HAL open science

# Régression et prédiction non-paramétrique spatiale

Sophie Dabo-Niang, Anne-Françoise Yao

## To cite this version:

Sophie Dabo-Niang, Anne-Françoise Yao. Régression et prédiction non-paramétrique spatiale. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494744

**HAL Id: inria-00494744**

**https://inria.hal.science/inria-00494744**

Submitted on 24 Jun 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Régression et prédiction non-paramétrique spatiale

Sophie Dabo-Niang[1] & Anne-Françoise Yao[2]

*1. Laboratoire GREMARS, Université Charles De Gaulle, Lille 3, maison de la recherche, domaine du pont de bois, BP 60149, 59653 Villeneuve d'ascq cedex, France.*
*sophie.dabo@univ-lille3.fr*
*2. Laboratoire LMGEM, Université Aix-Marseille 2, Campus de Luminy, case 901, 13288 Marseille cedex 09.*
*anne-francoise.yao@univmed.fr*

**Résumé**. Nous nous intéressons à l'estimation de la fonction de régression $r(x) = E\left(Y_{\mathbf{u}}|X_{\mathbf{u}} = x\right)$ à partir d'observations d'un processus $\left\{Z_{\mathbf{i}} = (X_{\mathbf{i}},\ Y_{\mathbf{i}}),\ \mathbf{i} \in \mathbb{Z}^N\right\}$. On suppose que les variables $Z_{\mathbf{i}}$ sont de même distribution que $Z = (X, Y)$, où $Y$ est une variable réelle, intégrable et $X$ un vecteur aléatoire à valeurs dans un espace séparable $\mathcal{E}$ muni (éventuellement de dimension infinie). Dans ce travail, la convergence nos estimateurs est étudiée sous conditions de mélange à partir d'observations dans une région rectangulaire de $\mathbb{Z}^N$. Nous illustrerons nos résultats par des simulations. L'application de nos méthodes à la prédiction spatiale sera également abordée.

**Mots clés :** *Estimateur à noyau, régression spatiale, champs aléatoire, régression en grande dimension, conditions de mélange.*

**Abstract**: We investigate kernel estimate of the regression function $r(x) = E\left(Y_{\mathbf{u}}|X_{\mathbf{u}} = x\right)$, based on observations of a strictly stationary spatial process $\left\{Z_{\mathbf{i}} = (X_{\mathbf{i}},\ Y_{\mathbf{i}}),\ \mathbf{i} \in \mathbb{Z}^N\right\}$. We suppose that the process is such that the $Z_{\mathbf{i}}'$s have the same distribution as a variable $Z = (X, Y)$, where $Y$ is a real-valued and integrable variable and $X$ valued in a separable space $\mathcal{E}$ (of eventually infinite dimension). The weak and the strong consistencies of the estimates are shown under sufficient conditions on mixing coefficients and the bandwidth, when the process is observed over a rectangular domain of $\mathbb{Z}^N$. We will illustrate our purpose through some simulation studies. A first approach of the kernel spatial prediction is given.

**Key Words:** *Kernel regression estimation, spatial regression, random fields, high dimensional regression, mixing conditions.*

# 1 Introduction

L'estimation de la fonction régression ainsi que prévision spatiale sont des problèmes importants dans de nombreuses applications où l'influence d'un vecteur de covariables sur certaines variables de réponse doit être étudiée dans un contexte de dépendance spatiale.

Dans de nombreux domaines : géologie, océanographie, économétrie, agronomie, épidémiologie, physique, environnement, gestion des risques, traitement d'images, etc... les données sont spatialement dépendantes. Leurs traitements requiert des outils spécifiques fourni par la statistique spatiale. Si la modélisation sous hypothèse paramétriques a fait l'objet de nombreuses applications, le traitement non paramétrique ce ces données est très limité.

A notre connaissance, très peu de modèles spatiaux sous hypothèses non-paramétriques de la dépendance spatiale existent dans la littérature : citons Lu et Chen (2004), Hallin et al. (2004, 2009), Carbon et al. (2007), Dabo-Niang et Yao (2007) en régression et Dabo-Niang et Yao (2007), Biau et Cadre (2004) et Menezes et al. (2009) en prédiction spatiales. Dans ce travail nous intéressons à l'estimation de la fonction de régression et la prédiction spatiale sous hypothèses hypothèse non-paramétrique. Nous illustrerons nos propos au travers de simulations.

