Composite likelihood estimation for a Gaussian process under fixed domain asymptotics

François Bachoc, Moreno Bevilacqua, Daira Velandia

COMPOSITE LIKELIHOOD ESTIMATION FOR A GAUSSIAN PROCESS UNDER FIXED DOMAIN ASYMPTOTICS

François Bachoc[1] Moreno Bevilacqua[2] and Daira Velandia[3]

[1] *Institut de Mathématiques de Toulouse, France.*
[2] *Universidad de Valparaso, Chile.*
[3] *Universidad de la Costa, Colombia.*

Abstract: We study composite likelihood estimation of the covariance parameters with data from 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, depending on the range of admissible parameter values in the likelihood optimization. On the contrary, the weighted pairwise conditional maximum likelihood estimator is always consistent. Both estimators are also asymptotically Gaussian when they are consistent, with asymptotic variance 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.

Key words and phrases: Gaussian processes; large data sets; exponential model; pairwise composite likelihood; fixed-domain asymptotics; microergodic parameters, consistency, asymptotic normality.

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 the estimation of 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) Stein (1999); Williams and Rasmussen (2006) of the covariance parameters of a Gaussian stochastic process observed in $\mathbb{R}^d$, $d \geq 1$, has been deeply studied in the last years in the two following asymptotic frameworks.

The fixed domain asymptotic framework, sometimes called infill asymptotics (Stein, 1999; Cressie, 1993), 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 and 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 Zhang and Zimmerman (2005).

Under increasing-domain asymptotics, generally speaking, for all (identifiable) 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 (asymptotic) Fisher information matrix (Mardia and Marshall (1984); Shaby and Ruppert (2012); Bachoc (2014)).

The situation is significantly different under fixed domain asymptotics. Indeed, two types of covariance parameters can be distinguished: microergodic and non-microergodic parameters Ibragimov and Rozanov (1978); Stein (1999). A covariance parameter is microergodic if, for two different values of it, the two corresponding Gaussian measures are orthogonal, see Ibragimov and Rozanov (1978); Stein (1999). It is non-microergodic if, even for two different values of it, the two corresponding Gaussian measures are equivalent. Non-microergodic parameters cannot be estimated consistently, but misspecifying them asymptotically results in the same statistical inference as specifying them correctly Stein (1988, 1990a,b); Zhang and Zimmerman (2005). In the case of isotropic Matérn covariance functions with $d \leq 3$, Zhang (2004) 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 Kaufman and Shaby (2013). Similar results for the special case of the exponential covariance function were obtained previously in Ying (1991).

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 Bachoc (2013, 2018)). 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 dataset of n observations, the computational burden is $O(n^3)$, making this method computationally impractical for large datasets. 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 Stein (2014) and the references therein for a review), sparse approximation Hensman and Fusi (2013), covariance tapering Furrer et al. (2006); Kaufman et al. (2008), Gaussian Markov Random Fields approximation Rue and Held (2005); Datta et al. (2016), submodel aggregation Hinton (2002); Tresp (2000); Cao and Fleet (2014); Deisenroth and Ng (2015); van Stein et al. (2015); Rullière et al. (2018) and composite likelihood (CL).

With CL we indicate a general class of objective functions based on the likelihood of marginal or conditional events Varin et al. (2011). This kind of estimation method has two important benefits: it is generally appealing when dealing with large data sets and/or it can be helpful when it is difficult to specify the full likelihood. In this paper we focus on a specific type of composite likelihood called pairwise likelihood. Examples of the use of pairwise likelihood in the Gaussian and non Gaussian case can be found in Heagerty and Lele (1998), Bevilacqua et al. (2012), Bevilacqua and Gaetan (2015), Guan (2006) and Feng et al. (2014) to name a few.

The increasing domain asymptotic properties of the weighted pairwise likelihood estimator in the Gaussian case can be found in Bevilacqua and Gaetan (2015). Under this setting for all (identifiable) covariance parameters, the pairwise likelihood estimator is consistent and asymptotically normal under some mild regularity conditions. The asymptotic covariance matrix is equal to the inverse of the (asymptotic) Godambe information matrix.

There are no results under fixed domain asymptotics for composite likelihood to the best of our knowledge. In this paper we study the asymptotic properties of the pairwise likelihood estimation method when estimating data from a one-dimensional Gaussian process with exponential covariance model. This covariance model is particularly amenable to theoretical analysis of the MLE, as its Markovian property yields an explicit (matrix-free) expression of the likelihood function Ying (1991). Thus, the MLE for this model has been studied in Ying (1991) Chen et al. (2000); Chang et al. (2017), and also in higher dimension in Ying (1993); van der Vaart (1996); Abt and Welch (1998).

