ELSEVIER

Composite likelihood estimation for a Gaussian process under fixed domain asymptotics

François Bachoc [a,*], Moreno Bevilacqua [b], Daira Velandia [c]

[a] *Institut de Mathématiques de Toulouse, France*
[b] *Departamento de Estadística, Universidad de Valparaíso, Millennium Nucleus Center for the Discovery of Structures in Complex Data, Chile*
[c] *Departamento de Estadística, Universidad de Valparaíso, Chile*

ARTICLE INFO

ABSTRACT

We study the problem of estimating the covariance parameters of a one-dimensional Gaussian process with exponential covariance function under fixed-domain asymptotics. We show that the weighted pairwise maximum likelihood estimator of the microergodic parameter can be consistent or inconsistent. This depends on the range of admissible parameter values in the likelihood optimization. On the other hand, the weighted pairwise conditional maximum likelihood estimator is always consistent. Both estimators are also asymptotically Gaussian when they are consistent. Their asymptotic variances are larger or strictly larger than that of the maximum likelihood estimator. A simulation study is presented in order to compare the finite sample behavior of the pairwise likelihood estimators with their asymptotic distributions. For more general covariance functions, an additional inconsistency result is provided, for the weighted pairwise maximum likelihood estimator of a variance parameter.

1. Introduction

Gaussian processes are widely used in statistics to model spatial data. When fitting a Gaussian field, one has to deal with the issue of estimating its covariance function. In many cases, it is assumed that this function belongs to a given parametric model or family of covariance functions, which turns the problem into a parametric estimation problem. Within this framework, the maximum likelihood estimator (MLE) [45,55] of the covariance parameters of a Gaussian stochastic process observed in $\mathbb{R}^d$, $d \geq 1$, has been deeply studied in the two following asymptotic frameworks.

The fixed-domain asymptotic framework, sometimes called infill asymptotics [16,45], corresponds to the case where more and more data are observed in some fixed bounded sampling domain (usually a region of $\mathbb{R}^d$). The increasing-domain asymptotic framework corresponds to the case where the sampling domain increases with the number of observed data. Furthermore, the distance between any two sampling locations is bounded away from 0. The asymptotic behavior of the MLE of the covariance parameters can be quite different under these two frameworks [60].

---

* Corresponding author.
  E-mail address: francois.bachoc@math.univ-toulouse.fr (F. Bachoc).

Under increasing-domain asymptotics, for all covariance parameters, the MLE is consistent and asymptotically normal under some mild regularity conditions. The asymptotic covariance matrix is equal to the inverse of the Fisher information matrix [3,33,41].

The situation is significantly different under fixed-domain asymptotics. Indeed, two types of covariance parameters can be distinguished: microergodic and non-microergodic parameters. A covariance parameter is microergodic if two Gaussian measures constructed from different values of it are orthogonal, see [26,45]. It is non-microergodic if two Gaussian measures constructed from different values of it are equivalent. Non-microergodic parameters cannot be estimated consistently, but misspecifying them asymptotically results in the same statistical inference as specifying them correctly [42–44,59]. In the case of isotropic Matérn covariance functions with $d \leq 3$, [59] shows that only a reparametrized quantity obtained from the scale and variance parameters is microergodic. The asymptotic normality of the MLE of this microergodic parameter is then obtained in [30]. Similar results for the special case of the exponential covariance function were obtained previously in [57].

The maximum likelihood method is generally considered the best option for estimating the covariance parameters of a Gaussian process (at least in the framework of the present paper, where the true covariance function does belong to the parametric model, see also [2,4]). Nevertheless, the evaluation of the likelihood function under the Gaussian assumption requires to solve a system of linear equations and to compute a determinant. For a data set of n observations, the computational burden is $O(n^3)$. This makes maximum likelihood computationally impractical for large data sets. This fact motivates the search for estimation methods with a good balance between computational complexity and statistical efficiency. Among these methods, we can mention low rank approximation (see [46] and the references therein for a review), sparse approximation [24], covariance tapering [21,29], Gaussian Markov random fields approximation [17,38], submodel aggregation [12,18,25,39,48,50] and composite likelihood (CL).

CL indicates a general class of objective functions based on the likelihood of marginal or conditional events [52]. This kind of estimation method has two important benefits. First, it is generally appealing when dealing with large data sets. Second, it can be helpful when it is difficult to specify the full likelihood. In this paper we focus on a specific type of CL called pairwise likelihood. Examples of the use of pairwise likelihood in the Gaussian and non Gaussian case can be found in [9,10,20,22,23] to name a few.

The increasing-domain asymptotic properties of various weighted pairwise likelihood estimators in the Gaussian case can be found in [9]. Under this setting, for all covariance parameters, these pairwise likelihood estimators are consistent and asymptotically normal under some mild regularity conditions.

There are no results under fixed-domain asymptotics for CL to the best of our knowledge. In this paper we study the fixed-domain asymptotic properties of the pairwise likelihood estimation method. A one-dimensional Gaussian process with exponential covariance model is considered. This covariance model is particularly amenable to theoretical analysis of the MLE. Indeed, its Markovian property yields an explicit (matrix-free) expression of the likelihood function [57]. Thus, the MLE for this model has been studied in [14,15,57], and also in higher dimension in [1,51,58].

Under the parametric model given by the one-dimensional exponential covariance function, the microergodic parameter is the product of the variance and the scale (see [57] and Section 2). Under the same setting as in [57], we study the consistency and asymptotic normality of the weighted pairwise maximum likelihood estimator (WPMLE) and of the weighted pairwise conditional maximum likelihood estimator (WPCMLE) of the microergodic parameter. Our results show that the WPMLE is inconsistent if the range of admissible values for the variance parameter is too large compared to that for the scale parameter (Theorem 1, (i) and (iii)). Conversely, they show that the WPMLE is consistent if the range of admissible values for the variance parameter is restricted enough compared to that for the scale parameter (Theorem 1, (ii) and (iv)). In contrasts, we prove that the WPCMLE is consistent regardless of the ranges of admissible values for the scale and variance (Theorem 2). Both estimators are also asymptotically Gaussian in the cases where they are consistent (Theorem 3). The asymptotic variance of these estimators is no smaller than that of the MLE. Furthermore, it is strictly larger for equispaced observation points, when non-zero weights are given to pairs of non successive observation points.

The proofs of the asymptotic results rely on Lemma 1. This lemma expresses the weighted pairwise (conditional) likelihood criteria as combinations of full likelihood criteria for subsets of the observations. This enables to exploit the asymptotic approximations obtained by [57] for the full likelihood. The consistency results are then obtained similarly as in [57]. The inconsistency results are obtained by a careful analysis of the limit weighted likelihood criterion, together with results on extrema of Gaussian processes (see the proof of Theorem 1). Finally, the asymptotic normality results are obtained by means of a central limit theorem for weakly dependent variables, together with the use of Lemma 1 and of the approximations in [57].

It turns out that a suitable choice of the symmetric weights used in the pairwise likelihood is crucial for improving the asymptotic statistical efficiency of the method. In particular, the choice of compactly supported weights with a specific cut-off generally decreases the asymptotic variance. As a special case, if the pairwise likelihood is constructed only from adjacent pairs, then the asymptotic efficiency of the MLE is attained.

These results are consistent with the increasing-domain setting. Indeed, it has been shown, via simulation, that weighting schemes downweighting observations that are far apart in time and/or in space, are preferable to the pairwise likelihood involving all possible pairs [9,27,53]. A simulation study compares the finite sample properties of the weighted pairwise likelihood estimators with their asymptotic distributions, using the MLE as a benchmark. Finally, we provide an additional inconsistency result for the WPMLE of a variance parameter, when the correlation function is known and allowed to be significantly more general than the exponential one.

The paper falls into the following parts. In Section 2, pairwise (conditional) likelihood estimation is introduced. In Section 3, the consistency and inconsistency results are provided. In Section 4, the asymptotic normality results are provided, together with the analysis of the asymptotic variance. In Section 5, the numerical results are discussed. Concluding remarks are given in Section 6. The proofs are postponed to Appendix A. The more general inconsistency result of the WPMLE is provided in Appendix B.

2. Pairwise (conditional) likelihood estimation

In the following, we will consider a stationary zero-mean Gaussian process $Z = \{Z(s), s \in [0, 1]\}$, with covariance function given by, for $s, t \in \mathbb{R}$,

$$\text{cov}\{Z(s), Z(t)\} = \sigma_0^2 e^{-\theta_0 |s-t|} \tag{1}$$

for fixed $\theta_0, \sigma_0^2 \in (0, \infty)$.

For $\theta, \sigma^2 \in (0, \infty)$, let $\psi = (\theta, \sigma^2)^\top$, call σ^2 the variance parameter and call θ the scale (inverse of correlation decay) parameter. Let $\psi_0 = (\theta_0, \sigma_0^2)^\top$. Throughout this paper, θ and σ^2 denote candidate covariance parameters, at which likelihood or pairwise likelihood criteria can be evaluated (see (2) and (3)), while θ_0 and σ_0^2 are the true covariance parameters, that provide the true covariance function of Z through (1).

The Gaussian process Z is known as the stationary Ornstein–Uhlenbeck process. Let $\left(s_i^{(n)}\right)_{n \in \mathbb{N}, i \in \{1, \ldots, n\}}$ be a triangular array of observations points in $[0, 1]$. For $n \in \mathbb{N}$, let $\{s_1, \ldots, s_n\} = \{s_1^{(n)}, \ldots, s_n^{(n)}\}$ for simplicity and assume without loss of generality that $s_1 < \cdots < s_n$. The observation vector can thus be written as $Z_n = (Z(s_1), \ldots, Z(s_n))^\top$.

For $s, t \in [0, 1]$, we let

$$\ell_{s,t}(\psi) = 2\ln(\sigma^2) + \ln(1 - e^{-2\theta|s-t|}) + \frac{1}{\sigma^2}Z(s)^2 + \frac{\{Z(t) - e^{-\theta|s-t|}Z(s)\}^2}{\sigma^2(1 - e^{-2\theta|s-t|})}$$

be the log-likelihood criterion associated to the bivariate random vector $(Z(s), Z(t))^\top$. The quantity $\ell_{s,t}(\psi)$ is -2 times the logarithm of the probability density function of the vector $(Z(s), Z(t))^\top$, when considering that Z has covariance function given by (1) with ψ_0 replaced by ψ.

We also let

$$\ell_{t|s}(\psi) = \ln(\sigma^2) + \ln(1 - e^{-2\theta|s-t|}) + \frac{\{Z(t) - e^{-\theta|s-t|}Z(s)\}^2}{\sigma^2(1 - e^{-2\theta|s-t|})}$$

be the conditional log-likelihood criterion of $Z(t)$ given $Z(s)$. The quantity $\ell_{t|s}(\psi)$ is -2 times the logarithm of the conditional probability density function of $Z(t)$ given $Z(s)$, when considering that Z has covariance function given by (1) with ψ_0 replaced by ψ.

We let $(w_{i,j}^{(n)})_{n \in \mathbb{N}, i,j \in \{1, \ldots, n\}, i \neq j}$ be a triangular array of weights in $[0, \infty)$. We let $w_{i,j}^{(n)} = w_{i,j}$ for simplicity. Then, define the weighted pairwise log-likelihood function $p\ell_n$ as

$$p\ell_n(\psi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \ell_{s_i, s_j}(\psi) \tag{2}$$

and the weighted pairwise conditional log-likelihood function $pc\ell_n$ as

$$pc\ell_n(\psi) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} w_{i,j} \ell_{s_j|s_i}(\psi). \tag{3}$$

Finally, for $J \subset (0, \infty)^2$, define the WPMLE as

$$\widehat{\psi}_{p\ell} = \underset{\psi \in J}{\text{argmin}} \; p\ell_n(\psi) \tag{4}$$

and the WPCMLE as

$$\widehat{\psi}_{pc\ell} = \underset{\psi \in J}{\text{argmin}} \; pc\ell_n(\psi). \tag{5}$$

We remark that the WPMLE and WPCMLE are two distinct estimators, which will be shown in Section 3 to have significantly different behaviors under fixed-domain asymptotics.

