbf{n}}} \Lambda_{\mathbf{i}}$.

LEMMA 4 *Under the hypotheses of Theorem 2.3.2, we obtain :*

$$\tilde{g}_2(x) \rightarrow g_2(x), \text{ in probability,}$$

$$\left(\frac{\hat{\mathbf{n}}h^d}{g_2(x)^2\sigma^2(x)} \right)^{1/2} V_{\mathbf{n}} \rightarrow 0,$$

and

$$\left(\frac{\hat{\mathbf{n}}h^d}{g_2(x)^2\sigma^2(x)} \right)^{1/2} V_{\mathbf{n}} (\tilde{g}_2(x) - E\tilde{g}_2(x)) \rightarrow 0, \text{ in probability.}$$

Proof of Lemma 4. For the first limit, we have, by Lemma 1

$$\mathbb{E} [\tilde{g}_2(x) - g_2(x)] \rightarrow 0$$

and by a similar argument as those used in the variance term in Lemma 2 we show that

$$Var [\tilde{g}_2(x)] \rightarrow 0$$

hence

$$\tilde{g}_2(x) - g_2(x) \rightarrow 0 \quad \text{in probability.}$$

Next, it is clear that the second limit is consequence for the last convergence. So, it suffices to treat the last one. For this, we use the fact that

$$Var [(\tilde{g}_2(x) - \mathbb{E}\tilde{g}_2(x))] = Var [\tilde{g}_2(x)] \rightarrow 0$$

and

$$V_{\mathbf{n}} = O(h^2) \quad (\text{see, Lemma 1})$$

The last part of Condition (H6') allows to deduce that

$$\left(\frac{\hat{\mathbf{n}}h^d}{g_1(x)^2\sigma^2(x)} \right)^{1/2} V_{\mathbf{n}} (\tilde{g}_2(x) - \mathbb{E}\tilde{g}_2(x)) \rightarrow 0 \quad \text{in probability.}$$

Acknowledgements

The authors greatly thank an Associate Editor and an anonymous referee for a careful reading of the paper.

References

- Bernhard F. A. and Stahlecker P. (2003) Relative squared error prediction in the generalized linear regression model. *Statist. Papers* **44**, 107-115.
- Bobbia M., Misiti M., Misiti Y., Poggi J.M. and Portier B. (2015) Spatial outlier detection in the PM10 monitoring network of Normandy. *Atmospheric Pollution Research* **6** , 476-483.
- Biau G. and Cadre B. (2004) Nonparametric spatial prediction. *Statist. Inference Stoch. Processes* **7**, 327-349.
- Carbon M., Tran L. T. and Wu, B. (1997) Kernel density estimation for random fields. *Statist. Probab. Lett.* **36**, 115-125.
- Carbon M., Francq C. and Tran L.T. (2007) Kernel regression estimation for random fields. *J. Statist. Plann. Inference* **137**, 778-798.
- Cressie N.A. *Statistics for spatial data*. Wiley, New York, 1993.
- Dabo-Niang S. and Thiam B. (2010). Robust quantile estimation and prediction for spatial processes. *Statist. Probab. Lett.* **80**, 1447-1458.
- Dabo-Niang S. and Yao A. F. (2007) Kernel regression estimation for continuous spatial processes. *Math. Method. Stat.* **16**, 1-20.
- Dabo-Niang S., Yao A., Pischedda L., Cuny P. and Gilbert F. (2009) Spatial

kernel mode estimation for functional random, with application to bioturbation problem. *Stochastic Environmental Research and Risk Assessment* **24**, 487-497.

Dabo-Niang S. Ould-Abdi S., Ould-Abdi, A., Diop, A. (2014) Consistency of a nonparametric conditional mode estimator for random fields. *Stat. Methods Appl.* **23** , 1-39.

Doukhan P., *Mixing : Properties and examples*, Lect. Notes in Statist., 85, Springer- Verlag, New York, 1994.

Diggle P. and Ribeiro P.J. (2007) *Model-Based Geostatistics*. Springer, New York, 2007.

El Machkouri M. and Stoica R. (2010) Asymptotic normality of kernel estimates in a regression model for random fields. *J. Nonparametric Statist.* **22** 955–971.

Filzmoser P., Ruiz-Gazen A. and Thomas-Agnan C. (2014) Identification of local multivariate outliers. *Statist. Papers* **55**, 29-47

Gheriballah A., Laksaci A. and Rouane, R. (2010). Robust nonparametric estimation for spatial regression, *J. Statist. Plann. Inference* **140**, 1656-1670.

Guyon, X. Estimation d'un champ par pseudo-vraisemblance conditionnelle : Etude asymptotique et application au cas Markovien. *Proceedings of the Sixth Franco-Belgian Meeting of Statisticians*, 1987.

Hallin M., Lu Z. and Yu K. (2009) Local linear spatial quantile regression. *Bernoulli* **15**, 659–686.

- Li J. and Tran L.T. (2009) Nonparametric estimation of conditional expectation. *J. Statist. Plann. Inference* **139**, 164-175.
- Liu X., Lu C.T. and Chen F. : Spatial outlier detection : Random walk based approaches. *Proceedings of the 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), San Jose, CA* , 2010.
- Lu Z. and Chen X. (2004) Spatial kernel regression : weak consistency. *Statist. Probab. Lett.* **68**, 125-136.
- Martínez J., Saavedra Á., García-Nieto P. J., Piñeiro J. I., Iglesias C., Taboada J., Sancho J. and Pastor J. (2014) Air quality parameters outliers detection using functional data analysis in the Langreo urban area (Northern Spain). *Appl. Math. Comput.* **241**, 1–10.
- Narula S.C. and Wellington, J.F. (1977) Prediction, linear regression and the minimum sum of relative errors. *Technometrics* **19**, 185–190.
- Omidi M. and Mohammadzadeh, M. (2015) A new method to build spatio-temporal covariance functions : analysis of ozone data. *Statist. Papers*, in presse DOI10.1007/s00362-015-0674-2
- Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *J. of Econometrics*, **165**, 5-19.
- Jones M. C., Park H., Shinb K. Vines S. K. and Jeong S.O. (2008) Relative error prediction via kernel regression smoothers. *J. Statist. Plann. Inference* **138**, 2887-2898

- Shen V.Y., Yu T. and Thebaut, S.M. (1985) Identifying error-prone software: An empirical study. *IEEE Trans. Software Eng.* **11**, 317–324.
- Yang, Y. and Ye, F. (2013) General relative error criterion and M-estimation. *Front. Math. China* **8**, 695–715.
- Tran, L.T. (1990). Kernel density estimation on random fields. *J. of Mult. Anal.* **34**, 37–53.
- Volker S. *Stochastic Geometry, Spatial Statistics and Random Fields : Models and Algorithms*. **2120**, Lecture Notes in Mathematics, Springer, 2014.
- Xu R. and Wang J. (2008) L_1 - estimation for spatial nonparametric regression. *J. Nonparametric Statist.* **20**, 523–537.

Chapitre 3

Cas fonctionnel

3.1 Construction of the spatial estimate

Consider $Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}})$, $\mathbf{i} \in \mathbb{N}^N$ be a $\mathcal{F} \times \mathbb{R}$ -valued measurable strictly stationary spatial process, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where \mathcal{F} is a semi-metric space, d denoting the semi-metric. We assume that the process under study $(Z_{\mathbf{i}})$ is observed over a rectangular domain $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} = (i_1, \dots, i_N) \in \mathbb{N}^N, 1 \leq i_k \leq n_k, k = 1, \dots, N\}$, $\mathbf{n} = (n_1, \dots, n_N) \in \mathbb{N}^N$. A point \mathbf{i} will be referred to as a *site*. We will write $\mathbf{n} \rightarrow \infty$ if $\min\{n_k\} \rightarrow \infty$ and $|\frac{n_j}{n_k}| < C$ for a constant C such that $0 < C < \infty$ for all j, k such that $1 \leq j, k \leq N$. For $\mathbf{n} = (n_1, \dots, n_N) \in \mathbb{N}^N$, we set $\hat{\mathbf{n}} = n_1 \times \dots \times n_N$. The nonparametric model studied in this paper, denoted by θ_x , is defined, by

$$\theta(x) = \frac{\mathbb{E}[Y^{-1}|X=x]}{\mathbb{E}[Y^{-2}|X=x]}.$$

Thus, a natural estimate of θ is

$$\tilde{\theta}(x) = \frac{\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-1} K(h^{-1}d(x, X_{\mathbf{i}}))}{\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}}^{-2} K(h^{-1}d(x, X_{\mathbf{i}}))} \quad (1)$$

