



HAL
open science

Data-driven neighborhood selection of a Gaussian field

Nicolas Verzelen

► **To cite this version:**

Nicolas Verzelen. Data-driven neighborhood selection of a Gaussian field. [Research Report] RR-6798, 2009, pp.25. inria-00353260v1

HAL Id: inria-00353260

<https://inria.hal.science/inria-00353260v1>

Submitted on 15 Jan 2009 (v1), last revised 1 Sep 2009 (v2)

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.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Data-driven neighborhood selection of a Gaussian field

Nicolas Verzelen

N° 6798

Janvier 2009

Thème COG

*R*apport
de recherche

Data-driven neighborhood selection of a Gaussian field

Nicolas Verzelen * †

Thème COG — Systèmes cognitifs
Équipes-Projets Select

Rapport de recherche n° 6798 — Janvier 2009 — 25 pages

Abstract: We study the non-parametric covariance estimation of a stationary Gaussian field X observed on a lattice. To tackle this issue, we have introduced a model selection procedure in a previous paper [Ver09]. This procedure amounts to selecting a neighborhood \hat{m} by a penalization method and estimating the covariance of X in the space of Gaussian Markov random fields (GMRFs) with neighborhood \hat{m} . Such a strategy is shown to satisfy oracle inequalities as well as minimax adaptive properties. However, it suffers several drawbacks which make the method difficult to apply in practice. On the one hand, the penalty depends on some unknown quantities. On the other hand, the procedure is only defined for toroidal lattices. Our contribution is threefold. We propose a data-driven algorithm for tuning the penalty function. Moreover, we extend the procedure to non-toroidal lattices. Finally, we study the numerical performances of this new method on simulated examples. These simulations suggest that Gaussian Markov random field selection is often a good alternative to variogram estimation.

Key-words: Gaussian field, Gaussian Markov random field, Data-driven calibration, model selection, pseudolikelihood.

* Laboratoire de Mathématiques UMR 8628, Université Paris-Sud, 91405 Osay

† INRIA Futurs, Projet SELECT, Université Paris-Sud, 91405 Osay

Sélection automatique de voisinage d'un champ gaussien

Résumé : Nous étudions l'estimation non-paramétrique d'un champ gaussien stationnaire X observé sur un réseau régulier. Dans ce cadre, nous avons précédemment introduit une procédure de sélection de modèle [Ver09]. Cette procédure revient à sélectionner un voisinage \hat{m} grâce une technique de pénalisation puis à estimer la covariance du champ X dans l'espace des champs de Markov gaussiens de voisinage \hat{m} . Une telle stratégie satisfait des inégalités oracles et des propriétés d'adaptation au sens minimax. En pratique, elle présente néanmoins quelques inconvénients. D'une part, la pénalité dépend de quantités inconnues. D'autre part, la procédure est uniquement définie pour des réseaux toriques. La contribution de cet article est triple. Nous proposons un algorithme automatique pour calibrer la pénalité. De plus, nous introduisons une extension à des réseaux non-toriques. Enfin, nous étudions les performances pratiques de la procédure sur des données simulées. Ces simulations suggèrent que la sélection de champs de Markov gaussiens est souvent une bonne alternative à l'estimation de variogramme.

Mots-clés : Champ gaussien, champ de Markov gaussien, calibration automatique, sélection de modèle, pseudo-vraisemblance.

1 Introduction

In this paper, we study the estimation of the distribution of a stationary Gaussian field $(X_{[i,j]})_{(i,j) \in \Lambda}$ indexed by the nodes of a rectangular lattice Λ of size $p_1 \times p_2$. This problem is often encountered in spatial statistics or in image analysis. Most statistical procedures amount to first subtracting a parametric form of the mean value. Henceforth, we assume that the field X is centered. Given a n -sample of the field X , the challenge is to infer the distribution of the field. In practice, the number n of observations often equals one. Different methods have been proposed to tackle this problem.

A first approach amounts to computing an empirical variogram and then fitting a suitable parametric variogram model such as the exponential or Matérn model (see [Cre93] Ch.2 for more details). If such methods perform well if one has chosen a *good* variogram model, it is not adaptive to any form of correlation.

A second approach amounts to estimating the covariance of X in a Gaussian Markov random field (GMRF) setting. Let us fix a node $(0,0)$ at the center of Λ and let m be a subset of $\Lambda \setminus \{(0,0)\}$. The field X is a GMRF with respect to the neighborhood m if conditionally to $(X_{[k,l]})_{(k,l) \in m}$, the variable $X_{[0,0]}$ is independent from all the remaining variables in Λ . We refer to Rue and Held [RH05] for a comprehensive introduction on GMRFs. If we know that X is a GMRF with respect to the neighborhood m , then we can estimate the covariance by applying likelihood or pseudolikelihood maximization. Such parametric procedures are well understood, at least from an asymptotic point of view (see for instance [Guy95] Sect.4). In practice, we do not know which neighborhood m we should choose. For instance, choosing the empty neighborhood amounts to assuming that all the components of X are independent. Alternatively, if we choose the complete neighborhood, which contains all the nodes of Λ except $(0,0)$, then the number of parameters is huge and estimation performances are poor.

In this paper, we tackle the problem of neighborhood selection from a practical point of view. The purpose is to define a data-driven procedure that picks a suitable neighborhood \hat{m} and then estimates the distribution of X in the space of GMRFs with neighborhood \hat{m} . This procedure should not require any knowledge on the distribution of X . Moreover, it has to achieve a fast rate of convergence if the distribution of X is a GMRF with respect to a small neighborhood. In statistical terms, the procedure has to be non-parametric and adaptive.

Besag and Kooperberg [BK95], Rue and Tjelmeland [RT02], Song *et al.* [SFG08], and Cressie and Verzelen [CV08] have considered the problem of *approximating* the distribution of a Gaussian field by a GMRF, but this approach requires the knowledge of the true distribution. Guyon and Yao have stated in [GY99] necessary conditions and sufficient conditions for a model selection procedure to choose asymptotically the true neighborhood of a GMRF with probability one. Our point of view is slightly different: we do not assume that the field X is a GMRF with respect to a sparse neighborhood and do not aim at estimating the true neighborhood, we rather want to select a neighborhood that allows to estimate *well* the distribution of X . The distinction between these two points of view has been nicely described in the first chapter of MacQuarrie and Tsai [MT98].

In [Ver09], we have introduced a neighborhood selection procedure based on pseudolikelihood maximization and penalization. Under mild assumptions, the procedure achieves optimal neighborhood selection. More precisely, it satisfies an oracle inequality and it is minimax adaptive to the sparsity of the neighborhood. To our knowledge, these are the first results of neighborhood selection in this spatial setting.

If the procedure exhibits appealing theoretical properties, it suffers several drawbacks from a practical perspective. First, the method constrains the largest eigenvalue of the estimated covariance to be smaller than some parameter ρ . In practice, it is difficult to choose ρ since we do not know the largest eigenvalue of the true covariance. Second, the penalty function $\text{pen}(\cdot)$ introduced in [Ver09] Sect.3 depends on the largest eigenvalue of the covariance of the field X . Hence, we need a practical method for tuning the penalty. Third, in [Ver09] we have only defined the procedure when the lattice Λ is a square torus.

Our contribution is twofold. On the one hand, we propose practical versions of our neighborhood selection procedure that overcome the previously-mentioned drawbacks:

- The procedure is extended to rectangular lattices.
- We do not constrain anymore the largest eigenvalue of the covariance.
- We provide an algorithm based on the so-called *slope heuristics* of Birgé and Massart [BM07] for tuning the penalty. Theoretical justifications for its use are also given.
- Finally, we extend the procedure to the case where the lattice Λ is not a torus.

On the other hand, we illustrate the performances of this new procedure on numerical examples. When Λ is a torus, we compare it with likelihood-based methods like AIC [Aka73] and BIC [Sch78], even if they were not studied in this setting. When Λ is not toroidal, likelihood methods become intractable. Nevertheless, our procedure still applies and often outperforms variogram-based methods.

The paper is organized as follows. In Section 2, we define a new version of the estimation procedure of [Ver09] that does not require anymore the choice of the constant ρ . We also discuss the computational complexity of the procedure. In Section 3, we connect this new procedure to the method introduced in [Ver09] and we recall some theoretical results. We provide an algorithm for tuning the penalty in practice in Section 4. In Section 5, we extend our procedure for handling non-toroidal lattices. This simulation studies are provided in Section 6. Section 7 summarizes our findings, while the proofs are postponed to Section 8.

Let us introduce some notations. In the sequel, X^v refers to the vectorialized version of X with the convention $X_{[i,j]} = X^v_{[(i-1) \times p_2 + j]}$ for any $1 \leq i \leq p_1$ and $1 \leq j \leq p_2$. Using this new notation amounts to “forgetting” the spatial structure of X and allows to get into a more classical statistical framework. We note $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ the n observations of the field X . The matrix Σ stands for the covariance matrix of X^v . For any matrix A , $\varphi_{\max}(A)$ and $\varphi_{\min}(A)$ respectively refer the largest eigenvalue and the smallest eigenvalues of A . Finally, I_r denotes the identity matrix of size r .

2 Neighborhood selection on a torus

In this section, we introduce the main concepts and notations for GMRFs on a torus. Afterwards, we describe our procedure based on pseudolikelihood maximization. Finally, we discuss some computational aspects. Throughout this section and the two following sections, the lattice Λ is assumed to be toroidal. Consequently, the components of the matrices X are taken modulo p_1 and p_2 .

2.1 GMRFs on the torus

The notion of conditional distribution is underlying the definition of GMRFs. By standard Gaussian derivations (see for instance [Lau96] App.C), there exists a unique $p_1 \times p_2$ matrix θ such that $\theta_{[0,0]} = 0$ and

$$X_{[0,0]} = \sum_{(i,j) \in \Lambda \setminus \{(0,0)\}} \theta_{[i,j]} X_{[i,j]} + \epsilon_{[0,0]} , \quad (1)$$

where the random variable $\epsilon_{[0,0]}$ follows a zero-mean normal distribution and is independent from the covariates $(X_{[i,j]})_{(i,j) \in \Lambda \setminus \{(0,0)\}}$. In the sequel, we note σ^2 the variance of $\epsilon_{[0,0]}$ and we call it the conditional variance of $X_{[0,0]}$.

Equation (1) describes the conditional distribution of $X_{[0,0]}$ given the remaining variables. By stationarity of the field X , it holds that $\theta_{[i,j]} = \theta_{[-i,-j]}$. The covariance matrix Σ is closely related to θ through the following equation:

$$\Sigma = \sigma^2 [I_{p_1 p_2} - C(\theta)]^{-1} , \quad (2)$$

where the $p_1 p_2 \times p_1 p_2$ matrix $C(\theta)$ is defined by $C(\theta)_{[(i_1-1)p_2+j_1, (i_2-1)p_2+j_2]} := \theta_{[i_2-i_1, j_2-j_1]}$ for any $1 \leq i_1, i_2 \leq p_1$ and $1 \leq j_1, j_2 \leq p_2$. The matrix $(I_{p_1 p_2} - C(\theta))$ is called the partial correlation matrix of the field X . The so-defined matrix $C(\theta)$ is symmetric block circulant with $p_2 \times p_2$ blocks. We refer to [RH05] Sect.2.6 or the book of Gray [Gra06] for definitions and main properties on circulant and block circulant matrices.

Identities (1) and (2) have two main consequences. First, estimating the $p_1 \times p_2$ matrix θ amounts to estimating the covariance matrix Σ up to a multiplicative constant. We shall therefore focus on θ . Second, by Equation (1), the field X is a GMRF with respect to the neighborhood defined by the support θ . The adaptive estimation issue of the distribution of X by neighborhood selection therefore reformulates as an adaptive estimation problem of the matrix θ via support selection.