In the rest of the paper, we will consider the case where the weights $(w_{i,j}^{(n)})$ are so that for any $n \in \mathbb{N}$ and $i, j \in \{1, \ldots, n\}, i \neq j$, $w_{i,j} = w_{|i-j|}$ with $(w_k)_{k \in \mathbb{N}}$ a fixed sequence of non-negative numbers, not all of which being zero. We let $K = \max\{i \in \mathbb{N}; w_i \neq 0\}$, where we allow for $K = +\infty$ if $(w_k)_{k \in \mathbb{N}}$ has infinitely many non-zero elements. The quantity K is fixed independently of n and $K < +\infty$ if an only if the sequence $(w_k)_{k \in \mathbb{N}}$ has finitely many non-zero elements. We remark that $K \geq 1$ since $(w_k)_{k \in \mathbb{N}}$ is not the zero sequence.

In the following lemma, convenient expressions are given for the weighted pairwise log-likelihood and weighted pairwise conditional log-likelihood functions.

Lemma 1. *Let, for $k \in \{1, \ldots, n-1\}$ and $a \in \{0, \ldots, k-1\}$, $x_j^{(k,a)} = s_{1+a+jk}$, for $j \in \mathbb{N}$ so that $1 + a + jk \le n$. Then, with $\lfloor x \rfloor$ the largest integer that is smaller or equal to x, and with the convention that $\sum_{i=a}^{b} r_i = 0$ if $b < a$,*

$$p\ell_n(\psi) = \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_j^{(k,a)}, x_{j+1}^{(k,a)}}(\psi) \tag{6}$$

and

$$pc\ell_n(\psi) = \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \left\{ \ell_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\psi) + \ell_{x_j^{(k,a)} | x_{j+1}^{(k,a)}}(\psi) \right\}. \tag{7}$$

The benefit of Lemma 1 is that the rightmost sum in (7) can be expressed as a (full) log likelihood for a subset (that depends on k and a in (7)) of $\{s_1, \ldots, s_n\}$ (see Lemma 1 in [57]). Similarly, the rightmost sum in (6) can be expressed conveniently as a function of this (full) log likelihood. We refer to the proofs of Theorems 1–3 for details.

3. Consistency and inconsistency

It is well-known that, in the case of the exponential covariance model given in (1), the parameters σ^2 and θ are non-microergodic and only the parameter $\sigma^2 \theta$ is microergodic [57,59]. Hence, we study the consistency, inconsistency and asymptotic normality for the estimators $\hat{\theta}_{p\ell} \hat{\sigma}_{p\ell}^2$ and $\hat{\theta}_{pc\ell} \hat{\sigma}_{pc\ell}^2$ of $\sigma_0^2 \theta_0$ given in (4) and (5).

In the next theorem, let J be of the form $[a, b] \times [c, d]$, $[a, b] \times (0, \infty)$ or $(0, \infty) \times [c, d]$. The next theorem characterizes, as a function of J, the cases where the WPMLE is consistent and those where it is inconsistent.

Theorem 1. *Assume that $K < +\infty$, that $J \subset (0, \infty)^2$ and that there exists $(\tilde{\theta}, \tilde{\sigma}^2)$ in J so that $\tilde{\theta} \tilde{\sigma}^2 = \theta_0 \sigma_0^2$. Then, for fixed $0 < a \le b < +\infty$ and $0 < c \le d < +\infty$,*

  (i) *If $J = [a, b] \times [c, d]$ with $ad > \theta_0 \sigma_0^2$ or $bc < \theta_0 \sigma_0^2$, then $\hat{\theta}_{p\ell} \hat{\sigma}_{p\ell}^2$ does not converge in probability to $\theta_0 \sigma_0^2$;*

  (ii) *If $J = [a, b] \times [c, d]$ with $ad \le \theta_0 \sigma_0^2$ and $bc \ge \theta_0 \sigma_0^2$, then $\hat{\theta}_{p\ell} \hat{\sigma}_{p\ell}^2$ converges to $\theta_0 \sigma_0^2$ almost surely;*

  (iii) *If $J = [a, b] \times (0, \infty)$ then $\hat{\theta}_{p\ell} \hat{\sigma}_{p\ell}^2$ does not converge in probability to $\theta_0 \sigma_0^2$;*

  (iv) *If $J = (0, \infty) \times [c, d]$ then $\hat{\theta}_{p\ell} \hat{\sigma}_{p\ell}^2$ converges to $\theta_0 \sigma_0^2$ almost surely.*

Remark 1. In Theorem 1, the parameter set J always contains a pair $(\tilde{\theta}, \tilde{\sigma}^2)$ so that $\tilde{\theta} \tilde{\sigma}^2 = \theta_0 \sigma_0^2$, where we recall that $\theta_0 \sigma_0^2$ is the unknown microergodic parameter. Thus, in the framework of Theorem 1, the parameter $\theta_0 \sigma_0^2$ is always consistently estimable, and indeed Theorem 2 shows that there exists an estimator, namely the WPCMLE, that is consistent. This does not mean that all the estimators of $\theta_0 \sigma_0^2$ are consistent, and indeed Theorem 1 shows that the WPMLE can be inconsistent.

Theorem 1 can be interpreted as follows: The pairwise likelihood criterion can be decomposed as

$$p\ell_n(\psi) = S_1(\sigma^2) + S_2(\theta, \sigma^2),$$

where $S_1(\sigma^2)$ is the first triple sum in (A.1) in the proof of Theorem 1 and $S_2(\theta, \sigma^2)$ is the second one. The minimizer of the random function S_2 of θ and σ^2 would yield a consistent estimator of $\theta_0 \sigma_0^2$. However, the random function S_1 has random fluctuations that do not become negligible as $n \to \infty$. The consequence of these fluctuations is that the WPMLE of $\theta_0 \sigma_0^2$ is inconsistent if the range of possible values for σ^2 is too large compared to that for θ (cases (i) and (iii)). On the contrary, if the range of possible values for θ is large enough compared to that for σ^2, then the WPMLE $\hat{\theta}_{p\ell}$ is able to compensate, so to speak, the non-negligible fluctuations of S_1. More precisely, independently of the value of $\hat{\sigma}_{p\ell}^2$, in the cases (ii) and (iv), because of the good behavior of S_2, the value of $\hat{\theta}_{p\ell}$ will be so that $\hat{\theta}_{p\ell} \hat{\sigma}_{p\ell}^2$ does converge to $\theta_0 \sigma_0^2$. This discussion is the basis of the proof of Theorem 1, to which we refer for more details. We remark that this discussion applies to the WPMLE, and not to the WPCMLE which is a different estimator and is always consistent, see Theorem 2.

It should be mentioned that, under increasing-domain asymptotics, the WPMLE of (θ_0, σ_0^2), and thus of $\theta_0 \sigma_0^2$, is consistent, independently of the form of J, provided J contains (θ_0, σ_0^2) (see the general results given in [9]). It can indeed be checked that $S_1(\sigma^2)$ has random fluctuations that are typically asymptotically negligible under increasing-domain asymptotics. Hence, we have an additional illustration of the important qualitative differences that can appear when comparing fixed and increasing-domain asymptotics (see also [60]).

From a practical standpoint, Theorem 1 provides a warning when using the WPMLE. In particular, a situation that often occurs in practice is the case (iii), where θ has a compact range of allowed values and where for any fixed θ, the value of σ^2 that optimizes the pairwise likelihood (without restriction) can be computed explicitly. This case (iii) is a case of inconsistent estimation of the microergodic parameter. Furthermore, in Appendix B, we provide another inconsistency result for the WPMLE, for more general covariance functions and when a single variance parameter is estimated.

The following theorem gives the consistency of the WPCMLE for the microergodic parameter, for all the cases for J under investigation in Theorem 1. Hence, Theorems 1 and 2 jointly provide an incentive to use pairwise conditional likelihood rather than pairwise (unconditional) likelihood in practice.

Theorem 2. *Assume that $K < +\infty$. Let $0 < a \leq b < +\infty$ and $0 < c \leq d < +\infty$ be fixed. Let $J = [a, b] \times [c, d]$ or $J = [a, b] \times (0, \infty)$ or $J = (0, \infty) \times [c, d]$ and assume that there exist $\tilde{\theta}, \tilde{\sigma}^2$ in J so that $\tilde{\theta}\tilde{\sigma}^2 = \theta_0\sigma_0^2$. Then, $\hat{\theta}_{pc\ell}\hat{\sigma}^2_{pc\ell}$ converges to $\theta_0\sigma_0^2$ almost surely.*

4. Asymptotic normality

Throughout this section, we assume that $\max_{i \in \{1,\ldots,n-1\}}(s_{i+1} - s_i) \to 0$ as $n \to \infty$. In order to express the asymptotic variance, in the asymptotic normality result in Theorem 3, let us define, for $i \in \mathbb{N}$ and $k \in \mathbb{N}$, with $i \geq 1$ and $i + k \leq n$,

$$W^2_{i,i+k} = \frac{\left\{Z(s_{i+k}) - e^{-\theta_0(s_{i+k}-s_i)}Z(s_i)\right\}^2}{\sigma_0^2(1 - e^{-2\theta_0(s_{i+k}-s_i)})}.$$

Let us also define

$$\tau_n^2 = \frac{1}{n}\mathrm{var}\left\{\sum_{i=1}^{n-1}\sum_{k=1}^{\min\{K,n-i\}} w_k\left(W^2_{i,i+k} - 1\right)\right\}. \tag{8}$$

Lemma 2 provides an asymptotic approximation of τ_n^2 when $K < +\infty$.

Lemma 2. *Assume that $K < +\infty$. Let, for $1 \leq i < j \leq n$ and $1 \leq k < \ell \leq n$, $b_{i,j,k,\ell}$ be defined by $b_{i,j,k,\ell} = b_{k,\ell,i,j}$ and, for $i \leq k$,*

$$b_{i,j,k,\ell} = \begin{cases} 0 & \text{if } j \leq k, \\ \frac{(s_j-s_k)^2}{(s_j-s_i)(s_\ell-s_k)} & \text{if } k \leq j \leq \ell, \\ \frac{s_\ell-s_k}{s_j-s_i} & \text{if } k \leq \ell \leq j. \end{cases} \tag{9}$$

We have as $n \to \infty$

$$\tau_n^2 = \left(\frac{2}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{\min(n,i+K)}\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{\min(n,k+K)} w_{j-i}w_{\ell-k}b_{i,j,k,\ell}\right) + o(1). \tag{10}$$

In Lemmas 3 and 4, we show that τ_n^2 is lower and upper bounded as $n \to \infty$.

Lemma 3. *Assume that $K < +\infty$. Then $\liminf_{n\to\infty} \tau_n^2 \geq 2(\sum_{k=1}^K w_k)^2$.*

Lemma 4. *Assume that $K < +\infty$. Then $\limsup_{n\to\infty} \tau_n^2 < +\infty$.*

Then, the following theorem establishes the asymptotic normality of the WPMLE (in the cases where it is consistent) and of the WPCMLE (in all the cases) of the microergodic parameter.