Considérons un processus spatial strictement stationnaire $Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}})$, $\mathbf{i} \in \mathbb{Z}^N$ $(N \geq 1)$, défini sur un espace probabilisé $(\Omega,\ \mathcal{A}, \mathbf{P})$ tel que les variables $Z_{\mathbf{i}}$ sont de même distribution que la variable $Z = (X, Y)$, où $Y$ est une variable réelle, intégrable et $X$ un vecteur d'un espace séparable $\mathcal{E}$ muni (éventuellement de dimension infinie).

# 2 Régression non-paramétrique spatiale : cadre fini-dimensionnel

Nous nous intéressons à l'estimateur non-paramétrique de la fonction de l'estimation, $r$ définie par $r(x) = E\left(Y|X = x\right)$ .

Dans le cas où $\mathcal{E} = \mathbb{R}^d$, l'estimateur à noyau obtenu à partir d'observations une région rectangulaire, $\mathcal{I}_{\mathbf{n}} = \left\{\mathbf{i} \in \mathbb{N}^N \ : \ 1 \leq i_k \leq n_k,\ k = 1, ..., N\right\}$ est donné par :

$$r_{\mathbf{n}}(x) = \begin{cases} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} W_{\mathbf{i},x} & \text{if } \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} W_{\mathbf{i},x} \neq 0; \\ \frac{1}{\widehat{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} & \text{sinon.} \end{cases}$$

où

$$W_{\mathbf{i},x} = \frac{K\left(\frac{X_{\mathbf{i}}-x}{h_{\mathbf{n}}}\right)}{\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} K\left(\frac{X_{\mathbf{j}}-x}{h_{\mathbf{n}}}\right)},$$

$\lim_{\mathbf{n}\to+\infty} h_{\mathbf{n}} = 0\,(+)$ , et le noyau $K$ est une fonction de $\mathbb{R}^d$ dans $\mathbb{R}^+$.

Nous rappelerons quelques résultats sur cet estimateur en dimension finie.

# 3 Estimation de la fonction de régression en grande dimension

Cette partie est consacrée au cas où, $\mathcal{E}$ est de grande dimension voire de dimension infinie. Dans ce cadre, nous proposons d'estimer la fonction de régression sous l'hypothèse :

$$r(x) = g(\Phi.x)$$

où $\Phi$ est application de $\mathcal{E}$ dans $\mathbb{R}^D$ avec $D \geq 1$ raisonnablement petit. Nous considérons les cas suivant :

## 3.1 L'application $\Phi$ est définie par $\Phi(x) = d(X, x)$ où $d$ est une semi-distance

Dans ce cas l'estimateur $r_{\mathbf{n}}(x)$ est tel que

$$W_{\mathbf{i},x} = \frac{K\left(\frac{d(X_{\mathbf{i}},x)}{h_{\mathbf{n}}}\right)}{\sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} K\left(\frac{d(X_{\mathbf{i}},x)}{h_{\mathbf{n}}}\right)}$$

avec $\lim_{\mathbf{n}\to+\infty} h_{\mathbf{n}} = 0\,(+)$ et le noyau tel que $K$ est une fonction de $\mathbb{R}^+$ dans $\mathbb{R}^+$. Nous présenterons des résultats de convergence de cet estimateur dans ce contexte.

## 3.2 L'application $\Phi$ est linéaire.

Nous considérons les cas suivants où $\mathcal{E}$ est muni d'un produit scalaire et $\Phi$ est un opérateur de projection dans un sous-espace $\mathcal{S} = \text{span}\{u_1, ..., u_D\}$ de $\mathcal{E}$ de dimension finie. Nous discuterons les cas suivants :

- $\mathcal{S}$ est connu.
- Les $u_i$ sont connus mais $D$ est inconnu.
- Les $u_i$ sont inconnus.

# 4 Applications

- Nous illustrerons l'intérêt de nos estimateurs à travers des applications sur des données simulées.
- Une application à la prédiction spatiales sera par également discutée.

## Bibliographie

[1] Biau, G. and Cadre, B. (2004). Nonparametric Spatial Prediction, *Statistical inference for Stochastic Processes*, **Vol. 7**, pp. 327-349.
[2] Dabo-Niang, S. and Yao, A.F. (2007). Kernel regression estimation for continuous spatial processes, *Mathematical Methods of Statistics.*, **16**, pp. 298-317.
[3] Hallin, M., Lu, Z. and Yu, K. (2009). Local linear spatial quantile regression. Bernoulli 15, 659-686.
[4] Hallin, M., Lu, and Tran, L.T. (2004). Local linear spatial regression, *Annal of Statistics*, **Vol. 32**, **No. 6**, pp. 2469-2500.
[5]Lu, Z., and Chen, X. 2004. Spatial kernel regression estimation : Weak concistency. Stat. proba. lett, 68, 125–136.