where K is a kernel and $h = h_{K,n}$ is a sequence of positive real numbers. Our main goal is to study the asymptotic proprieties (consistency and asymptotic normality) of the nonparametric estimate $\tilde{\theta}$ of θ when the random field $(Z_i, i \in \mathbb{N}^N)$ satisfies the following mixing condition :

$$\left\{ \begin{array}{l} \text{There exists a function } \varphi(t) \downarrow 0 \text{ as } t \rightarrow \infty, \text{ such that} \\ \forall E, E' \text{ subsets of } \mathbb{N}^N \text{ with finite cardinals} \\ \alpha(\mathcal{B}(E), \mathcal{B}(E')) = \sup_{B \in \mathcal{B}(E), C \in \mathcal{B}(E')} |\mathbb{P}(B \cap C) - \mathbb{P}(B)\mathbb{P}(C)| \\ \leq s(\text{Card}(E), \text{Card}(E')) \varphi(\text{dist}(E, E')), \end{array} \right. \quad (2)$$

where $\mathcal{B}(E)$ (*resp.* $\mathcal{B}(E')$) denotes the Borel σ -field generated by $(Z_i, i \in E)$ (*resp.* $(Z_i, i \in E')$), $\text{Card}(E)$ (*resp.* $\text{Card}(E')$) the cardinality of E (*resp.* E'), $\text{dist}(E, E')$ the Euclidean distance between E and E' and $s : \mathbb{N}^2 \rightarrow \mathbb{R}^+$ is a symmetric positive function nondecreasing in each variable such that :

$$s(n, m) \leq C \min(n, m), \quad \forall n, m \in \mathbb{N}. \quad (3)$$

We point out that condition (3) is used in Tran (1990), Carbon *et al* (1996) and is satisfied by a large class of spatial models (see Guyon (1987) for some examples). Noting that if $N = 1$, then Z_i is called strongly mixing (see Doukhan *et al.* (1994) for discussion on mixing and examples).

3.2 Main results

Throughout the paper, when no confusion will be possible, we will denote by C or/and C' some strictly positive generic constants. Furthermore, for (X, Y) a couple of random variables with same distribution as (X_i, Y_i) , we

denote by $g_l(\cdot) = E[Y^{-l}|X = \cdot]$ the conditional l -inverse moments of Y given X ; $l = 1, 2, 3, 4$

3.2.1 Consistency

Our first goal in this section, is to establish the uniform almost complete convergence¹ (a.co.) of $\tilde{\theta}(x)$ to $\theta(x)$ over compact S . To do that, we consider the following assumptions :

(H1) $\mathbb{P}(X \in \mathcal{B}(x, r)) = \phi_x(r) > 0$ où $\mathcal{B}(x, r)$ est la boule centrée en x et de rayon r .

(H2) The functions $g_l(x) = E[Y^{-l}|X = x]$, ($l = 1, 2$) such that

$$\forall x_1, x_2 \in \mathcal{N}_x, \quad |r_l(x_1) - r_l(x_2)| \leq C d^b(x_1, x_2)$$

where \mathcal{N}_x is a fixed neighborhood of x .

(H3) $\forall \mathbf{i} \neq \mathbf{j}$,

$$0 < \sup_{\mathbf{i} \neq \mathbf{j}} \mathbb{P}[(X_{\mathbf{i}}, X_{\mathbf{j}}) \in B(x, h) \times B(x, h)] \leq C(\phi_x(h))^{(a+1)/a}, \text{ for some } 1 < a < \delta N^{-1}.$$

(H4) The mixing coefficient defined in (2) satisfies

$$\sum_{i=1}^{\infty} i^{N-1} (\varphi(i))^a < \infty, \quad \text{for some } 0 < a < \frac{1}{2}.$$

1. Let $(z_n)_{n \in \mathbb{N}}$ be a sequence of real r.v.'s. We say that z_n converges almost completely (a.co.) toward zero if, and only if, $\forall \epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon) < \infty$. Moreover, we say that the rate of the almost complete convergence of z_n to zero is of order u_n (with $u_n \rightarrow 0$) and we write $z_n = O_{a.co.}(u_n)$ if, and only if, $\exists \epsilon > 0$ such that $\sum_{n=1}^{\infty} P(|z_n| > \epsilon u_n) < \infty$. This kind of convergence implies both almost sure convergence and convergence in probability.

(H6) K is a function with support $[0, 1]$ such that

$$C\mathbb{I}_{(0,1)}(\cdot) \leq K(\cdot) \leq C'\mathbb{I}_{(0,1)}(\cdot).$$

(H7) There exists $\eta_0 > 0$, such that, $C\widehat{\mathbf{n}}^{\frac{4N-\delta}{\delta}+\eta_0} \leq \phi_x(h)$.

(H8) $\exists C > 0$, such that $Y > C$ a.s.

We precise that our conditions are very usual in this context of nonparametric spatial analysis (see Tran(1990) or Li and Tran (2007)). Furthermore, Assumptions (H1) and (H2) are regularity condition permit to evaluate the bias term of the estimator (1). Condition (H3) is automatically fulfilled when the densities $f_{i,j}$ and f are bounded. While assumptions (H4)-(H7) are technical conditions imposed in order to attaint the brevity of proofs and to obtain a convergence rate comparable to the i.i.d case.

Theorem 3.2.1 *Under assumptions (H1)-(H7), we have :*

$$|\tilde{\theta}(x) - \theta(x)| = O(h^b) + O\left(\sqrt{\frac{\log \widehat{\mathbf{n}}}{\widehat{\mathbf{n}} \phi_x(h)}}\right) \quad a.co. \quad \text{as } \mathbf{n} \rightarrow \infty \quad (4)$$

Proof of Theorem 3.2.1 Let

$$\tilde{\theta}(x) = \frac{\tilde{g}_1(x)}{\tilde{g}_2(x)}$$

with

$$\tilde{g}_l(x) = \frac{1}{\widehat{\mathbf{n}}\mathbb{E}[K(h^{-1}d(x, X_i))]} \sum_{i \in \mathcal{I}_n} Y_i^{-l} K(h^{-1}d(x, X_i)) \quad \text{for } l = 1, 2.$$

Next, we use the following decomposition :

$$\tilde{\theta}(x) - \theta(x) = \frac{1}{\tilde{g}_2(x)} [\tilde{g}_1(x) - g_1(x)] + [g_2(x) - \tilde{g}_2(x)] \frac{\theta(x)}{\tilde{g}_2(x)} \quad (5)$$

where Thus, Theorem 3.2.1 is a consequence of the following intermediate results

LEMMA 5 *Under hypotheses (H1), (H2), (H4) and (H5), we have, for $l = 1, 2$, that :*

$$|E\tilde{g}_l(x) - g_l(x)| = O(h^b).$$

Proof of Lemma 5. By a change of variables, we get, for $l = 1, 2$,

$$\begin{aligned} \mathbb{E}[\tilde{g}_l(x)] - g_l(x) &= \mathbb{E}\left[\frac{1}{\hat{\mathbf{n}}\mathbb{E}[K(h^{-1}d(x, X_{\mathbf{i}}))]}Y^{-l}K(h^{-1}d(x, X_{\mathbf{i}}))\right] - g_l(x) \\ &= \frac{1}{\hat{\mathbf{n}}\mathbb{E}[K(h^{-1}d(x, X_{\mathbf{i}}))]} \mathbb{E}[K(h^{-1}d(x, X_{\mathbf{i}})Y^{-l})] - g_l(x) \\ &= \frac{1}{\hat{\mathbf{n}}\mathbb{E}[K(h^{-1}d(x, X_{\mathbf{i}}))]} \mathbb{E}[K(h^{-1}d(x, X_{\mathbf{i}})Y^{-l})\mathbb{1}_{B(x, h_k)}(X)\mathbb{E}(Y^{-l}/X)] - g_l(x) \end{aligned}$$

As

$$\mathbb{1}_{B(x, h_k)}(X)|g_l(X) - g_l(x)| \leq Ch^b.$$

The last result complete the proof of lemma. ■

LEMMA 6 *Under hypotheses (H3)-(H7), we have, for $l = 1, 2$, that :*

$$|\tilde{g}_l(x) - \mathbb{E}\tilde{g}_l(x)| = O\left(\sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}}\phi_x(h)}}\right) \quad a.co. \quad \text{as } \mathbf{n} \rightarrow \infty \quad (6)$$

Proof of Lemma 6. Consider

$$\Delta_{\mathbf{i}} = Y_{\mathbf{i}}^{-l}K(h^{-1}d(x_k, X_{\mathbf{i}})) - \mathbb{E}[Y^{-l}K(h^{-1}d(x_k, X))]$$

then

$$\widehat{r}_l(x) - \mathbb{E}[\widehat{r}_l(x)] = \frac{1}{\hat{\mathbf{n}}\mathbb{E}[K_{\mathbf{1}}(x)]} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \Delta_{\mathbf{i}}(x).$$