Theorem 3. *Under the cases (ii) and (iv) of Theorem 1, and under the same conditions as in this theorem,*

$$\frac{\sqrt{n}(\sum_{k=1}^K w_k)}{\tau_n\sigma_0^2\theta_0}(\hat{\sigma}^2_{p\ell}\hat{\theta}_{p\ell} - \sigma_0^2\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{11}$$

Furthermore, under the same conditions as in Theorem 2,

$$\frac{\sqrt{n}(\sum_{k=1}^K w_k)}{\tau_n\sigma_0^2\theta_0}(\hat{\sigma}^2_{pc\ell}\hat{\theta}_{pc\ell} - \sigma_0^2\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{12}$$

Because of Lemmas 3 and 4, the rate of convergence is $n^{1/2}$ in Theorem 3. Furthermore, it is shown in [57] that the asymptotic variance of the MLE of $\theta_0\sigma_0^2$ is $2(\theta_0\sigma_0^2)^2$. Hence, from Lemma 3, the WPMLE and WPCMLE have a larger asymptotic variance than the MLE, which is in agreement with the fact that the MLE is generally considered to be the most efficient statistically. Also, the asymptotic variance of the WPMLE and WPCMLE depends on the triangular array of observation points (see Lemma 2), while that of the MLE does not. A similar conclusion was obtained in [6] when considering a cross validation estimator in the same context.

Remark 2. It can be shown, from Lemma 1 and from Lemma 1 in [57], that the WPMLE and WPCMLE asymptotically coincide with the MLE when $K = 1$ (that is when $w_1 > 0$ and $w_k = 0$ for $k \geq 2$) and under the conditions of Theorem 3. When $K = 1$, it is easily seen from (8) that $\tau_n^2 = 2w_1^2 + o(1)$ so that, indeed, the WPMLE and WPCMLE have the same asymptotic variance as the MLE.

The following lemma shows that for equispaced observation points, the asymptotic variance of the WPMLE and WPCMLE is strictly larger than that of the MLE when one of the w_k, $k \geq 2$ is non-zero.

Lemma 5. *If $s_i = i/n$ for all $n \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$ and if $2 \leq K < +\infty$, then $\liminf_{n \to \infty} \tau_n^2 > 2 \left(\sum_{i=1}^{K} w_k \right)^2$.*

Remark 3. The sequence of weights $\{w_1, 0, \ldots\}$ with $w_1 > 0$ provides the smallest asymptotic variance because of the fact that, in this case, the log-likelihood function is asymptotically equal to a sum of pairwise conditional log likelihoods for consecutive observation points (see Lemma 1 in [57]). This is specific to the exponential covariance function in (1). For more general families of covariance function, it is possible that it is more efficient to use other weight configurations than $\{w_1, 0, \ldots\}$.

Remark 4. The log likelihood criterion can be evaluated with a $O(n)$ computational cost (see Lemma 1 in [57]). Hence, here the WPMLE, WPCMLE and MLE have the same computational cost but in general, the likelihood evaluation cost is $O(n^3)$ in time and $O(n^2)$ in storage, while this cost remains $O(n)$ for weighted pairwise likelihood criteria (when $K < +\infty$) irrespectively of the covariance function.

Remark 5. In Sections 3 and 4, we have provided consistency and asymptotic normality results under fixed-domain asymptotics. As discussed after Theorem 1, some of the conclusions obtained differ from those arising from an increasing-domain asymptotic analysis. Hence, given a finite set of observation points and observed values, the question of whether to consider increasing-domain or fixed-domain asymptotics as most relevant is important.

Let us discuss this question in the general setting of a stationary isotropic Gaussian random field on $\mathbb{R}^d$. Let us consider the correlation distance $\Delta \geq 0$, such that the correlation for two points at distance more than Δ is expected to be less than a given threshold, for instance 0.05. Let us also consider that there are n observation points $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, and consider the diameter $D = \max_{i,j \in \{1,\ldots,n\}} \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ of the point cloud. Then, a first rule of thumb is to consider that fixed-domain asymptotics is relevant when n is large compared to $(D/\Delta)^d$. Conversely, increasing-domain asymptotics is relevant if n is comparable to $(D/\Delta)^d$. Hence, fixed-domain asymptotics is more relevant when D decreases and Δ increases and conversely for increasing-domain asymptotics. We also remark that when n is large compared to $(D/\Delta)^d$ and D/Δ is large, that is the observation domain is large and the observation points are dense in this domain, then a third asymptotic setting called mixed-domain asymptotics [14] can be most relevant, although less theory is available in this case.

Beyond this rule of thumb, the reference [60] is, to the best of our knowledge, the only one that provides a quantitative comparison of the approximations of finite sample situations, brought by fixed-domain and increasing-domain asymptotics. The authors focus on maximum likelihood estimation of covariance parameters, and consider settings where two asymptotic distributions of the estimators can be obtained from both asymptotic frameworks. In simulations, they show that the fixed-domain asymptotic distributions are closer to the finite sample distributions than the increasing-domain ones, especially for the microergodic covariance parameters.

We can also mention other specific settings where fixed-domain asymptotics is more relevant. When studying the properties of sequential procedures, like in optimization [7,54], fixed-domain asymptotics is more natural because the definition domain of the Gaussian process is fixed, and more and more observation points are added sequentially. Similarly, Gaussian processes conditioned by inequality constraints (for instance boundedness or monotonicity) can be studied under fixed-domain asymptotics [5,32], but much less so under increasing-domain asymptotics, because the probability that a Gaussian process, defined on an unbounded domain, satisfy these types of constraints is generally zero.

Hence, generally speaking, it seems that fixed-domain asymptotics is perhaps more often relevant than increasing-domain asymptotics, as pointed out in the reference book [45]. On the other hand, the obtained theoretical results can be significantly more general, in terms of types of covariance functions and observation point locations, under increasing-domain asymptotics. Thus the joint study of both these frameworks is justified.

Finally, an ideal scenario is when both the increasing and fixed-domain asymptotic frameworks yield practical procedures that coincide, on fixed finite data sets. For instance, in [35], the asymptotic normal approximation given by the Fisher information matrix is used to derive efficient sampling schemes for covariance parameter estimation and prediction. Here, the Fisher information matrix is computed from finite sample data sets, so that the algorithm is motivated by both the increasing and fixed-domain frameworks. In this example, in order to be robust to the type of asymptotic setting used, we recommend to indeed compute the finite sample Fisher information matrix, rather than its theoretical limit under a given asymptotic framework.

5. Numerical experiments

The main goal of this section is to compare the finite sample behavior of the WPMLE and WPCMLE of the microergodic parameter of the exponential covariance model in (1) with the asymptotic distributions given in Theorem 3, using the MLE as a benchmark.

In a first simulation study we consider a set of points in $[0, 1]$ defined by $s_{i+1} = s_i + 0.02/L$ with $s_1 = 0$ and $s_n = 1$. We let $L = 1, 2, 4, 8, 16$, that is we consider $n = 51, 101, 201, 401, 801$ points.

For each L, with the Cholesky decomposition, we simulate 5000 realizations of a zero mean Gaussian process with covariance (1) setting $\sigma_0^2 = 1$ and $\theta_0 = 15$. For each simulation we estimate with MLE, WPMLE and WPCMLE choosing the 'optimal' weights $w_1 = 1$ and $w_k = 0$ if $k \geq 2$.

Table 1

Sample quantiles, mean, variance and kurtosis of $\left(\sqrt{n}/C_y\right)\left(\hat{\sigma}_{i,y}^2\hat{\theta}_{i,y}/\sigma_0^2\theta_0 - 1\right)$, $i \in \{1,\ldots,5000\}$ for y = WPMLE, WPCMLE, MLE when $n = 51, 101, 201, 401, 801$, compared with the associated theoretical values of the standard Gaussian distribution.

Method	n	5%	25%	50%	75%	95%	mean	var	Kur
WPMLE	51	−1.6186	−0.7006	0.0264	0.8950	2.3767	0.1703	1.5644	1.6348
	101	−1.5899	−0.6864	0.0278	0.8043	2.0501	0.1075	1.2520	0.2967
	201	−1.5733	−0.6702	−0.0029	0.7400	1.8353	0.0497	1.0862	0.0933
	401	−1.6121	−0.6767	0.0171	0.7275	1.7956	0.0419	1.0635	−0.0321
	801	−1.6275	−0.6728	0.0099	0.7229	1.7001	0.0303	1.0350	−0.0133
WPCMLE	51	−1.6186	−0.7006	0.0264	0.8950	2.3767	0.1703	1.5644	1.6348
	101	−1.5899	−0.6864	0.0278	0.8043	2.0501	0.1075	1.2520	0.2967
	201	−1.5733	−0.6702	−0.0029	0.7400	1.8353	0.0497	1.0862	0.0933
	401	−1.6121	−0.6767	0.0171	0.7275	1.7956	0.0419	1.0635	−0.0321
	801	−1.6275	−0.6728	0.0099	0.7229	1.7001	0.0303	1.0350	−0.0133
MLE	51	−1.5540	−0.7043	−0.0585	0.6733	1.8372	0.0197	1.0891	0.3874
	101	−1.5556	−0.7173	−0.0289	0.6759	1.8059	0.0215	1.0687	0.0327
	201	−1.5702	−0.7038	−0.0476	0.6478	1.7007	−0.0034	1.0062	0.0316
	401	−1.6194	−0.6909	−0.0254	0.6770	1.7012	0.0048	1.0288	−0.0607
	801	−1.6226	−0.6886	−0.0074	0.6880	1.6634	0.0052	1.0195	−0.0049
$\mathcal{N}(0,1)$		−1.6448	−0.6745	0.0000	0.6745	1.6448	0	1	0
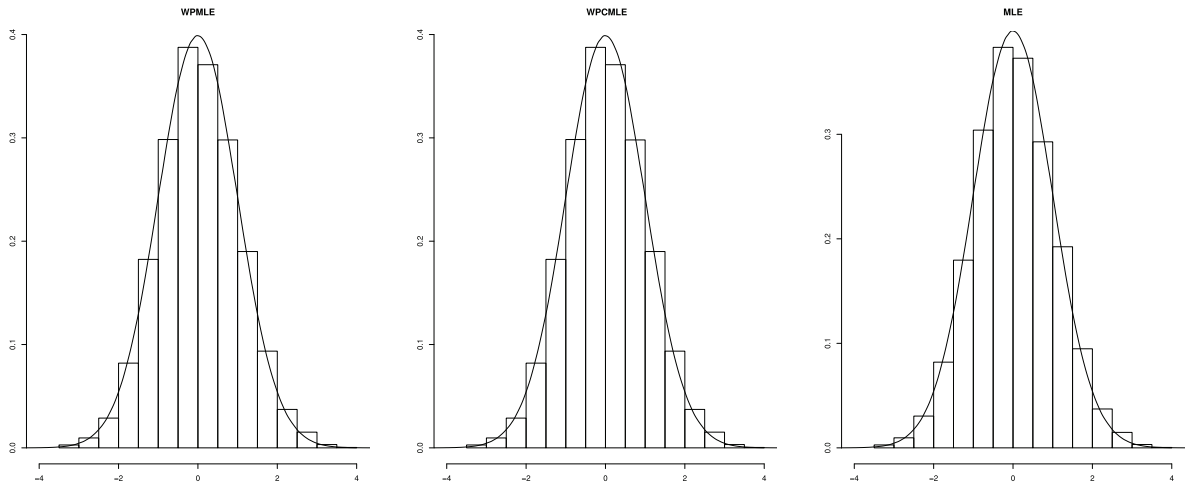
Fig. 1. Comparison of the histogram of $\left(\sqrt{n}/C_y\right)\left(\hat{\sigma}_{i,y}^2\hat{\theta}_{i,y}/\sigma_0^2\theta_0 - 1\right)$, $i \in \{1,\ldots,5000\}$ for y = WPMLE, WPCMLE, MLE (from left to right) with the standard Gaussian distribution when $n = 801$.