Let us now precise the set of possible values for θ . The set Θ denotes the vector space of the $p_1 \times p_2$ matrices that satisfy $\theta_{[0,0]} = 0$ and $\theta_{[i,j]} = \theta_{[-i,-j]}$, for any $(i, j) \in \Lambda$. Hence, a matrix $\theta \in \Theta$ corresponds to the distribution of a stationary Gaussian field if and only if the $p_1 p_2 \times p_1 p_2$ matrix $(I_{p_1 p_2} - C(\theta))$ is positive definite. This is why we define the convex subset Θ^+ of Θ by

$$\Theta^+ := \{\theta \in \Theta \text{ s.t. } [I_{p_1 p_2} - C(\theta)] \text{ is positive definite}\} . \quad (3)$$

The set of covariance matrices of stationary Gaussian fields on Λ with unit conditional variance is in one to one correspondence with the set Θ^+ . We sometimes assume that the field X is isotropic. The corresponding sets Θ^{iso} and $\Theta^{+, \text{iso}}$ for isotropic fields are introduced as:

$$\Theta^{\text{iso}} := \{\theta \in \Theta , \theta_{[i,j]} = \theta_{[-i,j]} = \theta_{[j,i]} , \forall (i, j) \in \Lambda\} \quad \text{and} \quad \Theta^{+, \text{iso}} := \Theta^+ \cap \Theta^{\text{iso}} .$$

2.2 Description of the procedure

Let $|(i, j)|_t$ refer to the toroidal norm defined by

$$|(i, j)|_t^2 := [i \wedge (p_1 - i)]^2 + [j \wedge (p_2 - j)]^2 ,$$

for any node $(i, j) \in \Lambda$.

In the sequel, a model m stands for a subset of $\Lambda \setminus \{(0, 0)\}$. It is also called a neighborhood. For the sake of simplicity, we shall only use the collection of models \mathcal{M}_1 defined below.

Definition 2.1. A subset $m \subset \Lambda \setminus \{(0,0)\}$ belongs to \mathcal{M}_1 if and only if there exists a number $r_m > 1$ such that

$$m = \{(i, j) \in \Lambda \setminus \{(0,0)\} \mid |(i, j)|_t \leq r_m\} . \quad (4)$$

In other words, the neighborhoods m in \mathcal{M}_1 are sets of nodes lying in a disc centered at $(0,0)$. Obviously, \mathcal{M}_1 is totally ordered with respect to the inclusion. Consequently, we order the models $m_0 \subset m_1 \subset \dots \subset m_i \dots$. For instance, m_0 corresponds to the empty neighborhood, m_1 stands for the neighborhood of size 4, and m_2 refers to the neighborhood with 8 neighbours. See Figure 1 for an illustration.

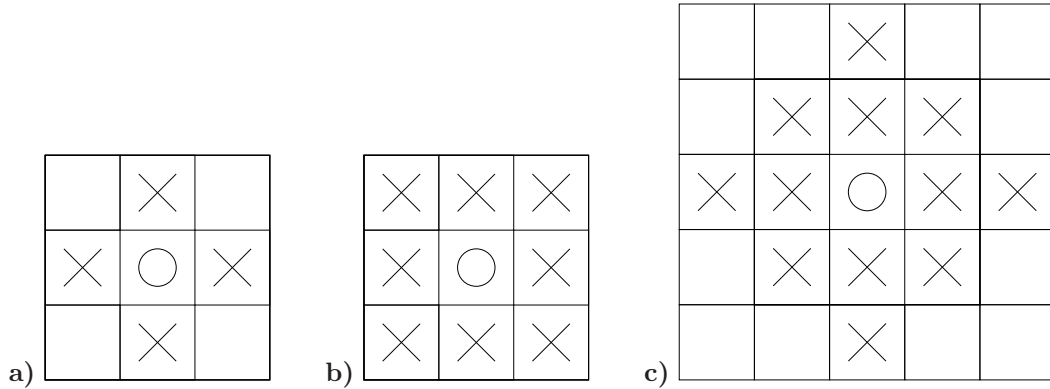


Figure 1: (a) Model m_1 with first order neighbors. (b) Model m_2 with second order neighbors. (c) Model m_3 with third order neighbors.

For any model $m \in \mathcal{M}_1$, the vector space Θ_m is the subset of matrices Θ whose support is included in m . Similarly Θ_m^{iso} is the subset of Θ^{iso} whose support is included in m . The dimensions of Θ_m and Θ_m^{iso} are respectively noted d_m and d_m^{iso} . Since we aim at estimating the positive matrix $(I_{p_1 p_2} - C(\theta))$, we also consider the convex subsets of Θ_m^+ and $\Theta_m^{+, \text{iso}}$ which correspond to non-negative matrices precision matrices.

$$\Theta_m^+ := \Theta_m \cap \Theta^+ \quad \text{and} \quad \Theta_m^{+, \text{iso}} := \Theta_m^{\text{iso}} \cap \Theta^{+, \text{iso}} . \quad (5)$$

For any $\theta' \in \Theta^+$, the conditional least-squares (CLS) criterion ([Guy87]) $\gamma_{n, p_1, p_2}(\theta')$ is defined by

$$\gamma_{n, p_1, p_2}(\theta') := \frac{1}{n p_1 p_2} \sum_{i=1}^n \sum_{(j_1, j_2) \in \Lambda} \left(\mathbf{X}_i[j_1, j_2] - \sum_{(l_1, l_2) \in \Lambda \setminus \{(0,0)\}} \theta'_{[l_1, l_2]} \mathbf{X}_i[j_1 + l_1, j_2 + l_2] \right)^2 . \quad (6)$$

The function $\gamma_{n, p_1, p_2}(\cdot)$ is a least squares criterion that allows us to perform the simultaneous linear regression of all $\mathbf{X}_i[j_1, j_2]$ with respect to the covariates $(\mathbf{X}_i[l_1, l_2])_{(l_1, l_2) \neq (k_1, k_2)}$. The advantage of this criterion is that it does not require the computation of a determinant as for the likelihood. We explain in [Ver09] Sect.2 its connection with the pseudolikelihood introduced by Besag [Bes75]. For any model $m \in \mathcal{M}_1$, the estimators are defined as the unique minimizers of $\gamma_{n, p_1, p_2}(\cdot)$ on the sets Θ_m^+ and $\Theta_m^{+, \text{iso}}$.

$$\hat{\theta}_m := \arg \min_{\theta' \in \Theta_m^+} \gamma_{n, p_1, p_2}(\theta') \quad \text{and} \quad \hat{\theta}_m^{\text{iso}} := \arg \min_{\theta' \in \Theta_m^{+, \text{iso}}} \gamma_{n, p_1, p_2}(\theta') , \quad (7)$$

where \bar{A} stands for the closure of A . Let us mention that the estimator $\hat{\theta}_m$ corresponds to the estimator $\hat{\theta}_{m,\rho_1}$ defined in [Ver09] Sect.2.2 with $\rho_1 = +\infty$. We further discuss the connection between $\hat{\theta}_m$ and $\hat{\theta}_{m,\rho_1}$ in Section 3.

Given a subcollection of models \mathcal{M} of \mathcal{M}_1 and a positive function $\text{pen} : \mathcal{M} \rightarrow \mathbb{R}^+$ called a penalty, we select a model as follows:

$$\hat{m} := \arg \min_{m \in \mathcal{M}} \left[\gamma_{n,p_1,p_2}(\hat{\theta}_m) + \text{pen}(m) \right] \quad \text{and} \quad \hat{m}^{\text{iso}} := \arg \min_{m \in \mathcal{M}} \left[\gamma_{n,p_1,p_2}(\hat{\theta}_m^{\text{iso}}) + \text{pen}(m) \right]. \quad (8)$$

For short, we write $\tilde{\theta}$ and $\tilde{\theta}^{\text{iso}}$ for $\hat{\theta}_{\hat{m}}$ and $\hat{\theta}_{\hat{m}^{\text{iso}}}^{\text{iso}}$. We discuss the choice of the penalty function in Section 4.

2.3 Computational aspects

Since the lattice Λ is a torus, the computation of the estimators $\hat{\theta}_m$ is performed efficiently thanks to the following lemma.

Lemma 2.1. *For any $p \times p$ matrix A and for any $1 \leq i \leq p_1$ and $1 \leq j \leq p_2$, $\lambda_{[i,j]}(A)$ is the (i, j) -th term of two-dimensional discrete Fourier transform of the matrix A :*

$$\lambda_{[i,j]}(A) := \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} A_{[i,j]} \exp \left[2i\pi \left(\frac{ki}{p_1} + \frac{jl}{p_2} \right) \right], \quad (9)$$

where $i^2 = -1$. The conditional least-squares criterion $\gamma_{n,p_1,p_2}(\theta')$ simplifies as

$$\gamma_{n,p_1,p_2}(\theta') = \frac{1}{np_1^2 p_2^2} \left\{ \sum_{i=1}^{p_1} \sum_{j=1}^{p_2} [1 - \lambda_{[i,j]}(\theta)]^2 \left[\sum_{k=1}^n \lambda_{[i,j]}(\mathbf{X}_k) \overline{\lambda_{[i,j]}(\mathbf{X}_k)} \right] \right\}.$$

A proof is given in Section 8. Optimization of $\gamma_{n,p_1,p_2}(\cdot)$ over the set Θ_m^+ is performed fastly using the fast Fourier transform (FFT). Nevertheless, this is not the privilege of CLS estimators, since maximum likelihood estimators are also computed fastly by FFT when Λ is a torus.

In Section 5, we mention that the computation of the CLS estimators $\hat{\theta}_m$ remains quite easy when Λ is not a torus whereas likelihood maximization becomes intractable.

3 Theoretical results

Throughout this section, Λ is assumed to be a toroidal square lattice and we note p its size. Let us mention that the restriction to square lattices made in [Ver09] allows to simplify the proofs but is not necessary so that the theoretical results hold. In this section, we first recall the procedure introduced in [Ver09] and we emphasize the differences with the one defined in the previous section. We also mention a result of optimality stated in [Ver09]. This will provide some insights for calibrating the penalty $\text{pen}(\cdot)$ in Section 4.

Given $\rho > 2$ be a positive constant, we define the subsets $\Theta_{m,\rho}^+$ and $\Theta_{m,\rho}^{+,\text{iso}}$ by

$$\begin{aligned} \Theta_{m,\rho}^+ &:= \{ \theta \in \Theta_m^+, \varphi_{\max} [I_{p_1 p_2} - C(\theta)] < \rho \} \\ \Theta_{m,\rho}^{+,\text{iso}} &:= \{ \theta \in \Theta_m^{+,\text{iso}}, \varphi_{\max} [I_{p_1 p_2} - C(\theta)] < \rho \}. \end{aligned} \quad (10)$$

Then, the corresponding estimators $\widehat{\theta}_{m,\rho}$ and $\widehat{\theta}_{m,\rho}^{\text{iso}}$ are defined as in (7), except that we now consider $\Theta_{m,\rho}^+$ instead of Θ_m^+ .

$$\widehat{\theta}_{m,\rho} := \arg \min_{\theta' \in \Theta_{m,\rho}^+} \gamma_{n,p,p}(\theta') \quad \text{and} \quad \widehat{\theta}_{m,\rho}^{\text{iso}} := \arg \min_{\theta' \in \Theta_{m,\rho}^{+, \text{iso}}} \gamma_{n,p,p}(\theta') .$$

Given a subcollection \mathcal{M} of \mathcal{M}_1 and a penalty function $\text{pen}(\cdot)$, we select the models \widehat{m}_ρ and $\widehat{m}_\rho^{\text{iso}}$ as in (8) except that we use $\widehat{\theta}_{m,\rho}$ and $\widehat{\theta}_{m,\rho}^{\text{iso}}$ instead of $\widehat{\theta}_m$ and $\widehat{\theta}_m^{\text{iso}}$. We also note $\widetilde{\theta}_\rho$ and $\widetilde{\theta}_\rho^{\text{iso}}$ for $\widehat{\theta}_{\widehat{m}_\rho,\rho}$ and $\widehat{\theta}_{\widehat{m}_\rho^{\text{iso}},\rho}^{\text{iso}}$.