Consider the spatial block decomposition (Tran [1990]) on the random's variables $\Delta_i(x)$ for fixed integers $p_{\mathbf{n}}$, as follows

$$\begin{aligned} U(1, \mathbf{n}, x, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+1 \\ k=1,\dots,N}}^{2j_k p_{\mathbf{n}}+p_{\mathbf{n}}} \Delta_i(x), \\ U(2, \mathbf{n}, x, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+1 \\ k=1,\dots,N-1}}^{2j_k p_{\mathbf{n}}+p_{\mathbf{n}}} \sum_{i_N=2j_N p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{(j_N+1)p_{\mathbf{n}}} \Delta_i(x), \\ U(3, \mathbf{n}, x, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+1 \\ k=1,\dots,N-2}}^{2j_k p_{\mathbf{n}}+p_{\mathbf{n}}} \sum_{i_{N-1}=2j_{N-1} p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{2(j_{N-1}+1)p_{\mathbf{n}}} \sum_{i_N=2j_N p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{2j_N p_{\mathbf{n}}+p_{\mathbf{n}}} \Delta_i(x), \\ U(4, \mathbf{n}, x, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+1 \\ k=1,\dots,N-2}}^{2j_k p_{\mathbf{n}}} \sum_{i_{N-1}=2j_{N-1} p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{2(j_{N-1}+1)p_{\mathbf{n}}} \sum_{i_N=2j_N p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{2(j_N+1)p_{\mathbf{n}}} \Delta_i(x), \end{aligned}$$

and so on. Finally

$$\begin{aligned} U(2^{N-1}, \mathbf{n}, x, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+p_{\mathbf{n}}+1 \\ k=1,\dots,N-1}}^{2(j_k+1)p_{\mathbf{n}}} \sum_{i_N=2j_N p_{\mathbf{n}}+p_{\mathbf{n}}+1}^{2j_N p_{\mathbf{n}}+p_{\mathbf{n}}} \Delta_i(x), \\ U(2^N, \mathbf{n}, x, \mathbf{j}) &= \sum_{\substack{i_k=2j_k p_{\mathbf{n}}+p_{\mathbf{n}}+1 \\ k=1,\dots,N}}^{2(j_k+1)p_{\mathbf{n}}} \Delta_i(x) \end{aligned} \tag{7}$$

Now, for $\mathcal{J} = \{0, \dots, r_1 - 1\} \times \dots \times \{0, \dots, r_N - 1\}$ where $r_l = n_l / 2p_{\mathbf{n}}$, $l = 1, \dots, N$. For $i = 1, \dots, 2^N$ we define

$$T(\mathbf{n}, x, i) = \sum_{\mathbf{j} \in \mathcal{J}} U(i, \mathbf{n}, x, \mathbf{j})$$

and we write,

$$\widehat{r}_l(x) - \mathbb{E}[\widehat{\Psi}_D(x)] = \frac{1}{\mathbf{n} \mathbb{E}[K_1(x)]} \sum_{i=1}^{2^N} T(\mathbf{n}, x, i). \tag{8}$$

Note that, as raises by Biau and Cadre (2004), if one does not have the equalities $n_{\mathbf{i}} = 2r_{\mathbf{i}} p_{\mathbf{n}}$, the term say $T(\mathbf{n}, x, 2^N + 1)$ (which contains the $\Delta_{\mathbf{i}}(x)$'s at the ends not included in the blocks above) can be added. This will not change the proof a lot. Using (8) we can say that, for all $\eta > 0$

$$\mathbb{P} (|\widehat{r}_l(x) - \mathbb{E}[\widehat{r}_l(x)]| \geq \eta) \leq 2^N \max_{i=1, \dots, 2^N} \mathbb{P} (T(\mathbf{n}, x, i) \geq \eta \widehat{\mathbf{n}} \mathbb{E} [K_1(x)]).$$

Finally, the desired result follow from the evaluation of the following quantities it suffices to compute

$$\mathbb{P} (T(\mathbf{n}, x, i) \geq \eta \widehat{\mathbf{n}} \mathbb{E} [K_1(x)]) , \quad \text{for all } i = 1, \dots, 2^N.$$

Without loss of generality, we will only consider the case $i = 1$. For this case, we enumerate the $M = \prod_{k=1}^N r_k = 2^{-N} \widehat{\mathbf{n}} p_{\mathbf{n}}^{-N} \leq \widehat{\mathbf{n}} p_{\mathbf{n}}^{-N}$ random variables $U(1, \mathbf{n}, x, \mathbf{j}) ; \mathbf{j} \in \mathcal{J}$ in the arbitrary way Z_1, \dots, Z_M . Thus, for each Z_j there exists a certain \mathbf{j} in \mathcal{J} such that

$$Z_j = \sum_{\mathbf{i} \in I(1, \mathbf{n}, x, \mathbf{j})} \Delta_{\mathbf{i}}(x)$$

where $I(1, \mathbf{n}, x, \mathbf{j}) = \{\mathbf{i} : 2j_k p_{\mathbf{n}} + 1 \leq i_k \leq 2j_k p_{\mathbf{n}} + p_{\mathbf{n}} \quad ; k = 1, \dots, N\}$. Clearly the sets $I(1, \mathbf{n}, x, \mathbf{j})$ contains $p_{\mathbf{n}}^N$ sites and are far apart by distant of $p_{\mathbf{n}}$ at least. So, according to Lemma 4.5 in [?] one can find independent random variables Z_1^*, \dots, Z_M^* having the same law as $Z_{j=1, \dots, M}$ and such that

$$\sum_{j=1}^M \mathbb{E}|Z_j - Z_j^*| \leq 2CM p_{\mathbf{n}}^N s(M-1) p_{\mathbf{n}}^N \varphi(p_{\mathbf{n}}).$$

Therefore, by the Bernstein and Markov inequalities we have

$$\mathbb{P} (T(\mathbf{n}, x, i) \geq \eta \widehat{\mathbf{n}} \mathbb{E} [K_1(x)]) \leq B_1 + B_2$$

where

$$B_1 = \mathbb{P} \left(\left| \sum_{j=1}^M Z_j^* \right| \geq \frac{M\eta \hat{\mathbf{n}} \mathbb{E}[K_1(x)]}{2M} \right) \leq 2 \exp \left(- \frac{(\eta \hat{\mathbf{n}} \mathbb{E}[K_1(x)])^2}{MVar[Z_1^*] + Cp_{\mathbf{n}}^N \eta \hat{\mathbf{n}} \mathbb{E}[K_1(x)]} \right)$$

and

$$\begin{aligned} B_2 &= \mathbb{P} \left(\sum_{j=1}^M |Z_j - Z_j^*| \geq \frac{\eta \hat{\mathbf{n}} \mathbb{E}[K_1(x)]}{2} \right) \\ &\leq \frac{1}{\eta \hat{\mathbf{n}} \mathbb{E}[K_1(x)]} \sum_{j=1}^M \mathbb{E}|Z_j - Z_j^*| \\ &\leq 2Mp_{\mathbf{n}}^N (\eta \hat{\mathbf{n}} \mathbb{E}[K_1(x)])^{-1} s((M-1)p_{\mathbf{n}}^N, p_{\mathbf{n}}^N) \varphi(p_{\mathbf{n}}). \end{aligned}$$

Since $\hat{\mathbf{n}} = 2^N Mp_{\mathbf{n}}^N$ and $s((M-1)p_{\mathbf{n}}^N, p_{\mathbf{n}}^N) \leq p_{\mathbf{n}}^N$ we get for $\eta = \eta_0 \sqrt{\frac{\log \hat{\mathbf{n}}}{\hat{\mathbf{n}} \phi_x(h)}}$

$$B_2 \leq \hat{\mathbf{n}} p_{\mathbf{n}}^N (\log \hat{\mathbf{n}})^{-1/2} (\hat{\mathbf{n}} \phi_x(h))^{-1/2} \varphi(p_{\mathbf{n}}).$$

As $p_{\mathbf{n}} = C \left(\frac{\hat{\mathbf{n}} \phi_x(h)}{\log \hat{\mathbf{n}}} \right)^{1/2N}$, we can write

$$B_2 \leq \hat{\mathbf{n}} \varphi(p_{\mathbf{n}}). \quad (9)$$

Consequently, from (H7), we have

$$\sum_{\mathbf{n}} \hat{\mathbf{n}} \varphi(p_{\mathbf{n}}) < \infty.$$

Let us focus now on B_1 . For this, we must evaluate asymptotically $Var[Z_1^*]$.