Optimization was carried out using the R [36] function `optim` with the method "L-BFGS-B" [11] which allows box constraints, that is each variable can be given a lower and/or upper bound. Specifically we set $[a,b] \times [c,d] = [0.01, 2500] \times [0.01, 5]$. From Theorems 1 and 2, under this setting, both the WPMLE and WPCMLE are consistent. Using the asymptotic distributions stated in Theorem 3, Table 1 compares the sample quantiles of order 0.05, 0.25, 0.5, 0.75, 0.95, mean, variance and kurtosis of the vector for $i \in \{1,\ldots,5000\}$

$$\frac{\sqrt{n}}{C_y}\left(\frac{\hat{\sigma}_{i,y}^2\hat{\theta}_{i,y}}{\sigma_0^2\theta_0} - 1\right),$$

where y = MLE, WPMLE and WPCMLE with the associated theoretical values of the standard Gaussian distribution. Here $C_{MLE}^2 = 2$ and $C_{WPMLE}^2 = C_{WPCMLE}^2 = \hat{\tau}_n^2/(\sum_{k=1}^K w_k)^2$ where $\hat{\tau}_n^2$ is computed using the approximation in (10), i.e.,

$$\hat{\tau}_n^2 = \frac{2}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{\min(n,i+K)}\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{\min(n,k+K)} w_{j-i}w_{\ell-k}b_{i,j,k,\ell}.$$

Under the optimal weights setting and following Remark 1 we have $C_{WPMLE}^2 = C_{WPCMLE}^2 \approx 2$ as in the MLE case.

It is apparent from Table 1 that the asymptotic distribution given in Theorem 3 is a satisfactory approximation of the sample distribution, visually improving when increasing n. It can be appreciated that the speed of convergence is

Table 2

Asymptotic variance versus sample variance of the WPMLE, WPCMLE and MLE of the microergodic parameter when increasing n and K.

	K	$n = 51$	$n = 101$	$n = 201$	$n = 401$	$n = 801$
Sample variance MLE	*	10.0650	4.8688	2.2809	1.1523	0.5749
Asymptotic variance		8.8230	4.4554	2.2388	1.1221	0.5618
Sample variance WPMLE	1	13.5326	5.5227	2.4197	1.1905	0.5807
Sample variance WPCMLE		13.5326	5.5227	2.4197	1.1905	0.5807
Asymptotic variance		8.6505	4.4113	2.2277	1.1193	0.5611
Sample variance WPMLE	10	33.5086	17.9724	7.9901	3.6650	1.7723
Sample variance WPCMLE		33.1477	17.7529	7.9465	3.6605	1.7720
Asymptotic variance		22.2798	12.3865	6.5138	3.3382	1.6895
Sample variance WPMLE	20	36.1910	24.7919	15.0243	7.1731	3.3934
Sample variance WPCMLE		35.9582	24.4611	14.8164	7.1312	3.3894
Asymptotic variance		32.9936	20.7760	11.5892	6.0239	3.0823
Sample variance WPMLE	30	36.6338	26.7294	19.7646	10.8759	5.1132
Sample variance WPCMLE		36.4643	26.4491	19.4406	10.7569	5.0969
Asymptotic variance		36.4154	25.3444	16.0245	8.6089	4.4547

slightly faster for MLE and that the asymptotic behavior of the WPMLE and WPCMLE is similar. Fig. 1 depicts a graphical comparison when $n = 801$.

In a second simulation study we focus only on the asymptotic variance $n^{-1}(\sigma_0^2 \theta_0 C_y)^2$ and we compare it with the sample variance of the WPMLE, WPCMLE and MLE of the microergodic parameter, under the same setting as in the first numerical experiment. In this case we consider increasing values of both K and n that is we consider also non optimal weights. The results are depicted in Table 2. For fixed K it can be appreciated that, as expected, the difference between the sample and theoretical variance is reduced when increasing n. On the other hand, for fixed n, both asymptotic and simulated variances increase when increasing K. Overall, as expected, the asymptotic variance slightly underestimates the simulated variance. Finally note that when $K \neq 1$ then the speed of convergence for the WPCMLE is slightly faster than for the WPMLE.

Let us conclude this section with a remark. For a fixed θ, the global minimizer of $p\ell_n(\theta, \sigma^2)$ and of $pc\ell_n(\theta, \sigma^2)$, over $\sigma^2 \in (0, +\infty)$, can be expressed explicitly. Hence, from a computational point of view, it is possible to compute explicitly $\min_{\sigma^2 \in (0,+\infty)} p\ell_n(\theta, \sigma^2)$ and $\min_{\sigma^2 \in (0,+\infty)} pc\ell_n(\theta, \sigma^2)$. These two latter functions can be minimized numerically over θ only, which is computationally beneficial. However, this case corresponds to the case (iii) in Theorem 1 so the WPMLE would be inconsistent in this setting.

6. Conclusions

In this paper, we provide, to the best of our knowledge, the first fixed-domain asymptotic properties of weighted pairwise likelihood estimators. We consider the exponential covariance function in dimension one. This enables us to exploit the results of [57]. By means of these results, and of specific new proof techniques, the following conclusions are obtained. The WPMLE can be inconsistent, if the range of allowed values for the variance is too large. This is in contrast with the increasing-domain asymptotic situation, and allows us to issue a warning for practical use. On the other hand, the WPCMLE is always consistent. In case of consistency, both these estimators are asymptotically Gaussian distributed, with an asymptotic variance that is larger than or equal to that of the MLE. This variance depends on the observation point locations and on the weight configurations.

Some of the conclusions obtained in this paper are specific to the exponential family of covariance functions. In particular, the MLE is here asymptotically equal to a WPCMLE based only on pairs of consecutive observation points. On the other hand, the inconsistency of the WPMLE that we have pointed out is not specific to the exponential covariance function. Indeed, Appendix B provides inconsistency results for the WPMLE for general covariance functions.

It would be interesting to study the consistency of the WPCMLE and its asymptotic normality for more general families of covariance functions than the exponential one, and in dimension larger than one. The present paper constitutes a first step in this more general direction. We remark that, in the maximum likelihood case, general results on fixed-domain asymptotics and for $d = 1, 2, 3$ can be found in [30] and [59] for the Matérn model and in [8] for the Generalized Wendland model.

As a referee pointed out, other types of CL functions have been proposed in the literature for the estimation of the covariance functions of Gaussian processes [19,28,47]. [31] highlights that, for a given estimation problem, the choice of a suitable CL function should be driven by statistical and computational considerations. In this paper we focused only on CL based in pairs since, from our perspective, there is a clear computational advantage when considering CL based only on pairs of observations. Moreover CL methods based on pairs are more robust (in the sense of [52] and [56]) with respect to other types of CL functions. Indeed, they require only the correct specification of bivariate random vectors. Nevertheless, we believe that for a one dimensional Gaussian process with exponential covariance function, the results in our paper

and in [57] are the starting points to study the consistency and the asymptotic distribution of other types of CL estimators under fixed-domain asymptotics.

Finally, it would be interesting to study the fixed-domain asymptotic properties of CL counterparts of the model selection criteria of [13], that are based on the full likelihood.

Acknowledgments

We thank the editor in chief, the associate editor and two anonymous referees for their constructive feedback and helpful suggestions.

Appendix A. Proofs for Sections 2 to 4

In the proofs, we let $0 < C_{\inf} < +\infty$ and $0 < C_{\sup} < +\infty$ be generic constants, which values can change from line to line.

Proof of Lemma 1. The principle of the proof is a change of indexation of the pairs of the form $s_i, s_j, i < j$. Instead of indexing such a pair by i and j, we index it by $j - i$ and by the remainder and quotient of the Euclidean division of i by $j - i$. We have

$$
p\ell_n(\psi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \ell_{s_i, s_j}(\psi) = \sum_{k=1}^{n-1} \sum_{i=1}^{n-1} w_k \mathbf{1}_{i+k \le n} \ell_{s_i, s_{i+k}}(\psi)
$$

$$
= \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{n} \mathbf{1}_{1+a+jk+k \le n} \ell_{s_{1+a+jk}, s_{1+a+jk+k}}(\psi) = \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_j^{(k,a)}, x_{j+1}^{(k,a)}}(\psi).
$$

The proof of (7) is the same, after observing that

$$
\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \ne i}}^{n} w_{i,j} \ell_{s_j|s_i}(\psi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \left\{ \ell_{s_j|s_i}(\psi) + \ell_{s_i|s_j}(\psi) \right\}. \quad \square
$$

Lemma 6. *Let m and M be fixed with $-\infty \le m < M \le +\infty$. Then, we have, with probability $\Pr > 0$ (not depending on n),*

$$
\forall t \in [0, 1], m \le Z(t) \le M.
$$

Proof. The lemma is a special case of Lemma A.3 in [32]. $\quad \square$

Proof of Theorem 1. Let us first consider the case (i). We have, from Lemma 1

$$
p\ell_n(\psi) = \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_j^{(k,a)}, x_{j+1}^{(k,a)}}(\psi)
$$

$$
= \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \left(\ln(\sigma^2) + \frac{1}{\sigma^2} Z\left(x_j^{(k,a)} \right)^2 \right) + \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\psi). \tag{A.1}
$$

We observe that, for any $k \in \{1, \ldots, K\}$ and $a \in \{0, \ldots, k-1\}$ the points $x_0^{(k,a)}, \ldots, x_{n_{k,a}}^{(k,a)}$, with $n_{k,a} = \lfloor (n-1-a-k)/k \rfloor = (n/k)(1 + o(1))$ satisfy the conditions of the setting of [57] (they have increasing two-by-two distinct values and are restricted to $[0, 1]$). Furthermore, the set $\{k \in \{1, \ldots, K\}, a \in \{0, \ldots, k-1\}\}$ is finite with cardinality not depending on n. Hence it can be seen that the proof of Theorem 1 in [57] (see in particular the offline equation at the top of Page 289 and Lemma 4) leads to

$$
\sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\psi) - \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\bar{\psi})
$$

$$
= \left[\sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} (n/k) \left\{ \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} - \frac{\theta_0 \sigma_0^2}{\bar{\theta} \bar{\sigma}^2} + \ln(\theta \sigma^2) - \ln(\bar{\theta} \bar{\sigma}^2) \right\} \right] + r_{1, \psi, \bar{\psi}},
$$

where $\bar{\psi} = (\bar{\theta}, \bar{\sigma}^2)$ and where $\sup_{\bar{\psi}, \psi \in J} |r_{1,\psi,\bar{\psi}}| = o(n)$ almost surely. Furthermore, since Z is almost surely continuous on $[0, 1]$, we have almost surely

$$\sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \left(\ln(\sigma^2) + \frac{1}{\sigma^2} Z\left(x_j^{(k,a)}\right)^2 \right) = n \left\{ \sum_{k=1}^{K} w_k \left(\ln(\sigma^2) + \frac{1}{\sigma^2} \frac{1}{n} \sum_{i=1}^{n} Z(s_i)^2 \right) \right\} + r_{2,\psi,\bar{\psi}},$$

where $\sup_{\bar{\psi}, \psi \in J} |r_{2,\psi,\bar{\psi}}| = o(n)$ almost surely. Hence, almost surely, as $n \to \infty$, we have

$$p\ell_n(\psi) - p\ell_n(\bar{\psi}) = n \left(\sum_{k=1}^{K} w_k \right) \left\{ \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} - \frac{\theta_0 \sigma_0^2}{\bar{\theta}\bar{\sigma}^2} + \ln(\theta \sigma^2) - \ln(\bar{\theta}\bar{\sigma}^2) \right.$$

$$+ \ln(\sigma^2) + \frac{1}{\sigma^2} \left(\frac{1}{n} \sum_{i=1}^{n} Z(s_i)^2 \right) - \ln(\bar{\sigma}^2) - \frac{1}{\bar{\sigma}^2} \left(\frac{1}{n} \sum_{i=1}^{n} Z(s_i)^2 \right) \right\} + r_{3,\psi,\bar{\psi}},$$

where $\sup_{\bar{\psi}, \psi \in J} |r_{3,\psi,\bar{\psi}}| = o(n)$ almost surely. Let $g(u) = (\theta_0 \sigma_0^2)/u + \ln(u)$, $S = (1/n) \sum_{i=1}^{n} Z(s_i)^2$ and $g_S(u) = \ln(u) + S/u$. We have

$$p\ell_n(\psi) - p\ell_n(\bar{\psi}) = n \left(\sum_{k=1}^{K} w_k \right) \left\{ g(\theta \sigma^2) - g(\bar{\theta}\bar{\sigma}^2) + g_S(\sigma^2) - g_S(\bar{\sigma}^2) \right\} + r_{3,\psi,\bar{\psi}}. \tag{A.2}$$