The only difference between the estimators $\widetilde{\theta}$ and $\widetilde{\theta}_\rho$ is that the largest eigenvalue of the precision matrix $(I_{p^2} - C(\widetilde{\theta}))$ is restricted to be smaller than ρ . We make this restriction in [Ver09] to facilitate the analysis.

In order to assess the performance of the penalized estimator $\widetilde{\theta}_\rho$ and $\widetilde{\theta}_\rho^{\text{iso}}$, we use the prediction loss function $l(\theta_1, \theta_2)$ defined by

$$l(\theta_1, \theta_2) := \frac{1}{p^2} \text{tr} [(C(\theta_1) - C(\theta_2))\Sigma(C(\theta_1) - C(\theta_2))] . \quad (11)$$

As explained in [Ver09] Sect.1.3, the loss $l(\theta_1, \theta_2)$ expresses in terms of conditional expectation

$$l(\theta_1, \theta_2) = \mathbb{E}_\theta \left\{ \left[\mathbb{E}_{\theta_1} (X_{[0,0]} | X_{\Lambda \setminus \{0,0\}}) - \mathbb{E}_{\theta_2} (X_{[0,0]} | X_{\Lambda \setminus \{0,0\}}) \right]^2 \right\} , \quad (12)$$

where $\mathbb{E}_\theta(\cdot)$ stands for the expectation with respect to the distribution $\mathcal{N}(0, \sigma^2(I_{p_1 p_2} - C(\theta))^{-1})$. Hence, $l(\widetilde{\theta}, \theta)$ corresponds the mean squared prediction loss of $X_{[0,0]}$ given the other covariates. A similar loss function is also used by Song *et al.* [SFG08], when approximation Gaussian fields by GMRFs. For any neighborhood $m \in \mathcal{M}$, we define the *projection* $\theta_{m,\rho}$ as the closest element of θ in $\Theta_{m,\rho}^+$ with respect to the loss $l(\cdot, \cdot)$.

$$\theta_{m,\rho} := \arg \min_{\theta' \in \Theta_{m,\rho}^+} l(\theta', \theta) \quad \text{and} \quad \theta_{m,\rho}^{\text{iso}} := \arg \min_{\theta' \in \Theta_{m,\rho}^{+, \text{iso}}} l(\theta', \theta) .$$

We call the loss $l(\theta_{m,\rho}, \theta)$ the bias of the set $\Theta_{m,\rho}^+$. This implies that $\widehat{\theta}_{m,\rho}$ cannot perform better than this loss.

Theorem 3.1. *Let $\rho > 2$, K be a positive number larger than an universal constant K_0 and \mathcal{M} be a subcollection of \mathcal{M}_1 . If every model $m \in \mathcal{M}$, it holds that*

$$\text{pen}(m) \geq K \rho^2 \varphi_{\max}(\Sigma) \frac{d_m + 1}{np^2} , \quad (13)$$

then for any $\theta \in \Theta^+$, the estimator $\widetilde{\theta}_\rho$ satisfies

$$\mathbb{E}_\theta[l(\widetilde{\theta}_\rho, \theta)] \leq L(K) \inf_{m \in \mathcal{M}} [l(\theta_{m,\rho}, \theta) + \text{pen}(m)] , \quad (14)$$

where $L(K)$ only depends on K . A similar bound holds if one replaces $\widetilde{\theta}_\rho$ by $\widetilde{\theta}_\rho^{\text{iso}}$, Θ^+ by $\Theta^{+, \text{iso}}$, $\theta_{m,\rho}$ by $\theta_{m,\rho}^{\text{iso}}$, and d_m by d_m^{iso} .

This nonasymptotic bound is provided in a slightly different version in [Ver09]. It states that $\tilde{\theta}_\rho$ achieves a trade-off between the bias and a variance term if the penalty is suitable chosen. In Theorem 3.1, we use the penalty $K\rho^2\varphi_{\max}(\Sigma)\frac{d_m+1}{np^2}$ instead of the penalty $K\rho^2\varphi_{\max}(\Sigma)\frac{d_m}{np^2}$ stated in [Ver09]. This makes the bound (14) simpler. Observe that these two penalties yield the same model selection since they only differ by a constant. Let us further discuss two points.

- In this paper, we use the estimator $\tilde{\theta}$ rather than $\tilde{\theta}_\rho$. Given a collection of models \mathcal{M} , there exists some finite $\rho > 2$, such that these two estimators coincide. Take for instance $\rho = \sup_{m \in \mathcal{M}} \sup_{\theta \in \Theta_m^+} \varphi_{\max}(I_{p_1 p_2} - C(\theta))$. Admittedly, the so-obtained ρ may be large, especially if there are large models in \mathcal{M} . The upper bound (14) on the risk therefore becomes worse. Nevertheless, we do not think that the dependency of (14) on ρ is sharp. Indeed, we illustrate in Section 6 that the risk of $\tilde{\theta}$ exhibits good statistical performances..
- Theorem 3.1 provides a suitable form of the penalty for obtaining oracle inequalities. However, this penalty depends on $\varphi_{\max}(\Sigma)$ which is not known in practice. This is why we develop a data-driven penalization method in the next section.

4 Slope Heuristics

Let us introduce a data-driven method for calibrating the penalty function $\text{pen}(\cdot)$. It is based on the so-called *slope heuristic* introduced by Birgé and Massart [BM07] in the fixed design Gaussian regression framework (see also [Mas07] Sect.8.5.2). This heuristic relies on the notion of minimal penalty. In short, assume that one knows that a good penalty has a form $\text{pen}(m) = NF(d_m)$ (where d_m is the dimension of the model and N is a tuning parameter). Let us define $\hat{m}(N)$ the selected model as a function of N . There exists a quantity \hat{N}_{\min} satisfying the following property: If $N > \hat{N}_{\min}$, the dimension of the selected model $d_{\hat{m}(N)}$ is reasonable and if $N < \hat{N}_{\min}$, the dimension of the selected model is huge. The function $\text{pen}_{\min}(\cdot) := \hat{N}_{\min}F(\cdot)$ is called the minimal penalty. In fact, a *dimension jump* occurs for $d_{\hat{m}(N)}$ at the point \hat{N}_{\min} . Thus, the quantity \hat{N}_{\min} is clearly observable for real data sets. In their Gaussian framework, Birgé and Massart have shown that twice the minimal penalty is nearly the optimal penalty. In other words, the model $\hat{m} := \hat{m}(2\hat{N}_{\min})$ yields an efficient estimator.

The slope heuristic method has been successfully applied for multiple change-point detection [Leb05]. Applications are also being developed in other frameworks such as mixture models [MM08], clustering [BCM08], estimation of oil reserves [Lep02], and genomic [Vil07].

If this method was originally introduced for fixed design Gaussian regression, Arlot and Massart [AM09] have proved more recently that a similar phenomenon occurs in the heteroscedastic random-design case. In the GMRF setting, we are only able to partially justify this heuristic. For the sake of simplicity, let us assume in the next proposition that the lattice Λ is a square of size p .

Proposition 4.1. *Consider $\rho > 2$, and $\eta < 1$ and suppose that p is larger than some numerical constant p_0 . Let m' be the largest model in \mathcal{M}_1 that satisfies $d_{m'} \leq \sqrt{np^2}$.*

For any model $m \in \mathcal{M}_1$, we assume that

$$\text{pen}(m') - \text{pen}(m) \leq K_1(1 - \eta)\sigma^2 \left\{ \varphi_{\min}(I_{p^2} - C(\theta)) \wedge [\rho - \varphi_{\max}(I_{p^2} - C(\theta))] \right\} \frac{d_{m'} - d_m}{np^2}, \quad (15)$$

where K_1 is a universal constant defined in the proof. Then, for any $\theta \in \Theta_{m', \rho}^+$, it holds that

$$\mathbb{P} \left\{ d_{\hat{m}_\rho} > L \left[\sqrt{np^2} \wedge p^2 \right] \right\} \geq \frac{1}{2},$$

where L only depends on η , ρ , $\varphi_{\min}(I_{p^2} - C(\theta))$, and $\varphi_{\max}(I_{p^2} - C(\theta))$.

The proof is postponed to Section 8. Let us define

$$N_1 := K_1 \sigma^2 \{ \varphi_{\min}(I_{p_1 p_2} - C(\theta)) \wedge [\rho - \varphi_{\max}(I_{p_1 p_2} - C(\theta))] \},$$

and let us consider penalty functions $\text{pen}(m) = N \frac{d_m}{np_1 p_2}$ for some $N > 0$. The proposition states that if N is smaller than N_1 , then the procedure selects a model of huge dimension with large probability, i.e $d_{\widehat{m}(N)}$ is huge. Alternatively, let us define

$$N_2 := K_0 \frac{\sigma^2 \rho^2}{\varphi_{\min}(I_{p_1 p_2} - C(\theta))} \frac{d_m}{np_1 p_2},$$

where the numerical constant K_0 is introduced in Theorem 3.1 in [Ver09]. By Theorem 3.1, choosing $N > N_2$ ensures that the risk of $\widetilde{\theta}_\rho$ achieves a type-oracle inequality and the dimension $d_{\widehat{m}_\rho(N)}$ is reasonable. The quantities N_1 and N_2 are different especially when the eigenvalues of $(I_{p_1 p_2} - C(\theta))$ are far from 1. Since we do not know the behavior of the selected model $\widehat{m}_\rho(N)$ when N is between N_1 and N_2 , we are not able to really prove a dimension jump as the fixed design Gaussian regression framework. Besides, we have mentioned in the preceding section that we are more interested to the estimator $\widetilde{\theta}$ than $\widetilde{\theta}_\rho$. Nevertheless, we clearly observe in simulation studies a dimension jump for some N between N_1 and N_2 even if we use the estimators $\widehat{\theta}_m$ instead of $\widehat{\theta}_{m,\rho}$. This lets us think that the slope heuristic is still valid in the GMRF framework.

Algorithm 4.1. (*Data-driven penalization with slope heuristic*). Let \mathcal{M} be a subcollection of \mathcal{M}_1 .

1. Compute the selected model $\widehat{m}(N)$ as a function of $N > 0$

$$\widehat{m}(N) \in \arg \min_{m \in \mathcal{M}} \left\{ \gamma_{n,p_1,p_2}(\widehat{\theta}_m) + N \frac{d_m}{np_1 p_2} \right\}.$$

2. Find $\widehat{N}_{\min} > 0$ such that the jump $d_{\widehat{m}([\widehat{N}_{\min}]_-)} - d_{\widehat{m}([\widehat{N}_{\min}]_+)}$ is maximal.

3. Select the model $\widehat{m} = \widehat{m}(2\widehat{N}_{\min})$.

The difference $f(x_-) - f(x_+)$ measures the discontinuity of a function f at the point x . Step 2 may need to introduce huge models in the collection \mathcal{M} all the other ones being considered as “reasonably small”. As the function $\widehat{m}(\cdot)$ is piecewise linear with at most $\text{Card}(\mathcal{M})$ jumps, so that steps 1-2 have a complexity $\mathcal{O}(\text{Card}(\mathcal{M}))^2$. We refer to App.A.1 of [AM09] for more details on the computational aspects of steps 1 and 2. Let us mention that there are other ways of estimating \widehat{N}_{\min} than choosing the largest jump as described in [AM09] App.A.2. Finally, the methodology described in this section straightforwardly extends to the case of isotropic GMRFs estimation by replacing $\widehat{m}(N)$ by $\widehat{m}^{\text{iso}}(N)$ and d_m by d_m^{iso} .

Algorithm 4.2. (*Data-driven penalization for isotropic models*). Let \mathcal{M} be a subcollection of \mathcal{M}_1 .