Indeed,

$$Var[Z_1^*] = Var \left[\sum_{\mathbf{i} \in I(1, \mathbf{n}, x, \mathbf{1})} \Delta_{\mathbf{i}}(x) \right] = \sum_{\mathbf{i}, \mathbf{j} \in I(1, \mathbf{n}, x, \mathbf{1})} |Cov(\Delta_{\mathbf{i}}(x), \Delta_{\mathbf{j}}(x))|.$$

$$\text{Let } Q_{\mathbf{n}} = \sum_{\mathbf{i} \in I(1, \mathbf{n}, x, \mathbf{1})} Var[\Delta_{\mathbf{i}}(x)] \quad \text{and} \quad R_{\mathbf{n}} = \sum_{\mathbf{i} \neq \mathbf{j} \in I(1, \mathbf{n}, x, \mathbf{1})} |Cov(\Delta_{\mathbf{i}}(x), \Delta_{\mathbf{j}}(x))|.$$

By assumptions (H2) and (H1), we have

$$Var[\Delta_{\mathbf{i}}(x)] \leq C(\phi_x(h) + (\phi_x(h))^2),$$

therefore

$$Q_{\mathbf{n}} = O(p_{\mathbf{n}}^N \phi_x(h)).$$

Concerning $R_{\mathbf{n}}$ we introduce the following sets :

$$S_1 = \{\mathbf{i}, \mathbf{j} \in I(1, \mathbf{n}, x, \mathbf{1}) : 0 < \|\mathbf{i} - \mathbf{j}\| \leq c_{\mathbf{n}}\}, \quad S_2 = \{\mathbf{i}, \mathbf{j} \in I(1, \mathbf{n}, x, \mathbf{1}) : \|\mathbf{i} - \mathbf{j}\| > c_{\mathbf{n}}\},$$

where $c_{\mathbf{n}}$ is a real sequence that converges to $+\infty$ and will be precise after.

Split this sum into two separate summations over sites in S_1 and S_2

$$\begin{aligned} R_{\mathbf{n}} &= \sum_{(\mathbf{i}, \mathbf{j}) \in S_1} |Cov(\Delta_{\mathbf{i}}(x), \Delta_{\mathbf{j}}(x))| + \sum_{(\mathbf{i}, \mathbf{j}) \in S_2} |Cov(\Delta_{\mathbf{i}}(x), \Delta_{\mathbf{j}}(x))| \\ &= R_{\mathbf{n}}^1 + R_{\mathbf{n}}^2. \end{aligned}$$

On one hand, we have :

$$\begin{aligned} R_{\mathbf{n}}^1 &\leq C \sum_{(\mathbf{i}, \mathbf{j}) \in S_1} |\mathbb{E}[K_{\mathbf{i}}(x)K_{\mathbf{j}}(x)] + \mathbb{E}[K_{\mathbf{i}}(x)]\mathbb{E}[K_{\mathbf{j}}(x)]| \\ &\leq Cp_{\mathbf{n}}^N c_{\mathbf{n}}^N \phi_x(h) \left((\phi_x(h))^{1/a} + \phi_x(h) \right) \\ &\leq Cp_{\mathbf{n}}^N c_{\mathbf{n}}^N \phi_x(h)^{(a+1)/a}. \end{aligned}$$

On the other hand, we have

$$R_{\mathbf{n}}^2 = \sum_{(\mathbf{i}, \mathbf{j}) \in S_2} |Cov(\Delta_{\mathbf{i}}(x), \Delta_{\mathbf{j}}(x))|.$$

As the random variables Y and $K_{\mathbf{j}}$ are bounded, we deduce from the covariance inequality that

$$|Cov(\Delta_{\mathbf{i}}(x), \Delta_{\mathbf{j}}(x))| \leq C\varphi(\|\mathbf{i} - \mathbf{j}\|),$$

thus

$$\begin{aligned} R_{\mathbf{n}}^2 &\leq C \sum_{(\mathbf{i}, \mathbf{j}) \in S_2} \varphi(\|\mathbf{i} - \mathbf{j}\|) \leq Cp_{\mathbf{n}}^N \sum_{\mathbf{i}: \|\mathbf{i}\| \geq c_{\mathbf{n}}} \varphi(\|\mathbf{i}\|) \\ &\leq Cp_{\mathbf{n}}^N c_{\mathbf{n}}^{-Na} \sum_{\mathbf{i}: \|\mathbf{i}\| \geq c_{\mathbf{n}}} \|\mathbf{i}\|^{Na} \varphi(\|\mathbf{i}\|). \end{aligned}$$

Let $c_{\mathbf{n}} = (\phi_x(h))^{-1/Na}$, then we have

$$\begin{aligned} R_{\mathbf{n}}^2 &\leq Cp_{\mathbf{n}}^N c_{\mathbf{n}}^{-Na} \sum_{\mathbf{i}: \|\mathbf{i}\| \geq c_{\mathbf{n}}} \|\mathbf{i}\|^{Na} \varphi(\|\mathbf{i}\|) \\ &\leq Cp_{\mathbf{n}}^N \phi_x(h) \sum_{\mathbf{i}: \|\mathbf{i}\| \geq c_{\mathbf{n}}} \|\mathbf{i}\|^{Na} \varphi(\|\mathbf{i}\|). \end{aligned}$$

Because of (3.2.1) and (H2) we get

$$R_{\mathbf{n}}^2 \leq Cp_{\mathbf{n}}^N \phi_x(h).$$

Furthermore, under this choose of $c_{\mathbf{n}}$ we have

$$R_{\mathbf{n}}^1 \leq Cp_{\mathbf{n}}^N \phi_x(h).$$

Hence

$$Var[Z_1^*] = O(p_{\mathbf{n}}^N \phi_x(h)).$$

By using this last result, together with the definitions of $p_{\mathbf{n}}$, M and η , we get

$$B_1 \leq \exp(-C(\eta_0) \log \hat{\mathbf{n}})$$

Consequently an appropriate choose of η_0 completes the proof of the first part of this lemma.

Corollary 3 *Under the hypotheses of Theorem 3.2.1, we obtain :*

$$\sum_n \mathbb{P} \left| \tilde{g}_2(x) \right| \leq \frac{g_2(x)}{2} \right) < \infty.$$

Proof of Corollary 3. It is clear that

$$\tilde{g}_2(x) < \frac{g_2(x)}{2} \Rightarrow |g_2(x) - \tilde{g}_2(x)| > \frac{g_2(x)}{2}.$$

Thus,

$$\mathbb{P} \left(|\tilde{g}_2(x)| \leq \frac{g_2(x)}{2} \right) \leq \mathbb{P} \left(|g_2(x) - \tilde{g}_2(x)| \geq \frac{g_2(x)}{2} \right).$$

The use of the results of Lemma 1 and Lemma 6 complete the proof of the corollary. \blacksquare

Now, we study the asymptotic normality of $\hat{t}_p(x)$. In this case, assumptions (H1) and (H4) must be replaced by the following hypotheses, respectively.

(H1') The concentration property (H1) holds with nonnegative differentiable function $\phi_x(\cdot)$. Besides, there exists a function $\beta_x(\cdot)$ such that

$$\forall s \in [0, 1], \quad \lim_{r \rightarrow 0} \phi_x(sr)/\phi_x(r) = \beta_x(s).$$

(H4') $K(\cdot)$ satisfies (H4) and is a differentiable function on $(0, 1)$, with derivative $K'(\cdot)$ such that

$$-C_7 I_{(0,1)}(\cdot) \leq K'(\cdot) \leq -C_8 I_{(0,1)}(\cdot).$$

We also assume that x belongs to $\mathcal{A} = \{z \in \mathcal{F}, f^z(t_p(z)) \neq 0 \text{ and } \beta_z(\cdot) \text{ is not identically zero}\}$. The function $\beta_x(\cdot)$ defined in (H1') plays a fundamental role in the asymptotic normality result, since it permits to give the variance term explicitly. Note that this function can be specified in several situations where the function $\phi_x(\cdot)$ is known and (H1') is fulfilled. We quote the following cases, which can be found in Ferraty *et al.* (2007) :

- (i) $\phi_x(r) = \phi_0(x)r^\gamma$ for some $\gamma > 0$ with $\beta_x(s) = s^\gamma$.
- (ii) $\phi_x(r) = \phi_0(x)r^{\gamma_1} \exp\{-Cr^{-\gamma_2}\}$ for some $\gamma_1 > 0$ and $\gamma_2 > 0$ with $\beta_x(\cdot)$ being Dirac's function.
- (iii) $\phi_x(r) = \phi_0(x)|\ln r|^{-1}$ with $\beta_x(s) = I_{(0,1]}(s)$.