We have

$$0 < C_{\inf} \leq \inf_{u \in [ac, bd]} g(u) \leq \sup_{u \in [ac, bd]} g(u) \leq C_{\sup} < \infty.$$

Consider now the subcase of (i) where $ad > \theta_0 \sigma_0^2$. Let $\epsilon > 0$, and consider θ, σ^2 so that $|\theta \sigma^2 - \theta_0 \sigma_0^2| \leq \epsilon$. We have

$$\sigma^2 \leq \frac{\theta_0 \sigma_0^2}{\theta} + \frac{\epsilon}{\theta} \leq \frac{\theta_0 \sigma_0^2}{a} + \frac{\epsilon}{a}.$$

Hence, for $\epsilon > 0$ small enough there exists a fixed $d' < d$ so that, for $|\theta \sigma^2 - \theta_0 \sigma_0^2| \leq \epsilon$ we have $\sigma^2 \leq d'$. It follows that, with $\psi_1 = (b, d)$, from (A.2)

$$\frac{1}{n \sum_{k=1}^{K} w_k} \liminf_{n \to \infty} \left\{ \inf_{\psi \in J; |\theta \sigma^2 - \theta_0 \sigma_0^2| \leq \epsilon} p\ell_n(\psi) - p\ell_n(\psi_1) \right\} \geq C_{\inf} + \ln(c) + \frac{1}{d'} S - C_{\sup} - \ln(d) - \frac{1}{d} S$$

$$= \left(\frac{1}{d'} - \frac{1}{d} \right) S + C_{\inf} - C_{\sup} + \ln(c) - \ln(d).$$

From Lemma 6, we can show that, with probability $\Pr > 0$ (not depending on n)

$$\inf_{t \in [0,1]} Z^2(t) \geq 2 \frac{-C_{\inf} + C_{\sup} - \ln(c) + \ln(d)}{\frac{1}{d'} - \frac{1}{d}}.$$

Hence, with probability $\Pr > 0$ we have, for fixed $\epsilon > 0$

$$\frac{1}{n \sum_{k=1}^{K} w_k} \liminf_{n \to \infty} \left\{ \inf_{\psi \in J; |\theta \sigma^2 - \theta_0 \sigma_0^2| \leq \epsilon} p\ell_n(\psi) - p\ell_n(\psi_1) \right\} > 0. \tag{A.3}$$

This implies that $\hat{\theta}_{p\ell} \hat{\sigma}_{p\ell}^2$ does not converge in probability to $\theta_0 \sigma_0^2$ in the case where $ad > \theta_0 \sigma_0^2$.

Let us now consider the subcase of (i) where $bc < \theta_0 \sigma_0^2$. Let $\epsilon > 0$, and consider θ, σ^2 so that $|\theta \sigma^2 - \theta_0 \sigma_0^2| \leq \epsilon$. We have

$$\sigma^2 \geq \frac{\theta_0 \sigma_0^2}{\theta} - \frac{\epsilon}{\theta} \geq \frac{\theta_0 \sigma_0^2}{b} - \frac{\epsilon}{a}.$$

Hence, for $\epsilon > 0$ small enough there exists a fixed $c' > c$ so that, for $|\theta \sigma^2 - \theta_0 \sigma_0^2| \leq \epsilon$ we have $\sigma^2 \geq c'$. One can check that the function g_S in (A.2) has a second derivative which is negative for $t \geq 2S$. Hence, if $\epsilon \leq c'/4$ and $S \leq c'/4$ we have, since $\sigma^2 \geq c'$

$$g_S(\sigma^2) - g_S(\sigma^2 - \epsilon) \geq \epsilon \inf_{t \in [c'-\epsilon, d]} g_S'(t) \geq \epsilon g_S'(d) \geq \epsilon \left(\frac{1}{d} - \frac{S}{d^2} \right) \geq \frac{\epsilon}{2d}$$

if we further assume that $S \leq d/2$. Also, we have (still considering $|\theta \sigma^2 - \theta_0 \sigma_0^2| \leq \epsilon$)

$$\left| g(\theta \sigma^2) - g(\theta(\sigma^2 - \epsilon)) \right| \leq b \epsilon \sup_{t \in [\theta_0 \sigma_0^2 + \epsilon, \theta_0 \sigma_0^2 - \epsilon - b\epsilon]} |g'(t)| = b \epsilon \sup_{t \in [\theta_0 \sigma_0^2 + \epsilon, \theta_0 \sigma_0^2 - \epsilon - b\epsilon]} \left| \frac{1}{t} \left(1 - \frac{\sigma_0^2 \theta_0}{t} \right) \right| \leq A_{\sup} \epsilon^2$$

when $\epsilon \le \nu$, where $A_{\sup} < \infty$ and $\nu > 0$ can be chosen so as to depend only on θ_0, σ_0^2, b. Hence, we have, for $\epsilon \le c'/4$, $\epsilon \le \nu$, $S \le c'/4$ and $S \le d/2$,

$$g(\theta\sigma^2) - g(\theta\{\sigma^2 - \epsilon\}) + g_S(\sigma^2) - g_S(\sigma^2 - \epsilon) \ge \frac{\epsilon}{2d} - A_{\sup}\epsilon^2.$$

From Lemma 6, we can show that with probability $\Pr > 0$, not depending on n, we have $\sup_{t \in [0,1]} Z^2(t) \le \min(c'/4, d/2)$. Hence, we can choose $\epsilon > 0$, not depending on n so that, from (A.2), with probability $\Pr > 0$

$$\frac{1}{n\sum_{k=1}^{K} w_k} \liminf_{n \to \infty} \left[\inf_{\psi \in J; |\theta\sigma^2 - \theta_0\sigma_0^2| \le \epsilon} p\ell_n(\psi) - p\ell_n\{(\theta, \sigma^2 - \epsilon)\} \right] > 0. \tag{A.4}$$

This implies that $\hat{\theta}_{p\ell} \hat{\sigma}_{p\ell}^2$ does not converge in probability to $\theta_0\sigma_0^2$ in the case where $bc < \theta_0\sigma_0^2$. Hence, the proof of the case (i) is concluded. The displays (A.3) and (A.4) also imply the case (iii).

Let us now address the case (ii). Let $\tilde{\sigma}^2$ be defined as

$$\tilde{\sigma}^2 = \begin{cases} S & \text{if } S \in [c, d] \\ c & \text{if } S \le c \\ d & \text{if } S \ge d. \end{cases}$$

Let $\tilde{\theta} = (\theta_0\sigma_0^2)/\tilde{\sigma}^2$ and observe that by assumption

$$\tilde{\theta} \in \left[\frac{\theta_0\sigma_0^2}{d}, \frac{\theta_0\sigma_0^2}{c} \right] \subset \left[\frac{ad}{d}, \frac{bc}{c} \right] = [a, b].$$

Then, from (A.1), from the last display of the proof of Theorem 1 in [57] and from similar arguments as before, we have for $\epsilon > 0$, almost surely, with $\tilde{\psi} = (\tilde{\theta}, \tilde{\sigma}^2) \in J$ with $\tilde{\theta}\tilde{\sigma}^2 = \theta_0\sigma_0^2$,

$$\frac{1}{n\sum_{k=1}^{K} w_k} \liminf_{n \to \infty} \inf_{\substack{|\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \ge \epsilon \\ (\theta,\sigma^2) \in J}} \left\{ p\ell_n(\psi) - p\ell_n(\tilde{\psi}) \right\} \ge \inf_{\substack{|\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \ge \epsilon \\ (\theta,\sigma^2) \in J}} \left\{ g(\theta\sigma^2) - g(\theta_0\sigma_0^2) + g_S(\sigma^2) - g_S(\tilde{\sigma}^2) \right\}$$

$$\ge \inf_{\substack{|\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \ge \epsilon \\ (\theta,\sigma^2) \in J}} \left\{ g(\theta\sigma^2) - g(\theta_0\sigma_0^2) \right\} > 0.$$

This concludes the proof for the case (ii).

Consider now the case (iv) where $J = (0, \infty) \times [c, d]$. Let $\tilde{\sigma}$ and $\tilde{\theta}$ be defined as in the proof for the case (ii). From the offline equation after (2.23) in the proof of Theorem 1 in [57], we have, with the notation and arguments in (A.1),

$$\sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\psi) - \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\tilde{\psi}) \ge \frac{n}{\theta} \left(\frac{\theta_0\sigma_0^2}{2d} \right) + \frac{1}{\theta} r_{4,\psi,\tilde{\psi}}, \tag{A.5}$$

where $\sup_{\theta \le \rho, c \le \sigma^2 \le d} |r_{4,\psi,\tilde{\psi}}| = o(n)$ almost surely, where $\rho > 0$ is a constant depending only on $\theta_0, \sigma_0^2, c, d$. Also, from (2.26) in the proof of Theorem 1 in [57], we have

$$\sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\psi) - \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\tilde{\psi}) \ge \nu n + r_{5,\psi,\tilde{\psi}}, \tag{A.6}$$

where $\sup_{\theta \ge B, c \le \sigma^2 \le d} |r_{5,\psi,\tilde{\psi}}| = o(n)$ almost surely, where $\nu > 0$ and $B < \infty$ can be chosen as functions of $\theta_0, \sigma_0^2, c, d$ only. Finally, as shown for obtaining (A.2),

$$\sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\psi) - \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\tilde{\psi}) = n \left\{ g(\theta\sigma^2) - g(\theta_0\sigma_0^2) \right\} + r_{6,\psi,\tilde{\psi}}, \tag{A.7}$$

where $\sup_{\rho \le \theta \le B, c \le \sigma^2 \le d} |r_{6,\psi,\tilde{\psi}}| = o(n)$ almost surely. We have for $\epsilon > 0$ $\inf_{|\theta\sigma^2 - \theta_0\sigma_0^2| \ge \epsilon} g(\theta\sigma^2) - g(\theta_0\sigma_0^2) > 0$. Hence, there exists a constant $A_{\inf} > 0$, not depending on n, so that, from (A.1) and (A.5)–(A.7)

$$\inf_{\substack{\theta \in (0,\infty), \sigma^2 \in [c,d] \\ |\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \ge \epsilon}} p\ell_n(\psi) - p\ell_n(\tilde{\psi}) \ge n \left(\sum_{k=1}^{K} w_k \right) \left\{ A_{\inf} + g_S(\sigma^2) - g_S(\tilde{\sigma}^2) \right\} + r_{7,\psi,\tilde{\psi}}, \ge n \left(\sum_{k=1}^{K} w_k \right) A_{\inf} + r_{7,\psi,\tilde{\psi}},$$

where $\sup_{0 < \theta < \infty, c \le \sigma^2 \le d} |r_{7,\psi,\tilde{\psi}}| = o(n)$ almost surely. This concludes the proof for the case (iv). $\square$