1. Compute the selected model $\widehat{m}^{\text{iso}}(N)$ as a function of $N > 0$

$$\widehat{m}^{\text{iso}}(N) \in \arg \min_{m \in \mathcal{M}} \left\{ \gamma_{n,p_1,p_2}(\widehat{\theta}_m^{\text{iso}}) + N \frac{d_m^{\text{iso}}}{np_1 p_2} \right\}.$$

2. Find $\widehat{N}_{\min}^{\text{iso}} > 0$ such that the jump $d_{\widehat{m}^{\text{iso}}}([\widehat{N}_{\min}^{\text{iso}}]_-) - d_{\widehat{m}^{\text{iso}}}([\widehat{N}_{\min}^{\text{iso}}]_+)$ is maximal.
3. Select the model $\widehat{m}^{\text{iso}} = \widehat{m}^{\text{iso}}(2\widehat{N}_{\min}^{\text{iso}})$.

In conclusion, the neighborhood selection procedure described in Algorithm 4.1 is completely data-driven and does not require any prior knowledge on the matrix Σ . Moreover, its computational burden remains small. We illustrate its efficiency in Section 6.

5 Extension to non-toroidal lattices

It is often artificial to consider the field X as stationary on a torus. However, we needed this hypothesis for deriving nonasymptotic properties of the estimator $\widehat{\theta}$ in [Ver09]. In many applications, it is more realistic to assume that we observe a small window of a Gaussian field defined on the plane \mathbb{Z}^2 . If we are unable to prove nonasymptotic risk bounds in this new setting. Nevertheless, Lakshman and Derin have shown in [LD93] that there is no phase transition within the valid parameter space for GMRFs defined on the plane \mathbb{Z}^2 . Let us briefly explain what this means: consider a GMRF defined on a square lattice of size p , but only observed on a square lattice of size p' . The absence of phase transition implies the distribution of this field observed on this fixed window of size p' does not asymptotically depend on the bound conditions when p goes to infinity. Consequently, it is reasonable to think that our estimation procedure still performs well to the price of slight modifications. In the sequel, we assume that the field X is defined on \mathbb{Z}^2 , but the data \mathbf{X} still correspond to n independent observations of the field X on the window Λ of size $p_1 \times p_2$. The conditional distribution of $X_{[0,0]}$ given the remaining covariates now decomposes as

$$X_{[0,0]} = \sum_{(i,j) \in \mathbb{Z}^2 \setminus \{(0,0)\}} \theta_{[i,j]} X_{[i,j]} + \epsilon_{[0,0]}, \quad (16)$$

where $\theta_{[.,.]}$ is an “infinite” matrix defined on \mathbb{Z}^2 and where $\epsilon_{[0,0]}$ is a centered Gaussian variable of variance σ^2 independent of $(X_{[i,j]})_{(i,j) \in \Lambda \setminus \{(0,0)\}}$. The distribution of the field X is uniquely defined by the function θ and positive number σ^2 . The set $\Theta^{+, \infty}$ of valid parameter for θ is now defined using the spectral density function. We refer to Rue and Held [RH05] Sect.2.7 for more details.

Definition 5.1. A function $\theta : \mathbb{Z}^2 \rightarrow \mathbb{R}$ belongs to the set $\Theta^{+, \infty}$ if and only if it satisfies the three following conditions:

1. $\theta_{[0,0]} = 0$.
2. For any $(i, j) \in \mathbb{Z}^2$, $\theta_{[i,j]} = \theta_{[-i, -j]}$.
3. For any $(\omega_1, \omega_2) \in [0, 2\pi)^2$, $1 - \sum_{(i,j) \in \mathbb{Z}^2} \theta_{[i,j]} \cos(i\omega_1 + j\omega_2) > 0$.

Similarly, we define the set $\Theta^{+, \infty, \text{iso}}$ for the isotropic GMRFs on the lattices. As done in Section 2 for toroidal lattices, we now introduce the parametric parameter sets. For any model $m \in \mathcal{M}_1$, the set $\Theta_m^{+, \infty}$ refers to the subset of matrices θ in $\Theta^{+, \infty}$ whose support is included in m . Analogously, we define the parametric set $\Theta_m^{+, \infty, \text{iso}}$ corresponding to isotropic GMRFs.

We cannot directly extend the CLS empirical contrast $\gamma_{n, p_1, p_2}(\cdot)$ defined in (6) in this new setting because we have to take the edge effect into account. Indeed, if we want to compute the

conditional regression of $\mathbf{X}_{i[j_1, j_2]}$, we have to observe *all* its neighbors with respect to m , i.e. $\{\mathbf{X}_{i[j_1+l_1, j_2+l_2]}, (l_1, l_2) \in m\}$. In this regard, we define the set Λ_m for any model $m \in \mathcal{M}_1$.

$$\Lambda_m := \{(i_1, i_2) \in \Lambda, (m + (i_1, i_2)) \subset \Lambda\},$$

where $(m + (i, j))$ denotes the set m of nodes translated by (i, j) . For instance, if we consider the model m_1 with four nearest neighbors, the edge effect size is one and Λ_m contains all the nodes that do not lie on the border. The model m_3 with 12 nearest neighbors yields an edge effect of size 2 and Λ_m contains all the nodes in Λ , except those which are at a (euclidean) distance strictly smaller than 2 from the border.

For any model $m \in \mathcal{M}_1$ and any θ' belonging to $\Theta_m^{+, \infty}$, we define $\gamma_{n, p_1, p_2}^{\Lambda_m}(\cdot)$ as an analogous of $\gamma_{n, p_1, p_2}(\cdot)$ except that it only relies on the conditional regression of the nodes in Λ_m .

$$\gamma_{n, p_1, p_2}^{\Lambda_m}(\theta') := \frac{1}{n \text{Card}(\Lambda_m)} \sum_{i=1}^n \sum_{(j_1, j_2) \in \Lambda_m} \left(\mathbf{X}_{i[j_1, j_2]} - \sum_{(l_1, l_2) \in m} \theta'_{[l_1, l_2]} \mathbf{X}_{i[j_1+l_1, j_2+l_2]} \right)^2.$$

Then, the *CLS* estimators $\hat{\theta}_m^{\Lambda_m}$ and $\hat{\theta}_m^{\Lambda_m, \text{iso}}$ are defined by

$$\hat{\theta}_m^{\Lambda_m} \in \arg \min_{\theta' \in \Theta_m^{+, \infty}} \gamma_{n, p_1, p_2}^{\Lambda_m}(\theta') \quad \text{and} \quad \hat{\theta}_m^{\Lambda_m, \text{iso}} \in \arg \min_{\theta' \in \Theta_m^{+, \infty, \text{iso}}} \gamma_{n, p_1, p_2}^{\Lambda_m}(\theta').$$

Contrary to $\hat{\theta}_m$, the estimator $\hat{\theta}_m^{\Lambda_m}$ is not necessarily unique especially if the size of Λ_m is smaller than d_m . Let us mention that it is quite classical in the literature to remove nodes to take edge effects or missing data into account (see e.g. [Guy95] Sect.4.3). We cannot use anymore fast Fourier transform for computing the parametric estimator. Nevertheless, $\hat{\theta}_m^{\Lambda_m}$ is still computationally amenable, since it minimizes a quadratic function on the closed convex set $\Theta_m^{+, \infty}$.

When considering toroidal networks, we have applied a linear penalization of the order $d_m/(n \text{Card}(\Lambda))$, i.e. the dimension of the model divided by the total number of nodes observed. Since $\gamma_{n, p_1, p_2}^{\Lambda_m}(\theta')$ only evaluates the conditional regression of the nodes in the subset Λ_m , we shall penalize our model by the quantity $\frac{d_m}{n \text{Card}(\Lambda_m)}$. In the sequel, $n \text{Card}(\Lambda_m)$ is called the *effective number* of observations.

We now provide a data-driven model selection procedure for choosing the neighborhood. It is based on the slope heuristic developed in the previous section. Suppose we are given a subcollection \mathcal{M} of \mathcal{M}_1 . Then, the penalized estimator $\hat{\theta}_{\hat{m}}^{\Lambda_{\hat{m}}}$ is computed as follows.

Algorithm 5.1. (*Data-driven penalization for non-toroidal lattice*).

1. Compute the selected model $\hat{m}(N)$ as a function of $N > 0$

$$\hat{m}(N) \in \arg \min_{m \in \mathcal{M}} \left\{ \gamma_{n, p_1, p_2}^{\Lambda_m}(\hat{\theta}_m^{\Lambda_m}) + N \frac{d_m}{n \text{Card}(\Lambda_m)} \right\}.$$

2. Find $\hat{N}_{\min} > 0$ such that the jump $d_{\hat{m}}([\hat{N}_{\min}]_-) - d_{\hat{m}}([\hat{N}_{\min}]_+)$ is maximal.

3. Select the model $\hat{m} = \hat{m}(2\hat{N}_{\min})$.

This procedure straightforwardly extends to the case of isotropic GMRFs estimation by replacing $\widehat{m}(N)$ by $\widehat{m}^{\text{iso}}(N)$ and d_m by d_m^{iso} . As for Algorithm 4.1, it is advised to introduce huge models in the collection \mathcal{M} in order to better detect the dimension jump. However, when the dimension of the models increases the size of Λ_m decreases and the estimator $\widehat{\theta}_m^{\Lambda_m}$ may become unreliable. The method therefore requires a reasonable number of data. In practice, Λ should not contain less than 100 nodes.

6 Simulation study

In the first simulation experiment, we compare the efficiency of our procedure with penalized maximum likelihood methods when the field is a torus. In the second experiment, we consider the estimation of a Gaussian field observed on a rectangle. The calculations are made with `R` www.r-project.org. Throughout these simulations, we only consider isotropic estimators.

6.1 Isotropic GMRF on a torus

First, we consider X an isotropic GMRF on the torus Λ of size $p = p_1 = p_2 = 20$. There are therefore 400 points in the lattice. The number of observations n equals one and the conditional variance σ^2 is one. We introduce a radius $r := \sqrt{17}$. Then, for any number $\phi > 0$, we define the $p \times p$ matrix θ^ϕ as:

$$\begin{cases} \theta^\phi_{[0,0]} & := 0, \\ \theta^\phi_{[i,j]} & := \phi \quad \text{if } |(i,j)|_t \leq r \text{ and } (i,j) \neq (0,0), \\ \theta^\phi_{[i,j]} & := 0 \quad \text{if } |(i,j)|_t > r. \end{cases}$$

In practice, we set ϕ to 0, 0.0125, 0.015, and 0.0175. Observe that these choices constrain $\|\theta^\phi\|_1 < 1$. The matrix θ^ϕ therefore belongs to the set $\Theta_{m_{10}}^{+, \text{iso}}$ of dimension 10 introduced in Definition 2.1.

In the sequel, we shall compare the efficiency of three model selection procedures. For each of them, we use the collection $\mathcal{M} := \{m_0, m_1, \dots, m_{20}\}$ whose maximal dimension $d_{m_{20}}^{\text{iso}}$ is 21. The estimator $\widehat{\theta}^{\text{iso}}$ is built using the CLS model selection procedure introduced in Algorithm 4.2. The two other procedures are based on likelihood maximization. In this regard, we first define the parametric maximum likelihood estimator $\widehat{\theta}_m^{\text{mle}}$ for any model $m \in \mathcal{M}$.

$$\left(\widehat{\theta}_m^{\text{mle}}, \widehat{\sigma}_m^{\text{mle}} \right) := \arg \min_{\theta' \in \Theta_m^{+, \text{iso}}, \sigma'} -\mathcal{L}_p(\theta', \sigma', \mathbf{X}),$$

where $\mathcal{L}_p(\theta', \mathbf{X})$ stands for the log-likelihood at the parameter θ' . We then select a model m applying either an AIC-type criterion [Aka73] or a BIC-type criterion [Sch78]:

$$\begin{aligned} \widehat{m}^{\text{AIC}} &:= \arg \min_{m \in \mathcal{M}} \left\{ -2\mathcal{L}_p(\widehat{\theta}_m^{\text{mle}}, \widehat{\sigma}_m^{\text{mle}}, \mathbf{X}) + 2d_m^{\text{iso}} \right\}, \\ \widehat{m}^{\text{BIC}} &:= \arg \min_{m \in \mathcal{M}} \left\{ -2\mathcal{L}_p(\widehat{\theta}_m^{\text{mle}}, \widehat{\sigma}_m^{\text{mle}}, \mathbf{X}) + \log(p^2)d_m^{\text{iso}} \right\}. \end{aligned}$$

For short, we write $\widehat{\theta}^{\text{AIC}}$ and $\widehat{\theta}^{\text{BIC}}$ for the two obtained estimators $\widehat{\theta}_{\widehat{m}^{\text{AIC}}}^{\text{mle}}$ and $\widehat{\theta}_{\widehat{m}^{\text{BIC}}}^{\text{mle}}$. Although AIC and BIC procedures are not justified in this setting, we still apply them as they are widely used in many frameworks. Moreover, their computation is performed efficiently thanks to fast Fourier transform described in Section 2.3.