The first condition defining the set \mathcal{A} is related to a nonvanishing conditional density. The second one means that a small amount a concentration is needed in order to ensure asymptotic normality.

Now we are now ready for establishing the asymptotic normality result.

Theorem 3.2.2 *Under hypotheses $(H1')$, $(H2)$, $(H3)$, $(H4')$, $\hat{\mathbf{n}}h^{2b_1}\phi_x(h) \rightarrow 0$ as $\mathbf{n} \rightarrow \infty$, and if*

$$\exists \theta_1 \in \left[\frac{1}{1+N+\delta}, \frac{1}{1+2N} \right], \text{ such that, } \hat{\mathbf{n}}^{-1/(1+2N)+\theta_1} \leq \phi_x(h) \quad (10)$$

then we have

$$\left(\frac{\hat{\mathbf{n}}\phi_x(h)}{\sigma^2(x)} \right)^{1/2} (\tilde{\theta}(x) - \theta(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } \hat{\mathbf{n}} \rightarrow \infty$$

where

$$\left\{ \begin{array}{l} \sigma^2(x) = \frac{(g_2(x) - 2\theta(x)E[Y^{-3}|X=x] + r^2(x)E[Y^{-4}|X=x])}{g_2^2(x)} \frac{a_2(x)}{a_1^2(x)} \\ \text{with } a_j(x) = K^j(1) - \int_0^1 (K^j)'(s)\beta_x(s)ds \quad \text{for } j = 1, 2, \end{array} \right.$$

and $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution.

Similarly to Theorem 2.3.2 our 3.2.2 is a consequence of the following results

LEMMA 7 *Under the hypotheses of Theorem 3.2.2, we obtain :*

$$\left(\frac{\hat{\mathbf{n}}\phi_x(h)}{g_2(x)\sigma} \right)^{1/2} (B_{\mathbf{n}} - E[B_{\mathbf{n}}]) \rightarrow N(0, 1).$$

Proof of Lemma 7. Once again, we use the same arguments as the vectorial case by taking

$$\Lambda_{\mathbf{i}} := \frac{1}{\sqrt{\phi_x(h)}} \left(K_i(Y_{\mathbf{i}}^{-1}g_2(x) - Y_{\mathbf{i}}^{-2}g_1(x)) - \mathbb{E} [K_i(Y_{\mathbf{i}}^{-1}g_2(x) - Y_{\mathbf{i}}^{-2}g_1(x))] \right).$$

So,

$$\sqrt{\widehat{\mathbf{n}}\phi_x(h)} [\sigma_1(x)]^{-1} \left(B_n - \mathbb{E}[B_n] \right) = (\widehat{\mathbf{n}}\sigma_1^2(x))^{-1/2} S_{\mathbf{n}}$$

where $S_{\mathbf{n}} = \sum_{i=1}^n \Lambda_i$. Thus, our claimed result is, now

$$(\widehat{\mathbf{n}}\sigma_1^2(x))^{-1/2} S_{\mathbf{n}} \rightarrow \mathcal{N}(0, 1). \quad (11)$$

where $\sigma_1^2(x) = (g_2(x))^2 \sigma^2$.

This last is shown by the blocking method, where the random variables $\Lambda_{\mathbf{j}}$ are grouped into blocks of different sizes defined similarly to Lemma 6 by

$$W(1, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1, k=1,\dots,N \\ j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}}}^{\substack{j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}}} \Lambda_{\mathbf{i}},$$

$$W(2, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1, k=1,\dots,N-1 \\ i_N=j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1}}^{\substack{j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}} \\ (j_N+1)(p_{\mathbf{n}}+q_{\mathbf{n}})}} \Lambda_{\mathbf{i}},$$

$$W(3, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1, k=1,\dots,N-2 \\ i_{N-1}=j_{N-1}(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1 \\ i_N=j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}}}^{\substack{j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}} \\ (j_{N-1}+1)(p_{\mathbf{n}}+q_{\mathbf{n}}) \\ j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}}} \Lambda_{\mathbf{i}},$$

$$W(4, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+1, k=1,\dots,N-2 \\ i_{N-1}=j_{N-1}(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1 \\ i_N=j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1}}^{\substack{j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}} \\ (j_{N-1}+1)(p_{\mathbf{n}}+q_{\mathbf{n}}) \\ (j_N+1)(p_{\mathbf{n}}+q_{\mathbf{n}})}} \Lambda_{\mathbf{i}},$$

and so on. The last two terms are

$$W(2^{N-1}, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1, k=1,\dots,N-1}}^{(j_k+1)(p_{\mathbf{n}}+q_{\mathbf{n}})} \sum_{i_N=j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+1}^{j_N(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}} \Lambda_i,$$

$$W(2^N, \mathbf{n}, x, \mathbf{j}) = \sum_{\substack{i_k=j_k(p_{\mathbf{n}}+q_{\mathbf{n}})+p_{\mathbf{n}}+1, k=1,\dots,N}}^{(j_k+1)(p_{\mathbf{n}}+q_{\mathbf{n}})} \Lambda_i$$

where $q_{\mathbf{n}} = o\left(\left[\widehat{\mathbf{n}}\phi_x(a_{\mathbf{n}})^{(1+2N)}\right]^{1/(2N)}\right)$ and $p_{\mathbf{n}} = \left[(\widehat{\mathbf{n}}\phi_x(h))^{1/(2N)}/s_{\mathbf{n}}\right]$ with $s_{\mathbf{n}} = o\left(\left[\widehat{\mathbf{n}}\phi_x(h)^{(1+2N)}\right]^{1/(2N)} q_{\mathbf{n}}^{-1}\right)$. Noting that, by (10) we can show all sequences $q_{\mathbf{n}}$, $p_{\mathbf{n}}$ and $s_{\mathbf{n}}$ tend to infinity.

Now, we adopt the same notations as those given in Lemma 6 and we define for each integer $i = 1, \dots, 2^N$,

$$T(\mathbf{n}, x, i) = \sum_{\mathbf{j} \in \mathcal{J}} W(i, \mathbf{n}, x, \mathbf{j}).$$

where $\mathcal{J} = \{0, \dots, r_1 - 1\} \times \dots \times \{0, \dots, r_N - 1\}$ with $r_k = n_k(p_{\mathbf{n}} + q_{\mathbf{n}})^{-N}$.

Then, we have

$$\begin{aligned} & \sqrt{\widehat{\mathbf{n}}\phi_x(h)} [g_2(x)\sigma(x)]^{-1} (B_n(x) - \mathbb{E}[B_n(x)]) \\ &= \left[\sqrt{\widehat{\mathbf{n}}} g_2(x)\sigma(x) \right]^{-1} \left(T(\mathbf{n}, x, 1) + \sum_{i=2}^{2N} T(\mathbf{n}, x, i) \right). \end{aligned}$$

Therefore, it suffices to proof

$$\text{the asymptotic normality of : } \left[\sqrt{\widehat{\mathbf{n}}} g_2(x)\sigma(x) \right]^{-1} (T(\mathbf{n}, x, 1)) \quad (12)$$

and

$$\text{the convergence in probability of : } \sqrt{\widehat{\mathbf{n}}}^{-1} \left(\sum_{i=2}^{2N} T(\mathbf{n}, x, i) \right). \quad (13)$$

Firstly, we begin by proving (13). Clearly it is sufficient to show that

$$\widehat{\mathbf{n}}^{-1} \mathbb{E} \left[\sum_{i=2}^{2^N} T(\mathbf{n}, x, i) \right]^2 \rightarrow 0.$$

We have

$$\widehat{\mathbf{n}}^{-1} \mathbb{E} \left[\sum_{i=2}^{2^N} T(\mathbf{n}, x, i) \right]^2 = \widehat{\mathbf{n}}^{-1} \left(\sum_{i=2}^{2^N} \mathbb{E} [T(\mathbf{n}, x, i)]^2 + \sum_{i,j=2, \dots, 2^N, i \neq j} \mathbb{E} [T(\mathbf{n}, x, i)T(\mathbf{n}, x, j)] \right).$$

By Cauchy-Schwartz inequality, we get :

$$\forall 2 \leq i \leq 2^N : \widehat{\mathbf{n}}^{-1} \mathbb{E} [T(\mathbf{n}, x, i)T(\mathbf{n}, x, j)] \leq (\widehat{\mathbf{n}}^{-1} \mathbb{E} [T(\mathbf{n}, x, i)]^2)^{1/2} (\widehat{\mathbf{n}}^{-1} \mathbb{E} [T(\mathbf{n}, x, j)]^2)^{1/2}.$$