Proof of Theorem 2. Let $n_{k,a} = \lfloor (n-1-a-k)/k \rfloor$, $R_\psi^{(k,a)} = [\sigma^2 e^{-\theta|x_i^{(k,a)} - x_j^{(k,a)}|}]_{0 \le i,j \le n_{k,a}+1}$ and $Z^{(k,a)} = (x_0^{(k,a)}, \ldots, x_{n_{k,a}+1}^{(k,a)})^\top$. We have for $\psi \in J$, from Lemma 1, by writing Gaussian likelihoods as products of conditional likelihoods, forward and

backward, and by using the Markovian property of Z (Lemma 1 in [57]),

$$
\begin{aligned}
pc\ell_n(\psi) &= \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \left\{ \ell_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\psi) + \ell_{x_j^{(k,a)} | x_{j+1}^{(k,a)}}(\psi) \right\} \\
&= \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \left(\ln(|R_\psi^{(k,a)}|) + (Z^{(k,a)})^\top R_\psi^{-1} Z^{(k,a)} - \ln(\sigma^2) - \frac{1}{\sigma^2} Z(x_0^{(k,a)})^2 \right. \\
&\qquad\qquad \left. + \ln(|R_\psi^{(k,a)}|) + (Z^{(k,a)})^\top R_\psi^{-1} Z^{(k,a)} - \ln(\sigma^2) - \frac{1}{\sigma^2} Z(x_{n_{k,a}+1}^{(k,a)})^2 \right) \\
&= \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \left[\left\{ 2 \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\psi) \right\} + \ln(\sigma^2) + \frac{1}{\sigma^2} Z(x_0^{(k,a)})^2 - \ln(\sigma^2) - \frac{1}{\sigma^2} Z(x_{n_{k,a}+1}^{(k,a)})^2 \right] \\
&= 2 \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \ell_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\psi) + \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \frac{1}{\sigma^2} \left(Z(s_{1+a})^2 - Z(s_{1+a+k(n_{k,a}+1)})^2 \right).
\end{aligned}
$$

Hence, we conclude the proof in the case $J = [a, b] \times [c, d]$ or $J = (0, \infty) \times [c, d]$ from (A.5)–(A.7). In the case $J = [a, b] \times (0, \infty)$, we can express $\hat{\sigma}^2(\theta) = \operatorname{argmin}_{\sigma^2 \in (0,\infty)} pc\ell_n(\theta, \sigma^2)$ explicitly, and conclude using similar techniques as in [57] and similar arguments as in (A.1). We skip the details. $\square$

Proof of Lemma 2. We have that for $1 \le i < j \le n$ and $1 \le k < \ell \le n$,

$$
\operatorname{cov}(W_{i,j}^2 - 1, W_{k,\ell}^2 - 1) = 2\operatorname{cov}(W_{i,j}, W_{k,\ell})^2.
$$

One can show, by computing explicitly the covariances $\operatorname{cov}(W_{i,j}, W_{k,\ell})$, by distinguishing the different cases in (9), and by considering Taylor expansions, that we have for $1 \le i < j \le \min(n, i + K)$ and $1 \le k < \ell \le \min(n, k + K)$,

$$
\operatorname{cov}(W_{i,j}, W_{k,\ell})^2 = b_{i,j,k,\ell} + r_{i,j,k,\ell}(s_{\max(i,j,k,\ell)} - s_{\min(i,j,k,\ell)}),
$$

where

$$
\sup_{\substack{n \in \mathbb{N} \\ i,j,k,\ell \in \{1,\dots,n\} \\ i+1 \le j \le i+K \\ k+1 \le \ell \le k+K \\ |k-i| \le K}} |r_{i,j,k,\ell}| = O(1) \quad \text{and} \quad r_{i,j,k,\ell} = 0 \ \text{ if } \ |k - i| > K.
$$

In the multiple sum

$$
\sum_{i=1}^{n-1} \sum_{j=i+1}^{\min(n,i+K)} \sum_{k=\max(1,i-K)}^{\min(n,i+K)} \sum_{\ell=k+1}^{\min(n,k+K)} (s_{\max(i,j,k,\ell)} - s_{\min(i,j,k,\ell)})
$$

one can see that for any $a \in \{2, \dots, n\}$, $\Delta_a = s_a - s_{a-1}$ appears less than $4K^4$ times. Hence, this sum is a $O\left(\sum_{a=2}^{n} \Delta_a \right)$. This concludes the proof. $\square$

Proof of Lemma 3. We have, for $n \ge K + 1$,

$$
\tau_n^2 = \left\{ \frac{2}{n} \sum_{k=1}^{K} \sum_{\ell=1}^{K} w_k w_\ell \left(\sum_{i=1}^{n-k} \sum_{j=1}^{n-\ell} b_{i,i+k,j,j+\ell} \right) \right\} + o(1).
$$

Let $k, \ell \in \{1, \dots, K\}$ be fixed. Without loss of generality, assume that $\ell \le k$. We have, for $i \in \{1, \dots, n-k\}$

$$
\sum_{j=1}^{n-\ell} b_{i,i+k,j,j+\ell} \ge \sum_{j=i}^{i+k-\ell} b_{i,i+k,j,j+\ell} = \sum_{j=i}^{i+k-\ell} \frac{s_{j+\ell} - s_j}{s_{i+k} - s_i} \ge \frac{s_{i+\ell} - s_i}{s_{i+k} - s_i} + \sum_{j=i+1}^{i+k-\ell} \frac{s_{j+\ell} - s_{j+\ell-1}}{s_{i+k} - s_i} = \frac{s_{i+k} - s_i}{s_{i+k} - s_i} = 1.
$$

Hence $\sum_{i=1}^{n-k} \sum_{j=1}^{n-\ell} b_{i,i+k,j,j+\ell} \ge n - k$. This concludes the proof. $\square$

Proof of Lemma 4. The lemma follows from Lemma 2, because, with the notation there, each $b_{i,j,k,\ell}$ is in $[0, 1]$ and, for any fixed i, there are less than $2K^3$ values of (j, k, ℓ) in (10) for which $b_{i,j,k,\ell}$ is non-zero. $\square$

Proof of Theorem 3. Because of the consistency results in Theorems 1 and 2, it is sufficient to consider the case where $J = [a, b] \times [c, d]$. Furthermore, we only write the proof of (11), since the proof of (12) is similar.

Let

$$\omega(\psi) = \frac{\partial}{\partial \theta} p\ell_n(\psi).$$

From (6), we can write

$$\omega(\psi) = \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-a-1-k}{k} \rfloor} \frac{\partial}{\partial \theta} \ell_{s_{jk+a+1}, s_{(j+1)k+a+1}}(\psi).$$

From (3.11) in [57] , we can write, with $\sup_{\psi \in [a,b] \times [c,d]} |r_{i,\psi}| = O_p(1)$ for $i \in \{1, 2, 3\}$,

$$
\begin{aligned}
\omega(\psi) &= \left\{ -\sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \frac{\sigma_0^2 \theta_0}{\sigma^2 \theta^2} \sum_{j=0}^{\lfloor \frac{n-a-1-k}{k} \rfloor} \left(W_{jk+a+1,(j+1)k+a+1}^2 + \left(\lfloor \frac{n-a-1-k}{k} \rfloor \right) \frac{1}{\theta} \right) \right\} + r_{1,\psi} \qquad (A.8) \\
&= \left\{ \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-a-1-k}{k} \rfloor} \left(\frac{-\sigma_0^2 \theta_0}{\sigma^2 \theta^2} W_{jk+a+1,(j+1)k+a+1}^2 + \frac{1}{\theta} \right) \right\} + r_{2,\psi} \\
&= \left\{ \sum_{i=1}^{n} \sum_{k=1}^{\min\{K, n-i\}} w_k \left(-\frac{\sigma_0^2 \theta_0}{\sigma^2 \theta^2} W_{i,i+k}^2 + \frac{1}{\theta} \right) \right\} + r_{3,\psi}.
\end{aligned}
$$

Now $\hat{\sigma}^2 \in [a, b]$ and by Theorem 1, $\hat{\sigma}_{p\ell}^2 \hat{\theta}_{p\ell} \xrightarrow{a.s} \theta_0 \sigma_0^2$ and in view of (A.8), we obtain

$$O_p(1) = \sum_{i=1}^{n} \sum_{k=1}^{\min\{K, n-i\}} w_k \left(-\sigma_0^2 \theta_0 W_{i,i+k}^2 + \hat{\sigma}_{p\ell}^2 \hat{\theta}_{p\ell} \right)$$

and

$$O_p(1) = \sum_{i=1}^{n} \sum_{k=1}^{\min\{K, n-i\}} w_k \left\{ -\sigma_0^2 \theta_0 \left(W_{i,i+k}^2 - 1 \right) \right\} + \sum_{i=1}^{n} \sum_{k=1}^{\min\{K, n-i\}} w_k (\hat{\sigma}_{p\ell}^2 \hat{\theta}_{p\ell} - \sigma_0^2 \theta_0).$$

Then

$$\sqrt{n} \left(\sum_{k=1}^{K} w_k \right) \{1 + o_p(1)\} \left(\hat{\sigma}_{p\ell}^2 \hat{\theta}_{p\ell} - \sigma_0^2 \theta_0 \right) = \left\{ \frac{\sigma_0^2 \theta_0}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{\min\{K, n-i\}} w_k \left(W_{i,i+k}^2 - 1 \right) \right\} + o_p(1).$$

For $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, \min(K, n-i)\}$, let $(W_{i,i+k}^2 - 1) = T_{(i-1)K+k}$. Then we can write

$$\sqrt{n} \left(\sum_{k=1}^{K} w_k \right) \left(\hat{\sigma}_{p\ell}^2 \hat{\theta}_{p\ell} - \sigma_0^2 \theta_0 \right) = \frac{\sigma_0^2 \theta_0}{\sqrt{n}} \left(\sum_{i=1}^{n} \sum_{k=1}^{\min\{K, n-i\}} w_k T_{(i-1)K+k} \right) + o_p(1). \qquad (A.9)$$

Since T_i is independent of T_j for $|i - j| \geq K(K+1)$ we can apply Theorem 2.1 in [34] for weakly dependent variables and we can establish a central limit theorem for $\{(\sqrt{n} \sum_{k=1}^{K} w_k)/(\tau_n \sigma_0^2 \theta_0)\} \left(\hat{\sigma}_{p\ell}^2 \hat{\theta}_{p\ell} - \sigma_0^2 \theta_0 \right)$.

Let us assume

$$\frac{\sqrt{n}(\sum_{k=1}^{K} w_k)}{\tau_n \sigma_0^2 \theta_0} \left(\hat{\sigma}_{p\ell}^2 \hat{\theta}_{p\ell} - \sigma_0^2 \theta_0 \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

By Lemmas 3 and 4, we can extract a subsequence $n' \to \infty$ so that $\tau_{K,n'}^2 \to \tau_K^2 \in (0, \infty)$ as $n' \to \infty$ and so that

$$\frac{\sqrt{n}(\sum_{k=1}^{K} w_k)}{\tau_{K,n'} \sigma_0^2 \theta_0} \left(\hat{\sigma}_{p\ell}^2 \hat{\theta}_{p\ell} - \sigma_0^2 \theta_0 \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \qquad (A.10)$$

The triangular array (T_i) satisfies the conditions of Theorem 2.1 in [34], thus we obtain

$$\frac{1}{\sqrt{n'}} \sum_{i=1}^{n'} \sum_{k=1}^{\min\{K, n'-i\}} w_k T_{(i-1)K+k} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_K^2).$$

Hence from (A.9) and Slutsky's lemma we are in contradiction with (A.10). This concludes the proof. $\square$

Proof of Lemma 5. From the proof of Lemma 3, it suffices to show that for any $2 \leq k \leq K$,

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-k} b_{i,i+k,j,j+k} > 1.$$

We have from (9), since $s_i = i/n$ for all $n \in \mathbb{N}$ and $i \in \{1, \ldots, n-k\}$

$$\sum_{j=1}^{n-k} b_{i,i+k,j,j+k} \geq 1 + b_{i,i+k,i+1,i+k+1} = 1 + \frac{(k-1)^2}{k^2}.$$

Hence

$$\frac{1}{n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-k} b_{i,i+k,j,j+k} \geq 1 + \frac{(k-1)^2}{k^2} + o(1). \quad \square$$

Appendix B. An inconsistency result for variance estimation by weighted pairwise log-likelihood, for general correlation functions

In this section, we extend the inconsistency results (i) and (iii) of Theorem 1 for the WPMLE, to more general covariance functions than the exponential ones as in (1). Furthermore, the input space of the Gaussian process under consideration is $[0, 1]^d$, where the dimension d is allowed to be larger than one. We consider the estimation of a variance parameter, where the correlation function is known.