The experiments are repeated $N = 1000$ times. The Gaussian field is simulated by using the fast Fourier transform. The quality of the estimations is assessed thanks to the prediction loss

function $l(\cdot, \cdot)$ defined in (11). For any ϕ and any of these three estimators, we evaluate the risks $\mathbb{E}_{\theta^\phi}[l(\hat{\theta}^{\text{AIC}}, \theta^\phi)]$, $\mathbb{E}_{\theta^\phi}[l(\hat{\theta}^{\text{BIC}}, \theta^\phi)]$, and $\mathbb{E}_{\theta^\phi}[l(\tilde{\theta}^{\text{iso}}, \theta^\phi)]$ as well as the corresponding empirical 95% confidence intervals by Monte-Carlo method. Thanks to the N experiments, we also estimate the risk of $\hat{\theta}_m^{\text{iso}}$ for each model $m \in \mathcal{M}$. It then allows to evaluate the oracle risks $\mathbb{E}_{\theta^\phi}[l(\hat{\theta}_{m^*}^{\text{iso}}, \theta^\phi)]$ and the risk ratios $\mathbb{E}_{\theta^\phi}[l(\tilde{\theta}^{\text{iso}}, \theta^\phi)]/\mathbb{E}_{\theta^\phi}[l(\hat{\theta}_{m^*}^{\text{iso}}, \theta^\phi)]$. The results are given in Table 1.

$\phi \times 10^2$	0	1.25	1.5	1.75
$\mathbb{E}_{\theta^\phi}[l(\hat{\theta}^{\text{AIC}}, \theta^\phi)] \times 10^2$	1.2 ± 0.2	3.1 ± 0.2	4.3 ± 0.2	6.4 ± 0.2
$\mathbb{E}_{\theta^\phi}[l(\hat{\theta}^{\text{BIC}}, \theta^\phi)] \times 10^2$	0.01 ± 0.01	1.9 ± 0.1	3.7 ± 0.1	9.7 ± 0.3
$\mathbb{E}_{\theta^\phi}[l(\tilde{\theta}^{\text{iso}}, \theta^\phi)] \times 10^2$	1.6 ± 0.2	3.2 ± 0.2	4.2 ± 0.1	7.2 ± 0.3
$\mathbb{E}_{\theta^\phi}[l(\tilde{\theta}^{\text{iso}}, \theta^\phi)]/\mathbb{E}_{\theta^\phi}[l(\hat{\theta}_{m^*}^{\text{iso}}, \theta^\phi)]$	$+\infty$	1.9 ± 0.7	1.3 ± 0.2	1.5 ± 0.3

Table 1: First simulation study. Estimates and 95% confidence intervals of the risks $\mathbb{E}_{\theta^\phi}[l(\hat{\theta}^{\text{AIC}}, \theta^\phi)]$, $\mathbb{E}_{\theta^\phi}[l(\hat{\theta}^{\text{BIC}}, \theta^\phi)]$, and $\mathbb{E}_{\theta^\phi}[l(\tilde{\theta}^{\text{iso}}, \theta^\phi)]$ and of the ratio $\mathbb{E}_{\theta^\phi}[l(\tilde{\theta}^{\text{iso}}, \theta^\phi)]/\mathbb{E}_{\theta^\phi}[l(\hat{\theta}_{m^*}^{\text{iso}}, \theta^\phi)]$.

The BIC criterion outperforms the other procedures when $\phi = 0, 0.0125$, or 0.015 but behaves bad for a large ϕ . Indeed, the BIC criterion has a tendency to overpenalize the models. For the two first values of ϕ the oracle model in \mathcal{M} is m_0 . Hence, overpenalizing increases the performance of estimation in this case. However, when ϕ increases, the dimension of the oracle model is larger and BIC therefore selects too small models.

In contrast, AIC and the CLS estimator exhibit similar behaviors. If we forget the case $\phi = 0$ for which the oracle risk is 0, the risk of $\tilde{\theta}^{\text{iso}}$ is close to the risk of the oracle model. Hence, the neighborhood choice for $\tilde{\theta}^{\text{iso}}$ is almost optimal.

In conclusion, $\tilde{\theta}^{\text{iso}}$ or $\hat{\theta}^{\text{AIC}}$ both exhibit good performances for estimating the distribution of a regular Gaussian field on a torus. The strength of our neighborhood selection procedure lies in the fact it easily generalizes to non-toroidal lattices as illustrated in the next section.

6.2 Isotropic Gaussian fields on \mathbb{Z}^2

First simulation experiment. We now consider X an isotropic Gaussian field defined on \mathbb{Z}^2 but only observed on a square Λ of size $p = p_1 = p_2 = 20$. This corresponds to the situation described in Section 5. The variance of $X_{[0,0]}$ is set to one and the distribution of the field is therefore uniquely defined by its correlation function $\rho(k, l) := \text{corr}(X_{[k,l]}, X_{[0,0]})$. Again, the number of replications n is chosen to be one. In the first experiment, we use four classical correlation functions: exponential, spherical, circular, and Matérn (e.g. [Cre93] Sect.2.3.1 and

[Mat86]).

$$\begin{aligned}
 \text{Exponential: } \rho(k, l) &= \exp\left(-\frac{d(k, l)}{r}\right) \\
 \text{Circular: } \rho(k, l) &= \begin{cases} 1 - \frac{2}{\pi} \left[\frac{d(k, l)}{r} \sqrt{1 - \left(\frac{d(k, l)}{r}\right)^2} + \sin^{-1}\left(\sqrt{\frac{d(k, l)}{r}}\right) \right] & \text{if } d(k, l) \leq r \\ 0 & \text{else} \end{cases} \\
 \text{Spherical: } \rho(k, l) &= \begin{cases} 1 - 1.5 \frac{d(k, l)}{r} + 0.5 \left(\frac{d(k, l)}{r}\right)^3 & \text{if } d(k, l) \leq r \\ 0 & \text{else} \end{cases} \\
 \text{Matérn: } \rho(k, l) &= \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{d(k, l)}{r}\right)^{\kappa} \mathcal{K}_{\kappa}\left(\frac{d(k, l)}{r}\right),
 \end{aligned}$$

where $d(k, l)$ denotes the euclidean distance from (k, l) to $(0, 0)$ and $\mathcal{K}_{\kappa}(\cdot)$ is the modified Bessel function of order κ . In a nutshell, the parameter r represents the range of correlation, whereas κ may be regarded as a smoothness parameter for the Matérn function. In this simulation experiment, we set r to 3. When considering the Matérn model, we take κ equal to 0.05, 0.25, 0.5, and 1.

The Gaussian fields are simulated using the function *GaussRF* in the library *RandomFields*. For each of experiments, we compute the estimator $\tilde{\theta}^{\text{iso}}$ based on Algorithm 5.1 with the collection $\mathcal{M} := \{m \in \mathcal{M}_1, d_m^{\text{iso}} \leq 18\}$. Since the lattice Λ is not a torus, methods based on likelihood maximization exhibit a prohibitive computational burden. Consequently, we do not use MLE in this experiment. We shall compare the efficiency of $\tilde{\theta}^{\text{iso}}$ with a variogram-based estimation method.

Let us recall that the matrix θ corresponds to the coefficients of the conditional regression of $X_{[0,0]}$ given $\{X_{[i,j]}, (i, j) \in \Lambda \setminus \{(0, 0)\}\}$ (Equation (1)). Hence, the linear combination $\sum_{(i,j) \in \Lambda \setminus \{(0,0)\}} \theta_{[i,j]} X_{[i,j]}$ is the kriging predictor of $X_{[0,0]}$ given the remaining variables. A natural method to estimate θ in this spatial setting amounts to estimating the variogram of the observed Gaussian field and then performing ordinary kriging at the node $(0, 0)$. More precisely, we first estimate the empirical variogram by applying the modulus estimator of Hawkes and Cressie (e.g. [Cre93] Eq.(2.2.8)) to the observed field of 400 points. Afterwards, we fit this empirical variogram to a variogram model using the reweighted least-squares suggested by Cressie [Cre85]. This procedure therefore requires the choice of a particular variogram model. In the first simulation study, we choose *the model* that has generated the data. Observe that this method is *not* adaptive since it requires the knowledge of the variogram model. In practice, we use Library *geoR* [JD01] implemented in *R* to estimate the parameters r , $\text{var}(X_{[0,0]})$ and eventually κ of the variogram model. Then, we compute the estimator $\hat{\theta}^K$ by performing ordinary kriging at the center node of Λ . For each of these estimations, we assume that the variogram model is known. For computational reasons, we use a kriging neighborhood of size 11×11 that contains 120 points. Previous simulations have indicated that this neighborhood choice does not decrease the precision of the estimation.

We again assess the performances of the procedures using the loss $l(\cdot, \cdot)$. Even, if this loss is defined in (11) for a torus, the alternative definition (12) clearly extends to this non-toroidal setting. Consequently, the loss $l(\hat{\theta}, \theta)$ measures the difference between the prediction error of $X_{[0,0]}$ when using $\sum_{(i,j) \in \Lambda \setminus \{(0,0)\}} \hat{\theta}_{[i,j]} X_{[i,j]}$ and the prediction error of $X_{[0,0]}$ when using the best predictor $\mathbb{E}[X_{[0,0]} | (X_{[i,j]})_{(i,j) \in \Lambda \setminus \{(0,0)\}}]$. In other words, $l(\hat{\theta}, \theta)$ is the difference of the kriging error

made with the estimated parameters $\widehat{\theta}$ and the kriging error made with the true parameter θ .

The experiments are repeated $N = 1000$ times. For any of the four correlation models previously mentioned, we evaluate the risks $\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)]$ and $\mathbb{E}_\theta[l(\widehat{\theta}^K, \theta)]$ by Monte-Carlo and also compare the risk of $\widehat{\theta}^{\text{iso}}$ with the oracle $\widehat{\theta}_{m^*}^{\text{iso}}$. As in Section 6.1, we evaluate the oracle risk $\mathbb{E}[l(\widehat{\theta}_{m^*}^{\text{iso}}, \theta)]$ by taking the minimum of the evaluations of the risks $\mathbb{E}[l(\widehat{\theta}_m^{\text{iso}}, \theta)]$ over all models $m \in \mathcal{M}$. Results of the simulation experiment are given in Table 2 and 3.

Observe that none of the fields considered in this study are GMRFs. Here, the GMRF models should only be viewed as a collection of approximation sets of the true distribution. This simulation experiment is in the spirit of Rue and Tjelmeland's study [RT02]. However, there are some major differences. Contrary to them, we perform estimation and not only approximation. Moreover, our lattice is not a torus. Finally, we use our prediction loss $l(.,.)$ to assess the performance, whereas they compare the correlation functions.

Model	Exponential	Circular	Spherical
$\mathbb{E}_\theta[l(\widehat{\theta}^K, \theta)] \times 10^2$	0.08 ± 0.03	9.1 ± 0.5	2.9 ± 0.1
$\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)] \times 10^2$	0.77 ± 0.04	6.2 ± 0.1	2.7 ± 0.1
$\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)]/\mathbb{E}_\theta[l(\widehat{\theta}_{m^*}^{\text{iso}}, \theta)]$	1.5 ± 0.1	1.1 ± 0.1	1.5 ± 0.1

Table 2: Estimates and 95% confidence intervals of the risks $\mathbb{E}_\theta[l(\widehat{\theta}^K, \theta)]$ and $\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)]$ and of the ratio $\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)]/\mathbb{E}_\theta[l(\widehat{\theta}_{m^*}^{\text{iso}}, \theta)]$ for the exponential, circular and spherical models.