Then, all what is left to be shown is to prove that

$$\widehat{\mathbf{n}}^{-1} \mathbb{E} [T(\mathbf{n}, x, i)]^2 \rightarrow 0 \quad ; \forall 2 \leq i \leq 2^N. \quad (14)$$

We will only prove (14) for $i = 2$, the others case is identique. Analogously to Lemma 6, we enumerate $W(2, \mathbf{n}, x, \mathbf{j})$ in the arbitrary way $\hat{W}_1, \dots, \hat{W}_M$, and we write

$$\begin{aligned} E [T(\mathbf{n}, x, 2)]^2 &= \sum_{i=1}^M Var [\hat{W}_i] + \sum_{i=1}^M \sum_{j=1, j \neq i}^M Cov (\hat{W}_i, \hat{W}_j) \\ &= A_1 + A_2. \end{aligned}$$

For the variance term we have

$$\begin{aligned} Var [\hat{W}_i] &= Var \left[\sum_{i_k=1, k=1, \dots, N-1}^{p_{\mathbf{n}}} \sum_{i_N=1}^{q_{\mathbf{n}}} \Lambda_{\mathbf{i}} \right] \\ &= p_{\mathbf{n}}^{N-1} q_{\mathbf{n}} Var [\Lambda_{\mathbf{i}}] \\ &\quad + \sum_{i_k=1, k=1, \dots, N-1}^{p_{\mathbf{n}}} \sum_{i_N=1}^{q_{\mathbf{n}}} \sum_{j_k=1, k=1, \dots, N-1, \mathbf{i} \neq \mathbf{j}}^{p_{\mathbf{n}}} \sum_{j_N=1}^{q_{\mathbf{n}}} \mathbb{E} [\Lambda_{\mathbf{i}} \Lambda_{\mathbf{j}}]. \end{aligned}$$

It is shown in Lemma 3.6 in Messabihi *et al.* (2015) that

$$\text{Var}[\Lambda_1] \rightarrow (g_2(x)\sigma(x))^2. \quad (15)$$

Moreover, employing Lemma in Carbon, to get

$$|E\Lambda_i\Lambda_j| \leq C\phi_x(h)^{-1}\varphi(\|\mathbf{i} - \mathbf{j}\|).$$

So, we deduce that

$$\begin{aligned} \text{Var}[\hat{W}_i] &\leq Cp_{\mathbf{n}}^{N-1}q_{\mathbf{n}} \left(1 + \phi_x(h)^{-1} \sum_{i_k=1, k=1,\dots,N-1}^{p_{\mathbf{n}}} \sum_{i_N=1}^{q_{\mathbf{n}}} (\varphi(\|\mathbf{i}\|)) \right) \\ &\leq Cp_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\phi_x(h)^{-1} \sum_{i_k=1, k=1,\dots,N-1}^{p_{\mathbf{n}}} \sum_{i_N=1}^{q_{\mathbf{n}}} (\varphi(\|\mathbf{i}\|)). \end{aligned}$$

It is shown in Lemma 3.6 in Messabihi *et al.* (2015) that

$$\text{Var}[\Lambda_1] \rightarrow (g_2(x)\sigma(x))^2. \quad (16)$$

Moreover, employing Lemma ??, to get

$$|E\Lambda_i\Lambda_j| \leq C\phi_x(h)^{-1}\varphi(\|\mathbf{i} - \mathbf{j}\|).$$

So, we deduce that

$$\begin{aligned} \text{Var}[\hat{W}_i] &\leq Cp_{\mathbf{n}}^{N-1}q_{\mathbf{n}} \left(1 + \phi_x(h)^{-1} \sum_{i_k=1, k=1,\dots,N-1}^{p_{\mathbf{n}}} \sum_{i_N=1}^{q_{\mathbf{n}}} (\varphi(\|\mathbf{i}\|)) \right) \\ &\leq Cp_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\phi_x(h)^{-1} \sum_{i_k=1, k=1,\dots,N-1}^{p_{\mathbf{n}}} \sum_{i_N=1}^{q_{\mathbf{n}}} (\varphi(\|\mathbf{i}\|)). \end{aligned}$$

Therefore

$$\widehat{\mathbf{n}}^{-1}A_1 \leq CMp_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\widehat{\mathbf{n}}^{-1}\phi_x(h)^{-1} \sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1}\varphi(i).$$

The definitions of M and $p_{\mathbf{n}}$ permit to get

$$\begin{aligned} CMp_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\widehat{\mathbf{n}}^{-1}\phi_x(h)^{-1} &= \widehat{\mathbf{n}}(p_{\mathbf{n}}+q_{\mathbf{n}})^{-N}p_{\mathbf{n}}^{N-1}q_{\mathbf{n}}\widehat{\mathbf{n}}^{-1}\phi_x(h)^{-1} \\ &\leq \left(\frac{q_{\mathbf{n}}}{p_{\mathbf{n}}}\right)\phi_x(h)^{-1} \\ &= q_{\mathbf{n}}s_{\mathbf{n}}(\widehat{\mathbf{n}}\phi_x(h))^{\frac{-1}{2N}}\phi_x(h)^{-1} \\ &= q_{\mathbf{n}}s_{\mathbf{n}}(\widehat{\mathbf{n}}\phi_x(h)^{(1+2N)})^{\frac{-1}{2N}}. \end{aligned}$$

Using the fact that $s_{\mathbf{n}} = o\left([\widehat{\mathbf{n}}\phi_x(h)^{(1+2N)}]^{1/(2N)}q_{\mathbf{n}}^{-1}\right)$ it is easy to see that the last term of (17) converges to $\rightarrow 0$. Furthermore, by (3.2.1) with $\delta > N$ (see, hypothesis (H2)) we show also that

$$\sum_{i=1}^{\infty} i^{N-1}\varphi(i) < \infty.$$

Finally, we deduce that

$$\widehat{\mathbf{n}}^{-1}A_1 \rightarrow 0.$$

We now proceed to evaluate A_2 . A simple computation shows that the sites of random variables $\Lambda_{\mathbf{i}}$ involved in two variables \hat{W}_i and \hat{W}_j with $i \neq j$ are far apart by distant of $q_{\mathbf{n}}$ at least. So, by covariance inequality in a spatial mixing variables (see, Lemma 5.2) we get

$$\begin{aligned} A_2 &\leq \sum_{j_k=1, k=1,\dots,N}^{n_k} \sum_{i_k=1, k=1,\dots,N}^{n_k} \|\mathbf{i} - \mathbf{j}\| > q_{\mathbf{n}}E\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}} \\ &\leq C\phi_x(h)^{-1}\widehat{\mathbf{n}} \sum_{i_k=1, k=1,\dots,N, \|\mathbf{i}\|>q_{\mathbf{n}}}^{n_k} \varphi(\|\mathbf{i}\|) \end{aligned}$$

and

$$\widehat{\mathbf{n}}^{-1}A_2 \leq C\phi_x(h)^{-1} \sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1}\varphi(i).$$

Observe that

$$\phi_x(h)^{-1} \sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1} \varphi(i) \leq \phi_x(h)^{-1} \sum_{i=q_{\mathbf{n}}}^{\infty} i^{N-1-\delta} \leq \phi_x(h)^{-1} \int_{q_{\mathbf{n}}}^{\infty} t^{N-1-\delta} dt = C \phi_x(h)^{-1} q_{\mathbf{n}}^{N-\delta}.$$

This latter goes to 0 by means of (10) and the definition of $q_{\mathbf{n}}$. So, we get

$$\widehat{\mathbf{n}}^{-1} A_2 \rightarrow 0.$$

This completes the proof of (13).