We consider a stationary correlation function $C : \mathbb{R}^d \to \mathbb{R}$ on $\mathbb{R}^d$, with $d \in \mathbb{N}$. We make the following assumptions on C.

Assumption. The function C is continuous. Let $d_C(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{2C(\boldsymbol{0}) - 2C(\boldsymbol{x} - \boldsymbol{x}')}$. Let $N([0, 1]^d, d_C, \rho)$ be the minimum number of balls with radius ρ (w.r.t. the distance d_C), required to cover $[0, 1]^d$. We have

$$\int_0^\infty \sqrt{\ln\{N([0, 1]^d, d_C, \rho)\}} \, d\rho < \infty. \tag{B.1}$$

Furthermore, the Fourier transform $\widehat{C}$ of C satisfies

$$\exists \, P < \infty \quad \text{so that as} \quad \|\boldsymbol{w}\| \to \infty, \quad \widehat{C}(\boldsymbol{w}) \|\boldsymbol{w}\|^P \to \infty. \tag{B.2}$$

Finally, $C(\boldsymbol{0}) = 1$ and $C(\boldsymbol{x}) < 1$ for $\boldsymbol{x} \neq \boldsymbol{0}$.

Many stationary correlation functions satisfy the above assumption. The condition (B.1) holds when C is Hölder continuous and the condition (B.2) holds when C belongs to the Matérn class of covariance functions [45]. We refer to the discussion after Condition A.1 in [32] for more insight on the above assumption.

We now consider a zero-mean Gaussian process Y on $[0, 1]^d$ with covariance function $\sigma_0^2 C$ with an unknown fixed $\sigma_0^2 \in (0, \infty)$. We assume that C is known. Consider a triangular array of weights $\{w_{i,j}^{(n)}\}_{n \in \mathbb{N}, 1 \leq i < j \leq n}$ in $[0, \infty)$ and let $w_{i,j}^{(n)} = w_{i,j}$. Assume that for all $n \in \mathbb{N}$, not all $w_{i,j}$ are zero. Consider also a sequence $(\boldsymbol{x}_i)_{i \in \mathbb{N}}$ of observation points on $[0, 1]^d$. Then, the pairwise log likelihood is

$$\ell_{C, \boldsymbol{x}_i, \boldsymbol{x}_j}(\sigma^2) = 2 \ln(\sigma^2) + \ln(1 - C(\boldsymbol{x}_i - \boldsymbol{x}_j)^2) + \frac{1}{\sigma^2} Y(\boldsymbol{x}_i)^2 + \frac{\{Y(\boldsymbol{x}_j) - C(\boldsymbol{x}_i - \boldsymbol{x}_j)Y(\boldsymbol{x}_i)\}^2}{\sigma^2(1 - C(\boldsymbol{x}_i - \boldsymbol{x}_j)^2)}$$

and the weighted pairwise log likelihood criterion is

$$p\ell_{C,n}(\sigma^2) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \ell_{C, \boldsymbol{x}_i, \boldsymbol{x}_j}(\sigma^2).$$

We can then define the WPMLE $\hat{\sigma}_{C,p\ell}^2$ of σ_0^2 as the solution of

$$p\ell_{C,n}(\hat{\sigma}_{C,p\ell}^2) = \min_{\sigma^2 \in (0,\infty)} p\ell_{C,n}(\sigma^2).$$

The next theorem gives the general inconsistency result for the WPMLE. It is proved at the end of Appendix B.

Theorem 4. *Assume that the above assumption holds. Then $\hat{\sigma}_{C,p\ell}^2$ does not converge to σ_0^2 in probability as $n \to \infty$.*

Because of its generality, Theorem 4 reinforces the warning given in (i) and (iii) of Theorem 1 against the WPMLE under fixed-domain asymptotics. Indeed, the estimation of a single variance parameter σ_0^2, with a known correlation function C, is arguably one of the most favorable one for covariance parameter estimation. For instance it is well-known that the

MLE of σ_0^2 is $n^{1/2}$-consistent in the setting of Theorem 4 (provided the matrix $[C(\boldsymbol{x}_i - \boldsymbol{x}_j)]_{i,j \in \{1,\ldots,n\}}$ is invertible for all n). Thus, it would not be a surprise if the WPMLE was shown to be inconsistent under fixed-domain asymptotics, in more general cases than in Theorem 4, where several covariance parameters have to be estimated, for instance a variance and d spatial correlation lengths [37].

In the next theorem, we show that, for some specific structures of the weights and of the observation points, the WPMLE of σ_0^2 converges to a non-constant random variable (that is given explicitly in the proof of this theorem).

Theorem 5. *Assume that the above assumption holds. Let $g : [-1, 1]^d \to \mathbb{R}^+$ be a continuous non-zero function satisfying $g(\boldsymbol{x}) = g(-\boldsymbol{x})$ for $\boldsymbol{x} \in [0, 1]^d$. Assume that $w_{i,j} = g(\boldsymbol{x}_j - \boldsymbol{x}_i)$ for $n \in \mathbb{N}$, $i, j \in \{1, \ldots, n\}$. Assume also that there exists $\delta \in (0, 1]$ and $a \in (0, \infty)$ so that $1 - C(\boldsymbol{t}) = a\|\boldsymbol{t}\|^\delta + o(\|\boldsymbol{t}\|^\delta)$ as $\boldsymbol{t} \to \boldsymbol{0}$. Let the observation points $(\boldsymbol{x}_i)_{i \in \mathbb{N}}$ be random, i.i.d. with bounded probability distribution function f on $[0, 1]^d$, with $(\boldsymbol{x}_i)_{i \in \mathbb{N}}$ independent of Y. Then there exists a non-negative random variable $\hat{\sigma}_\infty^2$ with non-zero variance so that $\mathrm{E}(\hat{\sigma}_\infty^2) = \sigma_0^2$ and so that, almost-surely, $\hat{\sigma}_{C,p\ell}^2 \to_{n \to \infty} \hat{\sigma}_\infty^2$.*

We remark that the WPMLE of σ_0^2 is thus an example of an estimator that is consistent under increasing-domain asymptotics and that is inconsistent and converges to a non-degenerate distribution under fixed-domain asymptotics. This type of behavior is also analyzed in [60].

Proof of Theorem 4. We can show that

$$p\ell_{C,n} = \ln(\sigma^2) \left(2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \right) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \ln(1 - C(\boldsymbol{x}_j - \boldsymbol{x}_j)^2)$$

$$+ \frac{1}{\sigma^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \frac{Y(\boldsymbol{x}_i)^2 + Y(\boldsymbol{x}_j)^2 - 2C(\boldsymbol{x}_j - \boldsymbol{x}_j)Y(\boldsymbol{x}_i)Y(\boldsymbol{x}_j)}{1 - C(\boldsymbol{x}_j - \boldsymbol{x}_j)^2}.$$

Hence, it follows that

$$\hat{\sigma}_{C,p\ell}^2 = \frac{1}{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \frac{Y(\boldsymbol{x}_i)^2 + Y(\boldsymbol{x}_j)^2 - 2C(\boldsymbol{x}_j - \boldsymbol{x}_j)Y(\boldsymbol{x}_i)Y(\boldsymbol{x}_j)}{1 - C(\boldsymbol{x}_j - \boldsymbol{x}_j)^2} \tag{B.3}$$

$$= \frac{1}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \frac{\left\{ Y(\boldsymbol{x}_i) - Y(\boldsymbol{x}_j) \right\}^2}{2(1 - C(\boldsymbol{x}_i - \boldsymbol{x}_j)^2)}$$

$$+ \frac{1}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} \frac{2 - 2C(\boldsymbol{x}_j - \boldsymbol{x}_j)}{2(1 - C(\boldsymbol{x}_j - \boldsymbol{x}_j)^2)} Y(\boldsymbol{x}_i)Y(\boldsymbol{x}_j).$$

Let $r(\boldsymbol{t}) = 1 - C(\boldsymbol{t})$. Then $r(\boldsymbol{t}) = o(1)$ as $\boldsymbol{t} \to \boldsymbol{0}$. Hence

$$\Delta := \inf_{\boldsymbol{t} \in [-1,1]^d} \frac{2 - 2C(\boldsymbol{t})}{2(1 - C(\boldsymbol{t})^2)} = \inf_{\boldsymbol{t} \in [-1,1]^d} \frac{2r(\boldsymbol{t})}{2(2r(\boldsymbol{t}) - r(\boldsymbol{t})^2)} > 0, \tag{B.4}$$

where we have used the continuity of C and the fact that $C(\boldsymbol{t}) < 1$ for $\boldsymbol{t} \in [-1, 1]^d \setminus \{\boldsymbol{0}\}$. Hence we have, if $\inf_{\boldsymbol{t} \in [0,1]^d} Y(\boldsymbol{t}) \geq 0$,

$$\hat{\sigma}_{C,p\ell}^2 \geq \Delta \inf_{\boldsymbol{t} \in [0,1]^d} Y(\boldsymbol{t})^2. \tag{B.5}$$

From Lemma A.3 in [32], we have

$$\inf_{\boldsymbol{t} \in [0,1]} Y(\boldsymbol{t})^2 \geq \frac{2\sigma_0^2}{\Delta}$$

with probability $\Pr > 0$ where $\Pr$ does not depend on n. Hence, from (B.5), $\hat{\sigma}_{C,p\ell}^2$ does not converge to σ_0^2 in probability. $\quad\square$

Proof of Theorem 5. From (B.3), we have

$$\hat{\sigma}_{C,p\ell}^2 = \frac{1}{\sum_{i,j=1}^{n} g(\boldsymbol{x}_j - \boldsymbol{x}_i)} \sum_{i,j=1}^{n} g(\boldsymbol{x}_j - \boldsymbol{x}_i) \frac{\left\{ Y(\boldsymbol{x}_i) - Y(\boldsymbol{x}_j) \right\}^2}{2(1 - C(\boldsymbol{x}_i - \boldsymbol{x}_j)^2)}$$

$$+ \frac{1}{\sum_{i,j=1}^{n} g(\boldsymbol{x}_j - \boldsymbol{x}_i)} \sum_{i,j=1}^{n} g(\boldsymbol{x}_j - \boldsymbol{x}_i) \frac{2 - 2C(\boldsymbol{x}_i - \boldsymbol{x}_j)}{2(1 - C(\boldsymbol{x}_i - \boldsymbol{x}_j)^2)} Y(\boldsymbol{x}_i)Y(\boldsymbol{x}_j)$$

$$= \frac{1}{\frac{1}{n^2} \sum_{i,j=1}^{n} g(\boldsymbol{x}_j - \boldsymbol{x}_i)} \frac{1}{n^2} \sum_{i,j=1}^{n} F(\boldsymbol{x}_i, \boldsymbol{x}_j) + \frac{1}{\frac{1}{n^2} \sum_{i,j=1}^{n} g(\boldsymbol{x}_j - \boldsymbol{x}_i)} \frac{1}{n^2} \sum_{i,j=1}^{n} G(\boldsymbol{x}_i, \boldsymbol{x}_j),$$

by defining $F(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $G(\boldsymbol{x}_i, \boldsymbol{x}_j)$ appropriately. In the above display, F and G are random functions from $[0, 1]^d \times [0, 1]^d$ to $\mathbb{R}$. From (B.4), and since Y has continuous realizations almost surely [49] and g is continuous, G is a bounded continuous function almost surely. Also, F is continuous on $[-1, 1]^d \backslash \{\mathbf{0}\}$ almost surely. From [49], for $\alpha < \delta$, almost surely, there exists $A < \infty$ for which, for any $\boldsymbol{x}, \boldsymbol{y} \in [0, 1]^d$, $|Y(\boldsymbol{x}) - Y(\boldsymbol{y})| \leq A\|\boldsymbol{x} - \boldsymbol{y}\|^{\alpha/2}$. Hence, almost surely, $F(\boldsymbol{x}, \boldsymbol{y}) \leq B\|\boldsymbol{x} - \boldsymbol{y}\|^{\alpha-\delta}$ with $B < \infty$. The function $(\boldsymbol{x}, \boldsymbol{y}) \to \|\boldsymbol{x} - \boldsymbol{y}\|^{\alpha-\delta}$ has a finite integral on $[-1, 1]^d$ for α close enough to δ. Hence, almost surely,

$$\int_{[0,1]^d} \int_{[0,1]^d} F(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{x}) f(\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y} < \infty.$$