κ	0.05	0.25	0.5	1
$\mathbb{E}_\theta[l(\widehat{\theta}^K, \theta)] \times 10^2$	13.0 ± 1.0	9.9 ± 1.0	9.0 ± 1.7	2.7 ± 0.4
$\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)] \times 10^2$	4.6 ± 0.2	1.7 ± 0.1	0.77 ± 0.04	0.36 ± 0.02
$\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)]/\mathbb{E}_\theta[l(\widehat{\theta}_{m^*}^{\text{iso}}, \theta)]$	2.3 ± 0.2	3.4 ± 0.5	1.5 ± 0.1	2.2 ± 0.2

Table 3: Estimates and 95% confidence intervals of the risks $\mathbb{E}_\theta[l(\widehat{\theta}^K, \theta)]$ and $\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)]$ and of the ratio $\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)]/\mathbb{E}_\theta[l(\widehat{\theta}_{m^*}^{\text{iso}}, \theta)]$ for Matérn model.

Comments on Tables 2 and 3: In both tables, the ratio $\mathbb{E}_\theta[l(\widehat{\theta}^{\text{iso}}, \theta)]/\mathbb{E}_\theta[l(\widehat{\theta}_{m^*}^{\text{iso}}, \theta)]$ stays close to one. Hence, the model selection is almost optimal from an efficiency point of view. Apart from the exponential model, the estimator $\widehat{\theta}^{\text{iso}}$ outperforms the estimator $\widehat{\theta}^K$ based on geostatistical methods. This is particularly striking for the Matérn correlation model because in that case the computation of $\widehat{\theta}^K$ requires the estimation of the additional parameter κ . Indeed, let us recall that the exponential model and the Matérn model with $\kappa = 0.5$ are equivalent. For $\kappa = 0.5$, the risk of $\widehat{\theta}^K$ is 100 times higher when κ has to be estimated than when κ is known.

Second simulation experiment. The kriging estimator $\widehat{\theta}^K$ requires the knowledge or the choice of a correlation model. In the second simulation experiment, the correlation of X is the Matérn function with range $r = 3$ and $\kappa = 0.05$. We now estimate θ using *different* variogram models, namely the exponential, the circular, the spherical and the Matérn model. The estimator $\widehat{\theta}^{\text{iso}}$ for such a field was already considered in Table 3. The experiment is repeated $N = 1000$ times.

Comments on Table 4: One observes, that circular and spherical models yield worse performances than Matérn model. In contrast, the exponential model behaves better. The choice

Model	Exponential	Circular	Spherical	Matèrn
$\mathbb{E}_\theta[l(\widehat{\theta}^K, \theta)] \times 10^2$	4.8 ± 0.2	44.8 ± 0.4	27.2 ± 0.3	13.0 ± 1.0

Table 4: Estimates and 95% confidence intervals of the risks $\mathbb{E}_\theta[l(\widehat{\theta}^K, \theta)]$ for Matèrn model with $\kappa = 0.05$ when using the exponential, circular, spherical, and Matèrn models.

of the variogram model therefore seems critical to get good performances. The model selection estimator $\widehat{\theta}^{\text{iso}}$ (Table 3) exhibits a smaller risk than the exponential model but the difference is not really significative.

Third simulation experiment. We consider a different correlation model which exhibits negative correlations. This *wave* model is used for instance when modelling ecological data. In the sequel, r still refers to a positive range parameter.

Wave:
$$\rho(k, l) = \frac{r}{d(k, l)} \sin\left(\frac{d(k, l)}{r}\right),$$

where we recall that $d(k, l)$ is the euclidean distance from (k, l) to $(0, 0)$. We consider the two values $r = 0.25$ and $r = 0.5$. Using $N = 1000$ observations of the field, we compare our neighborhood selection procedure with variogram-based estimators.

Range	Model selection	Exponential	Circular	Spherical	Matèrn	Wave
0.25	26.0 ± 0.2	34.2 ± 0.1	100.3 ± 0.1	67.9 ± 0.1	34.2 ± 0.1	81.8 ± 0.4
0.5	5.3 ± 0.1	46.4 ± 0.3	5.8 ± 0.2	9.8 ± 0.3	3.2 ± 6.9	80.0 ± 6.8

Table 5: Estimates and 95% confidence intervals of the risks $\mathbb{E}_\theta[l(\widehat{\theta}, \theta)] \times 100$ for the wave model with range 0.25 and 0.5.

Comments on Table 5: When $r = 0.25$, our procedure performs better, while exponential and Matèrn models provide rather good models. Observe that the wave variogram model provides terrible results. The conditional variance σ^2 for this particular field equals 0.658. We have mentioned earlier that the mean-squared prediction error is $\sigma^2 + \mathbb{E}_\theta[l(\widehat{\theta}, \theta)]$. Since the variance of $X_{[0,0]}$ is one, the mean-squared prediction error of the constant estimator $\widehat{\theta} = 0$ is one. The constant estimator therefore performs better than the circular, spherical, and the wave model-based estimators. It also performs as well as the exponential and Matèrn models. Only the neighborhood selection procedure exhibits a smaller error. When $r = 0.5$, the exponential and wave models are terrible. The best procedure seems to be the Matèrn-based estimator. Nevertheless, for the Matèrn model we only have considered 937 samples out of the 1000 because the procedure did not converge for 73 simulations. The neighborhood selection procedure performs quite well and the oracle ratio is smaller than two.

7 Conclusions

In this paper, we have extended a neighborhood selection procedure introduced in [Ver09]. On the one hand, an algorithm is provided for tuning the penalty in practice. On the other hand, the new method also handles non-toroidal lattices. The computational complexity remains reasonable even when the size of the lattice is large.

When one considers a stationary field on a torus, our neighborhood selection procedure exhibits a computational burden and statistical performances analogous to the AIC procedure. Even if this last method is not analysed yet from an efficiency point of view, this suggests that AIC may achieve an oracle inequality in this setting. Moreover, we have empirically checked that $\hat{\theta}$ performs almost as well as the oracle model since the oracle ratio $\mathbb{E}[l(\hat{\theta}, \theta)]/\mathbb{E}[l(\hat{\theta}_{m^*}, \theta)]$ remains close to one.

The strength of this neighborhood selection procedure lies in the fact it easily extends to non-toroidal lattices. We have illustrated that our method often outperforms variogram-based estimation methods in terms of the mean-squared prediction error. Moreover, the procedure behaves almost as well as the oracle. In contrast, variogram-based procedures may perform well for some covariances structure but also yield terrible results for other covariance structures. These results illustrate the *adaptivity* of the neighborhood selection procedure.

8 Proofs

Let us introduce some notations that shall be used throughout the proofs. For any $1 \leq k \leq n$, the vector \mathbf{X}_k^v denotes the vectorialized version of the k -th sample of X . Moreover, \mathbf{X}^v is the matrix of size $p_1 p_2 \times n$ of the n realisations of the vector \mathbf{X}_k^v . Throughout these proofs, L, L_1, L_2 denote constants that may vary from line to line. The notation $L(\cdot)$ specifies the dependency on some quantities. Finally, the $\gamma(\cdot)$ function stands for an infinite sampled version of the CLS criterion $\gamma_{n,p_1,p_2}(\cdot)$: $\gamma(\cdot) := \mathbb{E}[\gamma_{n,p_1,p_2}(\cdot)]$.

8.1 Proof of Lemma 2.1

Let us provide an alternative expression of $\gamma_{n,p_1,p_2}(\theta')$ in term of the factor $C(\theta')$ and the empirical covariance matrix $\overline{\mathbf{X}^v \mathbf{X}^{v*}}$.

$$\gamma_{n,p_1,p_2}(\theta') = \frac{1}{np_1 p_2} \text{tr} [(I_{p_1 p_2} - C(\theta')) \overline{\mathbf{X}^v \mathbf{X}^{v*}} (I_{p_1 p_2} - C(\theta'))] . \quad (17)$$

This is justified in [Ver09] Sect.2.2.

Lemma 8.1. *There exists an orthogonal matrix P which simultaneously diagonalizes every $p_1 p_2 \times p_1 p_2$ symmetric block circulant matrices with $p_2 \times p_2$ blocks. Let θ be a matrix of size $p_1 \times p_2$ such that $C(\theta)$ is symmetric. The matrix $D(\theta) = P^* C(\theta) P$ is diagonal and satisfies*

$$D(\theta)_{[(i-1)p_2+j, (i-1)p_2+j]} = \sum_{k=1}^{p_1} \sum_{l=1}^{p_2} \theta_{[k,l]} \cos[2\pi(ki/p_1 + lj/p_2)] , \quad (18)$$

for any $1 \leq i \leq p_1$ and $1 \leq j \leq p_2$.

This lemma is proved as in [RH05] Sect.2.6.2 to the price of a slight modification that takes into account the fact that P is orthogonal and not unitary. The difference comes from the fact that contrary to Rue and Held we also assume that $C(\theta)$ is symmetric. Lemma 8.1 states that all symmetric block circulant matrices are simultaneously diagonalizable. Observe that for any $1 \leq i \leq p_1$ and $1 \leq j \leq p_2$, it holds that $D(\theta)_{[(i-1)p_2+j, (i-1)p_2+j]} = \lambda_{[i,j]}(\theta)$ since $\theta_{[k,l]} = \theta_{[p_1-k, p_2-l]}$. Hence, Expression (17) becomes

$$\gamma_{n,p_1,p_2}(\theta') = \frac{1}{np_1 p_2} \left\{ \sum_{i=1}^{p_1} \sum_{j=1}^{p_2} [1 - \lambda_{[i,j]}(\theta)]^2 \left[\sum_{k=1}^n [P^* \mathbf{X}_k^v (\mathbf{X}_k^v)^* P]_{[(i-1)p_2+j, (i-1)p_2+j]} \right] \right\} ,$$

where \mathbf{X}_k^v is the vectorialized version of the k -th observation of the field X . Straightforward computations allow us to prove that the quantities

$$(P^* \mathbf{X}_k^v (\mathbf{X}_k^v)^* P)_{[(i-1)p_2+j, (i-1)p_2+j]} + (P^* \mathbf{X}_k^v (\mathbf{X}_k^v)^* P)_{[(p_1-i-1)p_2+p_2-j, (p_1-i-1)p_2+p_2-j]}$$

and

$$\frac{1}{\sqrt{p_1 p_2}} \lambda_{[i,j]}(\mathbf{X}_k^v) \overline{\lambda_{[i,j]}(\mathbf{X}_k^v)} + \frac{1}{\sqrt{p_1 p_2}} \lambda_{[p_1-i, p_2-j]}(\mathbf{X}_k^v) \overline{\lambda_{[p_1-i, p_2-j]}(\mathbf{X}_k^v)}$$

are equal for any $1 \leq i \leq p_1$ and $1 \leq j \leq p_2$. Here, the entries of the matrix $\lambda(\cdot)$ are taken modulo p_1 and p_2 and the entries of $[P^* \mathbf{X}_k^v (\mathbf{X}_k^v)^* P]$ are taken modulo $p_1 p_2$. The result of Lemma 2.1 follows.