Secondly, to prove the asymptotic normality (12) it is sufficient to show the three claim

$$Q_1 \equiv \left| \mathbb{E} [\exp [iuT(\mathbf{n}, x, 1)]] \right| \quad (17)$$

$$- \prod_{j_k=0, k=1, \dots, N}^{r_k-1} \mathbb{E} [\exp [iuW(1, \mathbf{n}, x, \mathbf{j})]] \rightarrow 0.$$

$$Q_2 \equiv \widehat{\mathbf{n}}^{-1} \sum_{\mathbf{j} \in \mathcal{J}} \mathbb{E} [W(1, \mathbf{n}, x, \mathbf{j})]^2 \rightarrow (g_2(x)\sigma(x))^2 \quad (18)$$

and

$$Q_3 \equiv \widehat{\mathbf{n}}^{-1} \sum_{\mathbf{j} \in \mathcal{J}} \mathbb{E} [(W(1, \mathbf{n}, x, \mathbf{j}))^2 \mathbf{1}_{\{|W(1, \mathbf{n}, x, \mathbf{j})| > \epsilon((g_2(x)\sigma(x))^2 \widehat{\mathbf{n}})^{1/2}\}}] \rightarrow 0 \quad (19)$$

for all $\epsilon > 0$

Proof of (17) : The proof of (17) is based on the Lemma (Nakh 1987) to the variable $(\exp(iu\tilde{W}_1), \dots, \exp(iu\tilde{W}_M))$ where $\tilde{W}_1, \dots, \tilde{W}_M$ are the random variables $W(1, \mathbf{n}, x, \mathbf{j})_{\mathbf{j} \in \mathcal{J}}$ enumerated in the arbitrary way. As $\left| \prod_{s=j+1}^M \exp[iu\tilde{W}_s] \right| \leq$

1, then

$$\begin{aligned}
Q_1 &= \left| \mathbb{E} [\exp [iuT(\mathbf{n}, x, 1)]] - \prod_{j_k=0, k=1,\dots,N}^{r_k-1} \mathbb{E} [\exp [iuW(1, \mathbf{n}, x, \mathbf{j})]] \right| \\
&= \left| \mathbb{E} \left[\prod_{j_k=0, k=1,\dots,N}^{r_k-1} \exp [iuW(1, \mathbf{n}, x, \mathbf{j})] \right] - \prod_{j_k=0, k=1,\dots,N}^{r_k-1} \mathbb{E} [\exp [iuW(1, \mathbf{n}, x, \mathbf{j})]] \right| \\
&\leq \sum_{k=1}^{M-1} \sum_{j=k+1}^M \left| \mathbb{E} (\exp[iu\tilde{W}_k] - 1) (\exp[iu\tilde{W}_j] - 1) \prod_{s=j+1}^M \exp[iu\tilde{W}_s] \right. \\
&\quad \left. - \mathbb{E} (\exp[iu\tilde{W}_k] - 1) \mathbb{E} (\exp[iu\tilde{W}_j] - 1) \prod_{s=j+1}^M \exp[iu\tilde{W}_s] \right| \\
&= \sum_{k=1}^{M-1} \sum_{j=k+1}^M \left| \mathbb{E} (\exp[iu\tilde{W}_k] - 1) (\exp[iu\tilde{W}_j] - 1) - \mathbb{E} (\exp[iu\tilde{W}_k] - 1) \mathbb{E} (\exp[iu\tilde{W}_j] - 1) \right| \\
&\quad \times \left| \prod_{s=j+1}^M \exp[iu\tilde{W}_s] \right| \\
&\leq \sum_{k=1}^{M-1} \sum_{j=k+1}^M \left| \mathbb{E} (\exp[iu\tilde{W}_k] - 1) (\exp[iu\tilde{W}_j] - 1) - \mathbb{E} (\exp[iu\tilde{W}_k] - 1) \mathbb{E} (\exp[iu\tilde{W}_j] - 1) \right|.
\end{aligned}$$

Let \tilde{I}_j be the set of sites such that $\tilde{W}_j = \sum_{\mathbf{i} \in \tilde{I}(1, \mathbf{n}, x, \mathbf{j})} \Lambda_{\mathbf{i}}$. Since the sets $\tilde{I}_{1 \leq j \leq M}$ contains $p_{\mathbf{n}}^N$ sites, we have by Lemma 5.2, under (2)

$$\left| \mathbb{E} (\exp[iu\tilde{W}_k] - 1) (\exp[iu\tilde{W}_j] - 1) - \mathbb{E} (\exp[iu\tilde{W}_k] - 1) \mathbb{E} (\exp[iu\tilde{W}_j] - 1) \right| \leq C \varphi(d(\tilde{I}_j, \tilde{I}_k)) p_{\mathbf{n}}^N.$$

Hence

$$\begin{aligned}
Q_1 &\leq Cp_{\mathbf{n}}^N \sum_{k=1}^{M-1} \sum_{j=k+1}^M \varphi(d(\tilde{I}_j, \tilde{I}_k)) \\
&\leq Cp_{\mathbf{n}}^N M \sum_{k=2}^M \varphi(d(\tilde{I}_1, \tilde{I}_k)) \\
&\leq Cp_{\mathbf{n}}^N M \sum_{i=1}^{\infty} \sum_{k: iq_{\mathbf{n}} \leq d(\tilde{I}_1, \tilde{I}_k) < (i+1)q_{\mathbf{n}}} \varphi(d(\tilde{I}_1, \tilde{I}_k)) \\
&\leq Cp_{\mathbf{n}}^N M \sum_{i=1}^{\infty} i^{N-1} \varphi(iq_{\mathbf{n}}).
\end{aligned}$$

It follows from (3.2.1) that

$$Q_1 \leq C\hat{\mathbf{n}} q_{\mathbf{n}}^{-\delta} \sum_{i=1}^{\infty} i^{N-1-\delta}.$$

The convergence result (17) is consequence of (10) and the definition of $q_{\mathbf{n}}$.

Proof of (18) : On the one hand

$$\begin{aligned}
\hat{\mathbf{n}}^{-1} \mathbb{E}[T(\mathbf{n}, x, 1)]^2 &= \hat{\mathbf{n}}^{-1} \sum_{j_k=0, k=1, \dots, N}^{r_k-1} \mathbb{E}[W(1, \mathbf{n}, x, \mathbf{j})]^2 \\
&+ \hat{\mathbf{n}}^{-1} \sum_{j_k=0, k=1, \dots, N}^{r_k-1} \sum_{i_k=0, k=1, \dots, N}^{r_k-1}, i_k \neq j_k \text{ for some } k \text{Cov}[W(1, \mathbf{n}, x, \mathbf{j}), W(1, \mathbf{n}, x, \mathbf{i})]
\end{aligned}$$

By the same arguments as those used for A_2 , this last term tend to zero.

Hence, the limit in (18) is equal to the limit of $\hat{\mathbf{n}}^{-1} E(T(\mathbf{n}, x, 1))^2$. On the other hand recall that

$$S_{\mathbf{n}} = T(\mathbf{n}, x, 1) + S''_{\mathbf{n}} = \sum_{i=2}^{2^N} T(\mathbf{n}, x, i)$$

where $S''_{\mathbf{n}} = \sum_{i=2}^{2^N} T(\mathbf{n}, x, i)$. Therefore

$$\widehat{\mathbf{n}}^{-1}\mathbb{E}[T(\mathbf{n}, x, 1)]^2 = \widehat{\mathbf{n}}^{-1}\mathbb{E}[S_{\mathbf{n}}^2] + \widehat{\mathbf{n}}^{-1}\mathbb{E}[S''_{\mathbf{n}}]^2 - 2\widehat{\mathbf{n}}^{-1}\mathbb{E}[S_{\mathbf{n}}S''_{\mathbf{n}}].$$

It is shown in (14) that $\widehat{\mathbf{n}}^{-1}\mathbb{E}[S''_{\mathbf{n}}]^2 \rightarrow 0$. Moreover, by Cauchy-Schwartz's inequality, we can write :

$$|\widehat{\mathbf{n}}^{-1}\mathbb{E}[S_{\mathbf{n}}S''_{\mathbf{n}}]| \leq \widehat{\mathbf{n}}^{-1}\mathbb{E}|S_{\mathbf{n}}S''_{\mathbf{n}}| \leq (\widehat{\mathbf{n}}^{-1}\mathbb{E}[S_{\mathbf{n}}]^2)^{1/2} \left(\widehat{\mathbf{n}}^{-1}\mathbb{E}[S''_{\mathbf{n}}]^2 \right)^{1/2}$$

Thus, all what it remains to compute is the limit of $\widehat{\mathbf{n}}^{-1}\mathbb{E}[S_{\mathbf{n}}]^2$ which can be writhen

$$\widehat{\mathbf{n}}^{-1}\mathbb{E}[S_{\mathbf{n}}]^2 = \widehat{\mathbf{n}}^{-1}Var[S_{\mathbf{n}}^2] = \widehat{\mathbf{n}}^{-1} \left(\sum_{\mathbf{i}} Var[\Lambda_{\mathbf{i}}] + \sum_{\mathbf{i} \neq \mathbf{j}} Cov[\Lambda_{\mathbf{i}}, \Lambda_{\mathbf{j}}] \right).$$

As indicated in (16) the variance terme is

$$Var[\Lambda_1] \rightarrow (g_2(x)\sigma(x))^2.$$

Let us evaluate the covariance term. Reasoning as in Lemma 7 we consider

$$E_1 = \{\mathbf{i}, \mathbf{j} \in \mathcal{I}_{\mathbf{n}} : 0 < \|\mathbf{i} - \mathbf{j}\| \leq c_{\mathbf{n}}\},$$

$$E_2 = \{\mathbf{i}, \mathbf{j} \in \mathcal{I}_{\mathbf{n}} : \|\mathbf{i} - \mathbf{j}\| > c_{\mathbf{n}}\}$$

where $c_{\mathbf{n}}$ is a sequence of integers that converges to infinite and that will be precise after.