Hence, for almost all the realization functions of Y (with respect to the distribution of Y), we can use Theorem A in Section 5.4 of [40] to obtain that

$$\hat{\sigma}^2_{C,p\ell} \to_{n\to\infty} \frac{1}{\int_{[0,1]^d} \int_{[0,1]^d} g(\boldsymbol{y} - \boldsymbol{x}) f(\boldsymbol{x}) f(\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y}} \int_{[0,1]^d} \int_{[0,1]^d} g(\boldsymbol{y} - \boldsymbol{x}) \frac{\{Y(\boldsymbol{x}) - Y(\boldsymbol{y})\}^2}{2(1 - C(\boldsymbol{x} - \boldsymbol{y})^2)} f(\boldsymbol{x}) f(\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y}$$

$$+ \frac{1}{\int_{[0,1]^d} \int_{[0,1]^d} g(\boldsymbol{y} - \boldsymbol{x}) f(\boldsymbol{x}) f(\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y}} \int_{[0,1]^d} \int_{[0,1]^d} g(\boldsymbol{y} - \boldsymbol{x}) \frac{2 - 2C(\boldsymbol{x} - \boldsymbol{y})}{2(1 - C(\boldsymbol{x} - \boldsymbol{y})^2)} Y(\boldsymbol{x}) Y(\boldsymbol{y}) f(\boldsymbol{x}) f(\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y}.$$

The above limit is a random variable that we write $\hat{\sigma}^2_\infty$ and that is almost surely finite (the integrals are defined in the L^2 sense, see [45]). One can show that the mean value of $\hat{\sigma}^2_\infty$ is σ^2_0. Finally, the variance of $\hat{\sigma}^2_\infty$ is non-zero because of Theorem 4. $\square$

References

[1] M. Abt, W. Welch, Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian Stochastic processes, Canad. J. Statist. 26 (1998) 127–137.

[2] F. Bachoc, Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model mispecification, Comput. Statist. Data Anal. 66 (2013) 55–69.

[3] F. Bachoc, Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes, J. Multivariate Anal. 125 (2014) 1–35.

[4] F. Bachoc, Asymptotic analysis of covariance parameter estimation for Gaussian processes in the misspecified case, Bernoulli 24 (2) (2018) 1531–1575.

[5] F. Bachoc, A. Lagnoux, A.F. López-Lopera, Maximum likelihood estimation for Gaussian processes under inequality constraints, 2018, arXiv: 1804.03378.

[6] F. Bachoc, A. Lagnoux, T.M.N. Nguyen, Cross-validation estimation of covariance parameters under fixed-domain asymptotics, J. Multivariate Anal. 160 (2017) 42–67.

[7] J. Bect, F. Bachoc, D. Ginsbourger, A supermartingale approach to Gaussian process based sequential design of experiments, Bernoulli (2019) (forthcoming).

[8] M. Bevilacqua, T. Faouzi, R. Furrer, E. Porcu, Estimation and prediction using generalized Wendland functions under fixed domain asymptotics, Ann. Statist. 47 (2) (2019) 828–856.

[9] M. Bevilacqua, C. Gaetan, Comparing composite likelihood methods based on pairs for spatial Gaussian random fields, Stat. Comput. 25 (2015) 877–892.

[10] M. Bevilacqua, C. Gaetan, J. Mateu, E. Porcu, Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach, J. Amer. Statist. Assoc. 107 (2012) 268–280.

[11] R.H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. 16 (1995) 1190–1208.

[12] Y. Cao, D.J. Fleet, Generalized product of experts for automatic and principled fusion of Gaussian process predictions, in: Modern Nonparametrics 3: Automating the Learning Pipeline workshop at NIPS, Montreal, 2014, arXiv preprint arXiv:1410.7827.

[13] C.-H. Chang, H.-C. Huang, C.-K. Ing, Asymptotic theory of generalized information criterion for geostatistical regression model selection, Ann. Statist. 42 (6) (2014) 2441–2468.

[14] C.-H. Chang, H.-C. Huang, C.-K. Ing, Mixed domain asymptotics for a stochastic process model with time trend and measurement error, Bernoulli 23 (1) (2017) 159–190.

[15] H. Chen, D. Simpson, Z. Ying, Infill asymptotics for a stochastic process model with measurement error, Statist. Sinica 10 (2000) 141–156.

[16] N. Cressie, Statistics for Spatial Data, J. Wiley, 1993.

[17] A. Datta, S. Banerjee, A.O. Finley, A.E. Gelfand, Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets, J. Amer. Statist. Assoc. 111 (514) (2016) 800–812.

[18] M.P. Deisenroth, J.W. Ng, Distributed Gaussian processes, in: Proceedings of the 32nd International Conference on Machine Learning, Lille, France, JMLR: W&CP vol. 37, 2015.

[19] J. Eidsvik, B.A. Shaby, B.J. Reich, M. Wheeler, J. Niemi, Estimation and prediction in spatial models with block composite likelihoods, J. Comput. Graph. Statist. 23 (2) (2014) 295–315.

[20] X. Feng, J. Zhu, P. Lin, M. Steen-Adams, Composite likelihood estimation for models of spatial ordinal data and spatial proportional data with zero/one values, Environmetrics 25 (8) (2014) 571–583.

[21] R. Furrer, M.G. Genton, D. Nychka, Covariance tapering for interpolation of large spatial datasets, J. Comput. Graph. Statist. 15 (3) (2006) 502–523.

[22] Y. Guan, A composite likelihood approach in fitting spatial point process models, J. Amer. Statist. Assoc. 111 (2006) 1502–1512.

[23] P. Heagerty, S.R. Lele, A composite likelihood approach to binary spatial data, J. Amer. Statist. Assoc. 93 (1998) 1099–1111.

[24] J. Hensman, N. Fusi, Gaussian processes for big data, in: Uncertainty in Artificial Intelligence, 2013, pp. 282–290.

[25] G.E. Hinton, Training products of experts by minimizing contrastive divergence, Neural Comput. 14 (8) (2002) 1771–1800.

[26] I.A. Ibragimov, Y.A. Rozanov, Gaussian Random Processes, Springer-Verlag New York, 1978.

[27] H. Joe, Y. Lee, On weighting of bivariate margins in pairwise likelihood, J. Multivariate Anal. 100 (2009) 670–685.

[28] M. Katzfuss, J. Guinness, W. Gong, Vecchia approximations of Gaussian-process predictions, 2018, arXiv:1805.03309.

[29] C.G. Kaufman, M.J. Schervish, D.W. Nychka, Covariance tapering for likelihood-based estimation in large spatial data sets, J. Amer. Statist. Assoc. 103 (484) (2008) 1545–1555.
[30] C.G. Kaufman, B.A. Shaby, The role of the range parameter for estimation and prediction in geostatistics, Biometrika 100 (2013) 473–484.
[31] B.G. Lindsay, G.Y. Yi, J. Sun, Issues and strategies in the selection of composite likelihoods, Statist. Sinica 21 (2011) 71–105.
[32] A.F. Lopez-Lopera, F. Bachoc, N. Durrande, O. Roustant, Finite-dimensional Gaussian approximation with linear inequality constraints, SIAM/ASA J. Uncertain. Quantificat. 6 (3) (2018) 1224–1255.
[33] K.V. Mardia, R.J. Marshall, Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika 71 (1984) 135–146.
[34] M. Neumann, A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics, ESAIM Probab. Stat. 17 (2013) 120–134.
[35] L. Pronzato, W.G. Müller, Design of computer experiments: space filling and beyond, Stat. Comput. 22 (3) (2012) 681–701.
[36] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016, http://www.R-project.org.
[37] O. Roustant, D. Ginsbourger, Y. Deville, DiceKriging, DiceOptim: two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization, J. Stat. Softw. 51 (1) (2012) 1–55.
[38] H. Rue, L. Held, Gaussian Markov Random Fields, Theory and Applications, Chapman & Hall, 2005.
[39] D. Rullière, N. Durrande, F. Bachoc, C. Chevalier, Nested kriging predictions for datasets with a large number of observations, Stat. Comput. 28 (4) (2018) 849–867.
[40] R.J. Serfling, Approximation Theorems of Mathematical Statistics, vol. 162, John Wiley & Sons, 2009.
[41] B.A. Shaby, D. Ruppert, Tapered covariance: Bayesian estimation and asymptotics, J. Comput. Graph. Statist. 21 (2) (2012) 433–452.
[42] M. Stein, Asymptotically efficient prediction of a random field with a misspecified covariance function, Ann. Statist. 16 (1988) 55–63.
[43] M. Stein, Bounds on the efficiency of linear predictions using an incorrect covariance function, Ann. Statist. 18 (1990) 1116–1138.
[44] M. Stein, Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure, Ann. Statist. 18 (1990) 850–872.
[45] M.L. Stein, Interpolation of spatial data: some theory for kriging, in: Springer Series in Statistics, Springer-Verlag, New York, 1999.
[46] M.L. Stein, Limitations on low rank approximations for covariance matrices of spatial data, Spat. Stat. 8 (2014) 1–19.
[47] M. Stein, Z. Chi, L. Welty, Approximating likelihoods for large spatial data sets, J. Royal Stat. Soc. B 66 (2004) 275–296.
[48] B. van Stein, H. Wang, W. Kowalczyk, T. Bäck, M. Emmerich, Optimally weighted cluster kriging for big data regression, in: International Symposium on Intelligent Data Analysis, Springer, 2015, pp. 310–321.
[49] P. Strait, Sample function regularity for Gaussian processes with the parameter in a Hilbert space, Pacific J. Math. 19 (1) (1966) 159–173.
[50] V. Tresp, A Bayesian committee machine, Neural Comput. 12 (11) (2000) 2719–2741.
[51] A.W. van der Vaart, Maximum likelihood estimation under a spatial sampling scheme, Ann. Statist. 24 (5) (1996) 2049–2057.
[52] C. Varin, N. Reid, D. Firth, An overview of composite likelihood methods, Statist. Sinica 21 (2011) 5–42.
[53] C. Varin, P. Vidoni, Pairwise likelihood inference for ordinal categorical time series, Comput. Statist. Data Anal. 51 (2006) 2365–2373.
[54] E. Vazquez, J. Bect, Convergence properties of the expected improvement algorithm with fixed mean and covariance functions, J. Stat. Plan. Inference 140 (11) (2010) 3088–3095.
[55] C.K. Williams, C.E. Rasmussen, Gaussian Processes for Machine Learning, MIT Press, 2006.
[56] X. Xu, N. Reid, On the robustness of maximum composite likelihood estimate, J. Statist. Plann. Inference 141 (9) (2011) 3047–3054.
[57] Z. Ying, Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process, J. Multivariate Anal. 36 (1991) 280–296.
[58] Z. Ying, Maximum likelihood estimation of parameters under a spatial sampling scheme, Ann. Statist. 21 (1993) 1567–1590.
[59] H. Zhang, Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics, J. Amer. Statist. Assoc. 99 (2004) 250–261.
[60] H. Zhang, D.L. Zimmerman, Towards reconciling two asymptotic frameworks in spatial statistics, Biometrika 92 (2005) 921–936.