8.2 Proof of Proposition 4.1

Proof of Proposition 4.1. We only consider the anisotropic case, since the proof for isotropic estimation is analogous. For any model $m \in \mathcal{M}_1$, we define

$$\Delta(m, m') := \gamma_{n,p,p}(\widehat{\theta}_{m,\rho}) + \text{pen}(m) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) - \text{pen}(m').$$

We aim at showing that with large probability, the quantity $\Delta(m, m')$ is positive for all small dimensional models m . Hence, we would conclude that the dimension of \widehat{m} is large. In this regard, we bound the deviations of the differences

$$\begin{aligned} \gamma_{n,p,p}(\widehat{\theta}_{m,\rho}) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) &= \left[\gamma_{n,p,p}(\widehat{\theta}_{m,\rho}) - \gamma_{n,p,p}(\theta_{m,\rho}) \right] + \left[\gamma_{n,p,p}(\theta_{m,\rho}) - \gamma_{n,p,p}(\theta) \right] \\ &\quad + \left[\gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) \right]. \end{aligned}$$

Lemma 8.2. *Let K_2 be some universal constant that we shall define in the proof. With probability larger than $3/4$,*

$$\gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\theta_{m,\rho}) \leq \frac{K_2}{2} \rho^2 \varphi_{\max}(\Sigma) \frac{d_m \vee 1}{np^2} \quad \text{and} \quad \gamma_{n,p,p}(\theta_{m,\rho}) - \gamma_{n,p,p}(\widehat{\theta}_{m,\rho}) \leq \frac{K_2}{2} \rho^2 \varphi_{\max}(\Sigma) \frac{d_m}{np^2}$$

for all models $m \in \mathcal{M}_1$.

Lemma 8.3. *Assume that p is larger than some numerical constant p_0 . With probability larger than $3/4$, it holds that*

$$\gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) \geq K_3 \sigma^2 \left[\varphi_{\min} [I_{p^2} - C(\theta)] \wedge [\rho - \varphi_{\max} (I_{p^2} - C(\theta))] \right] \frac{d_{m'}}{np^2},$$

where K_3 is a universal constant defined in the proof.

Let us take K_1 to be exactly K_3 . Gathering the two last lemma with Assumption (15), there exists an event Ω of probability larger than $1/2$ such that

$$\Delta(m, m') \geq \frac{\sigma^2}{np^2} \left\{ K_1 \eta d_{m'} \left[\varphi_{\min} (I_{p^2} - C(\theta)) \wedge [\rho - \varphi_{\max} (I_{p^2} - C(\theta))] \right] - K_2 (d_m \vee 1) \frac{\rho^2}{\varphi_{\min} [I_{p^2} - C(\theta)]} \right\},$$

for all models $m \in \mathcal{M}_1$. Thus, conditionally to Ω , $\Delta(m, m')$ is positive for all models $m \in \mathcal{M}_1$ that satisfy

$$\frac{d_m \vee 1}{d_{m'}} \leq \frac{K_3 \eta}{K_2 \rho^2} \varphi_{\min}(I_{p^2} - C(\theta)) \{ \varphi_{\min}(I_{p^2} - C(\theta)) \wedge [\rho - \varphi_{\max}(I_{p^2} - C(\theta))] \} .$$

By Lemma 7.13 of [Ver09], the dimension $d_{m'}$ is larger than $\frac{1}{2} \left[\sqrt{np^2} \wedge (p^2 - 1) \right]$. We conclude that

$$d_{\widehat{m}_\rho} \vee 1 \geq \left[\sqrt{np^2} \wedge p^2 - 1 \right] \frac{K_3 \eta}{K_2 \rho^2} \varphi_{\min}(I_{p^2} - C(\theta)) \{ \varphi_{\min}(I_{p^2} - C(\theta)) \wedge [\rho - \varphi_{\max}(I_{p^2} - C(\theta))] \} ,$$

with probability larger than $1/2$. \square

Proof of Lemma 8.2. In the sequel, $\bar{\gamma}_{n,p,p}(\cdot)$ denotes the difference $\gamma_{n,p,p}(\cdot) - \gamma(\cdot)$. Given a model m , we consider the difference

$$\gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\theta_{m,\rho}) = \bar{\gamma}_{n,p,p}(\theta) - \bar{\gamma}_{n,p,p}(\theta_{m,\rho}) - l(\theta_{m,\rho}, \theta) .$$

Upper bounding the difference of $\gamma_{n,p,p}$ therefore amounts to bounding the difference of $\bar{\gamma}_{n,p,p}$. By definition of $\gamma_{n,p,p}$ and γ , it expresses as

$$\bar{\gamma}_{n,p,p}(\theta) - \bar{\gamma}_{n,p,p}(\theta_{m,\rho}) = \frac{1}{p^2} \text{tr} \{ [(I_{p^2} - C(\theta))^2 - (I_{p^2} - C(\theta_{m,\rho}))^2] (\overline{\mathbf{X}^v \mathbf{X}^{v*}} - \Sigma) \} .$$

The matrices Σ , $(I_{p^2} - C(\theta))$, and $(I_{p^2} - C(\theta_{m,\rho}))$ are symmetric block circulant. By Lemma 8.1, they are jointly diagonalizable in the same orthogonal basis. If we note P an orthogonal matrix associated to this basis, then $C(\theta_{m,\rho})$, $C(\theta)$, and Σ respectively decompose in

$$C(\theta_{m,\rho}) = P^* D(\theta_{m,\rho}) P , C(\theta) = P^* D(\theta) P \text{ and } \Sigma = P^* D(\Sigma) P ,$$

where the matrices $D(\theta_{m,\rho})$, $D(\theta)$, and $D(\Sigma)$ are diagonal.

$$\begin{aligned} \bar{\gamma}_{n,p,p}(\theta) - \bar{\gamma}_{n,p,p}(\theta_{m,\rho}) = \\ \frac{1}{p^2} \text{tr} \{ (D(\theta_{m,\rho}) - D(\theta)) [2I_{p^2} - D(\theta) - D(\theta_{m,\rho})] D_\Sigma (\overline{\mathbf{Y} \mathbf{Y}^*} - I_{p^2}) \} , \end{aligned} \quad (19)$$

where the matrix \mathbf{Y} is defined as $P \sqrt{\Sigma^{-1}} \mathbf{X}^v P^*$. Its components follow independent standard Gaussian distributions. Since the matrices involved in (19) are diagonal, Expression (19) is a linear combination of centered χ^2 random variables. We apply the following lemma to bound its deviations.

Lemma 8.4. *Let (Y_1, \dots, Y_D) be i.i.d. standard Gaussian variables. Let a_1, \dots, a_D be fixed numbers. We set*

$$\|a\|_\infty := \sup_{i=1, \dots, D} |a_i|, \quad \|a\|_2^2 := \sum_{i=1}^D a_i^2$$

Let T be the random variable defined by

$$T := \sum_{i=1}^D a_i (Y_i^2 - 1) .$$

Then, the following deviation inequality holds for any positive x

$$\mathbb{P} [T \geq 2\|a\|_2 \sqrt{x} + 2\|a\|_\infty x] \leq e^{-x} .$$

This result is very close to Lemma 1 of Laurent and Massart in [LM00]. The only difference lies in the fact that they constrain the coefficients a_i to be non-negative. Nevertheless, their proof easily extends to our situation. Let us define the matrix a of size $n \times p^2$ as

$$a^i[j] := \frac{D_{\Sigma[i,i]}(D(\theta_{m,\rho})[i,i] - D(\theta)[i,i])(2 - D(\theta)[i,i] - D(\theta_{m,\rho})[i,i])}{np^2},$$

for any $1 \leq i \leq n$ and any $1 \leq j \leq p^2$. Since the matrices $I - C(\theta)$ and $I - C(\theta_{m,\rho})$ belong to the set Θ_{ρ}^+ , their largest eigenvalue is smaller than ρ . By Definition (11) of the loss function $l(\cdot, \cdot)$, $\|a\|_2 \leq 2\rho\sqrt{\frac{\varphi_{\max}(\Sigma)l(\theta_{m,\rho},\theta)}{np^2}}$ and $\|a\|_{\infty} \leq 4\rho^2\frac{\varphi_{\max}(\Sigma)}{np^2}$. By Applying Lemma 8.4 to Expression (19), we conclude that

$$\mathbb{P}\left[\bar{\gamma}_{n,p,p}(\theta) - \bar{\gamma}_{n,p,p}(\theta_{m,\rho}) \geq l(\theta_{m,\rho}, \theta) + 12\rho^2\frac{\varphi_{\max}(\Sigma)}{np^2}x\right] \leq e^{-x},$$

for any $x > 0$. Consequently, for any $K > 0$, the difference of $\gamma_{n,p,p}(\cdot)$ satisfies

$$\gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\theta_{m,\rho}) \leq \frac{K}{2}\rho^2\varphi_{\max}(\Sigma)\frac{d_m \vee 1}{np^2},$$

simultaneously for all models $m \in \mathcal{M}_1$ with probability larger than $1 - \sum_{m \in \mathcal{M}_1 \setminus \emptyset} e^{-K(d_m \vee 1)/24}$. If K is chosen large enough, the previous upper bound holds on an event of probability larger than $7/8$. Let us call K'_2 such a value.

Let us now turn to the second part of the result. As previously, we decompose the difference of empirical contrasts

$$\gamma_{n,p,p}(\theta_{m,\rho}) - \gamma_{n,p,p}(\hat{\theta}_{m,\rho}) = \bar{\gamma}_{n,p,p}(\theta_{m,\rho}) - \bar{\gamma}_{n,p,p}(\hat{\theta}_{m,\rho}) - l(\hat{\theta}_{m,\rho}, \theta_{m,\rho})$$

Arguing as in the proof of Theorem 3.1 in [Ver09], we obtain an upper bound analogous to Eq.(59) in [Ver09]

$$\bar{\gamma}_{n,p,p}(\theta_{m,\rho}) - \bar{\gamma}_{n,p,p}(\hat{\theta}_{m,\rho}) \leq l(\hat{\theta}_{m,\rho}, \theta_{m,\rho}) + \rho^2 \left\{ \sup_{R \in \mathcal{B}_{m^2, m^2}^{\mathcal{H}'_2}} \frac{1}{p^2} \text{tr} [RD_{\Sigma}(\overline{\mathbf{Y}\mathbf{Y}^*} - I_{p^2})] \right\}^2.$$

The set $\mathcal{B}_{m^2, m^2}^{\mathcal{H}'_2}$ is defined in the proof of Lemma 8.5 in [Ver09]. Its precise definition is not really of interest in this proof. Coming back to the difference of $\gamma_{n,p,p}(\cdot)$, we get

$$\gamma_{n,p,p}(\theta_{m,\rho}) - \gamma_{n,p,p}(\hat{\theta}_{m,\rho}) \leq \rho^2 \left\{ \sup_{R \in \mathcal{B}_{m^2, m^2}^{\mathcal{H}'_2}} \frac{1}{p^2} \text{tr} [RD_{\Sigma}(\overline{\mathbf{Y}\mathbf{Y}^*} - I_{p^2})] \right\}^2.$$

We consecutively apply Lemma 8.6 and 8.7 in [Ver09] to bound the deviation of this supremum. Hence, for any positive number α ,

$$\gamma_{n,p,p}(\theta_{m,\rho}) - \gamma_{n,p,p}(\hat{\theta}_{m,\rho}) \leq L_1(1 + \alpha/2)\rho^2\varphi_{\max}(\Sigma)\frac{d_m}{np^2}. \quad (20)$$

with probability larger than $1 - \exp\left[-L_2\sqrt{d_m}\left(\frac{\alpha}{\sqrt{1+\alpha/2}} \wedge \frac{\alpha^2}{1+\alpha/2}\right)\right]$. Thus, there exists some numerical constant α_0 such that the upper bound (20) with $\alpha = \alpha_0$ holds simultaneously for all

models $m \in \mathcal{M}_1 \setminus \emptyset$ with probability larger than $7/8$. Choosing K_2 to be the supremum of K'_2 and $2L_1(1 + \alpha_0/2)$ allows to conclude. \square