Now, we write

$$\sum_{i \neq j} Cov[\Lambda_i, \Lambda_j] = \sum_{(i,j) \in E_1} Cov[\Lambda_i, \Lambda_j] + \sum_{(i,j) \in E_2} Cov[\Lambda_i, \Lambda_j].$$

For the first sum on E_1 , proceeding as in (R_n^1) in Lemma 6, we get from (3.2.1)

$$|Cov [\Lambda_i, \Lambda_j]| \leq C \left(\phi_x(h) + (\phi_x(h))^{1/a} \right) \leq C(\phi_x(h))^{1/a}.$$

It follows that,

$$\sum_{E_1} Cov (\Lambda_i, \Lambda_j) \leq C \widehat{\mathbf{n}} c_n^N \phi_x(h)^{1/a}.$$

Next, on E_2 we apply Lemma lemme 4.1 (Carbon) and $|\Lambda_i| \leq C\phi_x(h)^{-1/2}$, permit ro write that :

$$|Cov (\Lambda_i, \Lambda_j)| \leq C\phi_x(h)^{-1} \varphi (\|\mathbf{i} - \mathbf{j}\|)$$

and

$$\begin{aligned} \sum_{E_2} Cov (\Lambda_i, \Lambda_j) &\leq C\phi_x(h)^{-1} \sum_{(\mathbf{i}, \mathbf{j}) \in E_2} \varphi (\|\mathbf{i} - \mathbf{j}\|) \\ &\leq C \widehat{\mathbf{n}} \phi_x(h)^{-1} \sum_{\mathbf{i}: \|\mathbf{i}\| > c_n} \varphi (\|\mathbf{i}\|) \\ &\leq C \widehat{\mathbf{n}} \phi_x(h)^{-1} c_n^{-\delta} \sum_{\mathbf{i}: \|\mathbf{i}\| > c_n} \|\mathbf{i}\|^\delta \varphi (\|\mathbf{i}\|). \end{aligned}$$

Finally, we have :

$$\sum Cov (\Lambda_i, \Lambda_j) \leq \left(C \widehat{\mathbf{n}} c_n^N \phi_x(h)^{1/a} + C \widehat{\mathbf{n}} \phi_x(h)^{-1} c_n^{-\delta} \sum_{\mathbf{i}: \|\mathbf{i}\| > c_n} \|\mathbf{i}\|^\delta \varphi (\|\mathbf{i}\|) \right).$$

Let $c_n = \phi_x(h)^{-\alpha}$ for some $(\delta)^{-1} < \alpha < (Na)^{-1}$, then we have :

$$\sum Cov (\Lambda_i, \Lambda_j) \leq \left(C \widehat{\mathbf{n}} \phi_x(h)^{-\alpha N + 1/a} + C \widehat{\mathbf{n}} \phi_x(h)^{\alpha \delta - 1} \sum_{\mathbf{i}: \|\mathbf{i}\| > c_n} \|\mathbf{i}\|^\delta \varphi (\|\mathbf{i}\|) \right).$$

Hence, we obtain that

$$\sum Cov (\Lambda_i, \Lambda_j) = o (\widehat{\mathbf{n}}).$$

In conclusion, we have

$$\widehat{\mathbf{n}}^{-1} \sum_{\mathbf{j} \in \mathcal{J}} E \left[W(1, \mathbf{n}, x, \mathbf{j}) \right]^2 \rightarrow g_2(x) \sigma \text{ when } \mathbf{n} \rightarrow \infty.$$

Proof of 19 : Because $|\Lambda_i| \leq C\phi_x(h)^{-1/2}$, then $|W(1, \mathbf{n}, x, \mathbf{j})| \leq Cp_{\mathbf{n}}^N \phi_x(h)^{-1/2}$.

Thus

$$Q_4 \leq Cp_{\mathbf{n}}^{2N} \phi_x(h)^{-1} \widehat{\mathbf{n}}^{-1} \sum_{j_k=0, k=1, \dots, N}^{r_k-1} \mathbb{P} \left[|W(1, \mathbf{n}, x, \mathbf{j})| > \epsilon ((g_2(x)\sigma(x))^2 \widehat{\mathbf{n}})^{1/2} \right].$$

Since $p_{\mathbf{n}} = \left[(\widehat{\mathbf{n}}\phi_x(h))^{1/(2N)} / s_{\mathbf{n}} \right]$ and $s_{\mathbf{n}} \rightarrow \infty$ then

$$\begin{aligned} |W(1, \mathbf{n}, x, \mathbf{j})| / \left(((g_2(x)\sigma(x))^2 \widehat{\mathbf{n}})^{1/2} \right) &\leq Cp_{\mathbf{n}}^N (\widehat{\mathbf{n}}\phi_x(h))^{-1/2} \\ &= C(s_{\mathbf{n}})^{-N} \rightarrow 0. \end{aligned}$$

So, for all ϵ and $\mathbf{j} \in \mathcal{J}$; if \mathbf{n} is great enough, we have

$\mathbb{P} \left[W(1, \mathbf{n}, x, \mathbf{j}) > \epsilon ((g_2(x)\sigma(x))^2 \widehat{\mathbf{n}})^{1/2} \right] = 0$. Then $Q_4 = 0$ for \mathbf{n} great enough. This yields the proof. \blacksquare

LEMMA 8 Under the hypotheses of Theorem 3.2.2, we obtain :

$\tilde{g}_2(x) \rightarrow g_2(x)$, in probability,

$$\left(\frac{\widehat{\mathbf{n}}\phi_x(h)}{g_2(x)^2 \sigma^2(x)} \right)^{1/2} V_{\mathbf{n}} \rightarrow 0,$$

and

$$\left(\frac{\widehat{\mathbf{n}}\phi_x(h)}{g_2(x)^2 \sigma^2(x)} \right)^{1/2} V_{\mathbf{n}} (\tilde{g}_2(x) - E\tilde{g}_2(x)) \rightarrow 0, \text{ in probability.}$$

Proof of Lemma 8. For the first limit, we have, by Lemma (1)

$$E [\tilde{g}_2(x) - g_2(x)] \rightarrow 0$$

and by a similar argument as those used in the variance term in Lemma (6) we show that

$$\text{Var} [\tilde{g}_2(x)] \rightarrow 0$$

hence

$$\tilde{g}_2(x) - g_2(x) \rightarrow 0 \quad \text{in probability.}$$

Next, it is clear that the second limit is consequence for the last convergence.

So, it suffices to treat the last one. For this, we use the fact that

$$\text{Var} [(\tilde{g}_2(x) - E\tilde{g}_2(x))] = \text{Var} [\tilde{g}_2(x)] \rightarrow 0$$

and

$$V_{\mathbf{n}} = O(h^b) \quad (\text{see, Lemma (1)})$$

The last part of Condition (H6') allows to deduce that

$$\left(\frac{\widehat{\mathbf{n}}\phi_x(h)}{g_1(x)^2\sigma^2} \right)^{1/2} V_{\mathbf{n}} (\tilde{g}_2(x) - E\tilde{g}_2(x)) \rightarrow 0 \quad \text{in probability.}$$

3.3 Références

Carbon, Michel ; Hallin, Marc ; Tran, Lanh Tat. Kernel density estimation for random fields : the L_1 theory. J. Nonparametr. Statist. 6 (1996), no. 2-3, 157–170.

Doukhan, Paul ; Massart, Pascal ; Rio, Emmanuel. The functional central limit theorem for strongly mixing processes. Ann. Inst. H. Poincaré Probab. Statist. 30 (1994), no. 1, 63-82.

Guyon, Xavier. Estimation d'un champ par pseudo-vraisemblance conditionnelle : étude asymptotique et application au cas markovien. (French) [Estimation of a field by pseudoconditional likelihood : asymptotic study and application to the Markov case] Spatial processes and spatial time series analysis (Brussels, 1985), 15–62, Travaux Rech., 11, Publ. Fac. Univ. Saint-Louis, Brussels, 1987.

Li, Jiexiang; Tran, Lanh Tat. Hazard rate estimation on random fields. *J. Multivariate Anal.* 98 (2007), no. 7, 1337–1355.

Tran, Lanh Tat. Kernel density estimation under dependence. *Statist. Probab. Lett.* 10 (1990), no. 3, 193–201.