Proof of Lemma 8.3. Thanks to the definition (17) of $\gamma_{n,p,p}(\cdot)$ we obtain

$$\gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) = \frac{1}{p^2} \sup_{\theta' \in \Theta_{m',\rho}^+} \text{tr} [(C(\theta') - C(\theta)) (2I_{p^2} - C(\theta) - C(\theta')) \Sigma \overline{\mathbf{Z}\mathbf{Z}^*}] ,$$

where the $p^2 \times n$ matrix \mathbf{Z} is defined by $\mathbf{Z} := \sqrt{\Sigma}^{-1} \mathbf{X}^v$. We recall that the matrices Σ , $C(\theta)$ and $C(\theta')$ commute since they are jointly diagonalizable by Lemma 8.1. Let $(\Theta_{m',\rho}^+ - \theta)$ be the set $\Theta_{m',\rho}^+$ translated by θ . Since $C(\theta) + C(\theta') = C(\theta + \theta')$, we lower bound the difference of $\gamma_{n,p,p}(\cdot)$ as follows

$$\begin{aligned} \gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) &= \frac{1}{p^2} \sup_{\theta' \in (\Theta_{m',\rho}^+ - \theta)} 2\sigma^2 \text{tr} [C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*}] - \text{tr} [C(\theta')^2 \Sigma \overline{\mathbf{Z}\mathbf{Z}^*}] \\ &\geq \frac{\sigma^2}{p^2} \sup_{\theta' \in (\Theta_{m',\rho}^+ - \theta)} \{2\text{tr} [C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*}] - \varphi_{\min}^{-1} [I_{p^2} - C(\theta)] \text{tr} [C(\theta')^2 \overline{\mathbf{Z}\mathbf{Z}^*}]\} . \end{aligned}$$

Let us consider $\Psi_{i_1, j_1}, \dots, \Psi_{i_{d_{m'}}, j_{d_{m'}}}$ a basis of the space $\Theta_{m'}$ defined in Eq.(14) of [Ver09]. Let α be a positive number that we shall define later. We then introduce θ' as

$$\theta' := \varphi_{\min} [I_{p^2} - C(\theta)] \frac{\alpha}{p^2} \sum_{k=1}^{d_{m'}} \text{tr} [C(\Psi_{i_k, j_k}) \overline{\mathbf{Z}\mathbf{Z}^*}] \Psi_{i_k, j_k} .$$

Since θ is assumed to belong to $\Theta_{m',\rho}^+$, the parameter θ' belongs to $(\Theta_{m',\rho}^+ - \theta)$ if $\varphi_{\max}[C(\theta')] \leq \varphi_{\min}(I_{p^2} - C(\theta))$ and $\varphi_{\min}[C(\theta')] \geq -\rho + \varphi_{\max}(I_{p^2} - C(\theta))$. The largest eigenvalue of $C(\theta')$ is smaller than $\|\theta'\|_1$ whereas its smallest eigenvalue is larger than $-\|\theta'\|_1$. Let us upper bound the l_1 norm of θ' :

$$\begin{aligned} \|\theta'\|_1 &= 2\varphi_{\min} [I_{p^2} - C(\theta)] \frac{\alpha}{p^2} \sum_{k=1}^{d_{m'}} |\text{tr} [C(\Psi_{i_k, j_k}) \overline{\mathbf{Z}\mathbf{Z}^*}]| \\ &\leq 2\sqrt{\frac{\alpha}{p^2} \varphi_{\min} [I_{p^2} - C(\theta)] d_{m'} \text{tr} [C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*}]} . \end{aligned} \quad (21)$$

Hence, θ' belongs to $(\Theta_{m',\rho}^+ - \theta)$ if

$$\|\theta'\|_1 \leq \varphi_{\min} (I_{p^2} - C(\theta)) \wedge [\rho - \varphi_{\max} (I_{p^2} - C(\theta))] . \quad (22)$$

Thus, we get the lower bound

$$\gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) \geq \frac{\sigma^2}{p^2} \{2\text{tr} [C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*}] - \varphi_{\min}^{-1} [I_{p^2} - C(\theta)] \text{tr} [C(\theta')^2 \overline{\mathbf{Z}\mathbf{Z}^*}]\} , \quad (23)$$

as soon as Condition (22) is satisfied.

Let us now bound the deviations of the two random variables involved in (21) and (23) by applying Markov's and Tchebychev's inequality. For the sake of simplicity, we assume that $d_{m'}$ is smaller than $(p^2 - 2p)/2$. In such a case, all the nodes in m' are different from their symmetric in Λ . We omit the proof for $d_{m'}$ larger than $(p^2 - 2p)/2$ because the approach is analogous but the computations are slightly more involved. Straightforwardly, we get

$$\mathbb{E} [tr (C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*})] = 4\alpha\varphi_{\min} [I_{p^2} - C(\theta)] \frac{d_{m'}}{n} ,$$

since the neighborhood m' only contains points (i, j) whose symmetric $(-i, -j)$ is different. A cumbersome but pedestrian computation leads to the upper bound

$$\text{var} [tr (C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*})] \leq L_1\alpha^2\varphi_{\min}^2 [I_{p^2} - C(\theta)] \frac{d_{m'}}{n^2} ,$$

where L_1 is a numerical constant. Similarly, we upper bound the expectation of $tr [C(\theta')^2 \overline{\mathbf{Z}\mathbf{Z}^*}]$

$$\mathbb{E} [tr (C(\theta')^2 \overline{\mathbf{Z}\mathbf{Z}^*})] \leq L_2\alpha^2\varphi_{\min}^2 [I_{p^2} - C(\theta)] \frac{d_{m'}}{n} .$$

Let us respectively apply Tchebychev's inequality and Markov's inequality to the variables $tr [C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*}]$ and $tr [C(\theta')^2 \overline{\mathbf{Z}\mathbf{Z}^*}]$. Hence, there exists an event Ω of probability larger than $3/4$ such that

$$\begin{aligned} 2tr [C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*}] - \varphi_{\min}^{-1} [I_{p^2} - C(\theta)] tr [C(\theta')^2 \overline{\mathbf{Z}\mathbf{Z}^*}] \geq \\ \varphi_{\min} [I_{p^2} - C(\theta)] \frac{d_{m'}}{n} \left\{ 8\alpha \left(1 - \sqrt{\frac{L_1'}{d_{m'}}} \right) - \alpha^2 L_2' \right\} \end{aligned}$$

and

$$tr [C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*}] \leq 4\alpha\varphi_{\min} [I_{p^2} - C(\theta)] \frac{d_{m'}}{n} \left(1 + \sqrt{\frac{L_1'}{d_{m'}}} \right) .$$

In the sequel, we assume that p is larger than some universal constant p_0 , which ensures the dimension $d_{m'}$ to be larger than $4L_1'$. Gathering (21) with the upper bound on $tr [C(\theta') \overline{\mathbf{Z}\mathbf{Z}^*}]$ yields

$$\|\theta'\|_1 \leq 2\sqrt{2}\alpha\varphi_{\min} [I_{p^2} - C(\theta)] \frac{d_{m'}}{\sqrt{np^2}} \leq 2\sqrt{2}\alpha\varphi_{\min} [I_{p^2} - C(\theta)] ,$$

since $d_{m'} \leq p\sqrt{n}$. If $2\sqrt{2}\alpha$ is smaller than $1 \wedge \{[\rho - \varphi_{\max} (I_{p^2} - C(\theta))] \varphi_{\min}^{-1} [I_{p^2} - C(\theta)]\}$, then Condition (22) is fulfilled on the event Ω and it follows from (23) that

$$\mathbb{P} \left\{ \gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) \geq 4\sigma^2\varphi_{\min} [I_{p^2} - C(\theta)] \frac{d_{m'}}{np^2} [\alpha - \alpha^2 L_2'/4] \right\} \geq \frac{3}{4} .$$

Choosing $\alpha = \frac{2}{L_2'} \wedge \frac{\sqrt{2}}{4} \wedge \sqrt{2} \frac{\rho - \varphi_{\max}(I_{p^2} - C(\theta))}{4\varphi_{\min}(I_{p^2} - C(\theta))}$, we get

$$\mathbb{P} \left\{ \gamma_{n,p,p}(\theta) - \gamma_{n,p,p}(\widehat{\theta}_{m',\rho}) \geq K_3\sigma^2 \left\{ \varphi_{\min} [I_{p^2} - C(\theta)] \wedge [\rho - \varphi_{\max} (I_{p^2} - C(\theta))] \right\} \frac{d_{m'}}{np^2} \right\} \geq \frac{3}{4} ,$$

where K_3 is an universal constant. \square

Acknowledgements

I gratefully thank Pascal Massart and Liliane Bel for many fruitful discussions.

References

- [Aka73] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281. Akadémiai Kiadó, Budapest, 1973.
- [AM09] S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *J. Mach. Learn. Res. (to appear)*, 2009.
- [BCM08] J.P. Baudry, G. Celeux, and J.M. Marin. Selecting models focussing the modeller’s purpose. In *Compstat 2008: Proceedings in Computational Statistics*. Springer-Verlag, 2008.
- [Bes75] J. E. Besag. Statistical Analysis of Non-Lattice Data. *The Statistician*, 24(3):179–195, 1975.
- [BK95] J. E. Besag and C. Kooperberg. On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746, 1995.
- [BM07] L. Birgé and P. Massart. Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields*, 138(1-2):33–73, 2007.
- [Cre85] N. Cressie. Fitting variogram models by weighted least squares. *Mathematical Geology*, 17:563–586, 1985.
- [Cre93] N. A. C. Cressie. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1993.
- [CV08] N. A. C. Cressie and N. Verzelen. Conditional-mean least-squares of Gaussian Markov random fields to Gaussian fields. *Comput. Statist. Data Analysis*, 52(5):2794–2807, 2008.
- [Gra06] R.M. Gray. *Toeplitz and Circulant Matrices: A Review*. Now Publishers, Norwell, Massachusetts, rev. edition, 2006.
- [Guy87] X. Guyon. Estimation d’un champ par pseudo-vraisemblance conditionnelle: étude asymptotique et application au cas markovien. In *Spatial processes and spatial time series analysis (Brussels, 1985)*, volume 11 of *Travaux Rech.*, pages 15–62. Publ. Fac. Univ. Saint-Louis, Brussels, 1987.
- [Guy95] X. Guyon. *Random fields on a network*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.
- [GY99] X. Guyon and J.F. Yao. On the underfitting and overfitting sets of models chosen by order selection criteria. *J. Multivariate Anal.*, 70(2):221–249, 1999.
- [JD01] P.J. Ribeiro Jr. and P.J. Diggle. geoR: A package for geostatistical analysis. *R-NEWS*, 1(2):1609–3631, 2001.

- [Lau96] S. L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York, 1996. Oxford Science Publications.
- [LD93] S. Lakshmanan and H. Derin. Valid parameter space for 2-D Gaussian Markov random fields. *IEEE Trans. Inform. Theory*, 39(2):703–709, 1993.
- [Leb05] E. Lebarbier. Detecting multiple change-points in the mean of a gaussian process by model selection. *Signal processing*, 85(4):717–736, 2005.
- [Lep02] V. Lepez. *Some estimation problems related to oil reserves*. PhD thesis, University Paris XI, 2002.
- [LM00] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 2000.
- [Mas07] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.
- [Mat86] B. Matérn. *Spatial variation*, volume 36 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, second edition, 1986. With a Swedish summary.
- [MM08] C. Maugis and B. Michel. Slope heuristics for variable selection and clustering via gaussian mixtures. Technical Report RR-6550, INRIA, 2008.
- [MT98] A. D. R. McQuarrie and C.-L. Tsai. *Regression and time series model selection*. World Scientific Publishing Co. Inc., River Edge, NJ, 1998.
- [RH05] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, London, 2005.
- [RT02] H. Rue and H. Tjelmeland. Fitting Gaussian Markov random fields to Gaussian fields. *Scand. J. Statist.*, 29(1):31–49, 2002.
- [Sch78] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.
- [SFG08] H.-R. Song, M. Fuentes, and S. Ghosh. A comparative study of gaussian geostatistical models and gaussian markov random field models. *Journal of Multivariate Analysis*, 99:1681–1697, 2008.
- [Ver09] N. Verzelen. Adaptive estimation of regular gaussian markov random fields. Technical Report RR-6797, INRIA, 2009.
- [Vil07] F. Villers. *Tests et sélection de modèles pour l'analyse de données protéomiques et transcriptomiques*. PhD thesis, University Paris XI, December 2007.



Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399