Under the same setting as in Ying (1991), 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). We 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. Conversely, we 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. In contrasts, we prove that the WPCMLE is consistent regardless of the ranges of admissible values for the scale and variance. Both estimators are also asymptotically Gaussian in the cases where they are consistent. The asymptotic variance of these estimators is no smaller than that of the MLE, and 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, that 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 Ying (1991) for the full likelihood. The consistency results are then obtained similarly as in Ying (1991). 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 Ying (1991).

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 decrease the asymptotic variance and, as 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 since it has been shown, by means of simulation studies, 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 (Varin and Vidoni, 2006; Joe and Lee, 2009; Bevilacqua and Gaetan, 2015).

A simulation study compares the finite sample properties of the weighted pairwise likelihood estimators with their asymptotic distributions, using maximum likelihood as a benchmark.

The paper falls into the following parts. In Section 2 we introduce pairwise (conditional) likelihood estimation. In Section 3 we provide the consistency and inconsistency results. In Section 4 we provide the asymptotic normality results, together with the analysis of the asymptotic variance. In Section 5 we discuss the numerical results. Concluding remarks are given in Section 6. The proofs are postponed to the appendix.

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 $s,t \in \mathbb{R}$,

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

for fixed $\theta_0, \sigma_0^2 \in (0,\infty)$. For $\theta, \sigma^2 \in (0,\infty)$ we let $\psi = (\theta, \sigma^2)$, we call σ^2 the variance parameter and we call θ the scale (correlation decay) parameter. We let $\psi_0 = (\theta_0, \sigma_0^2)$.

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

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

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

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

We also let

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

be the conditional log-likelihood criterion of $Z(t)$ given $Z(s)$. The quantity $l_{t|s}(\psi)$ 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 $(w_{i,j}^{(n)})_{n\in\mathbb{N}, i,j=1,\dots,n, i\neq j}$ be a triangular array of weights in $[0,\infty)$. We let $w_{i,j}^{(n)} = w_{i,j}$ for simplicity. Then, we define the weighted pairwise log-likelihood function pl_n as

$$\mathrm{pl}_n(\psi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} l_{s_i,s_j}(\psi) \tag{2}$$

and the weighted pairwise conditional log-likelihood function pcl_n as

$$\text{pcl}_n(\psi) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} w_{i,j} l_{s_j|s_i}(\psi). \tag{3}$$

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

$$\widehat{\psi}_{\text{pl}} = \underset{\psi \in J}{\text{argmin}}\ \text{pl}_n(\psi), \qquad \widehat{\psi}_{\text{pcl}} = \underset{\psi \in J}{\text{argmin}}\ \text{pcl}_n(\psi). \tag{4}$$

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 $i,j = 1,\ldots,n, i \neq j$, $w_{i,j} = w_{|i-j|}$ 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.

We now give convenient expressions for the weighted pairwise log-likelihood and weighted pairwise conditional log-likelihood functions in the following lemma.

Lemma 1. *Let, for $k = 1,...,n-1$ and $a = 0,...,k-1$, $x_j^{(k,a)} = s_{1+a+jk}$, for $j \in \mathbb{N}$ so that $1+a+jk \leq n$. Then we have, 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 $b < a$,*

$$\text{pl}_n(\psi) = \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_j^{(k,a)}, x_{j+1}^{(k,a)}}(\psi) \tag{5}$$

and

$$\text{pcl}_n(\psi) = \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \left(l_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\psi) + l_{x_j^{(k,a)}|x_{j+1}^{(k,a)}}(\psi) \right). \tag{6}$$

The benefit of Lemma 1 is that the rightmost sum in (6) can be expressed as a (full) log likelihood for a subset (that depends on k and a in (6)) of $(s_1,\ldots,s_n)$ (see Lemma 1 in Ying (1991)). Similarly, the rightmost sum in (5) can be expressed conveniently as a function of this (full) log likelihood. We refer to the Proofs of Theorems 1, 2 and 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 Ying (1991); Zhang (2004). Hence, we study the consistency, inconsistency and asymptotic normality for the estimators $\hat{\theta}_{\mathrm{pl}}\hat{\sigma}^2_{\mathrm{pl}}$ and $\hat{\theta}_{\mathrm{pcl}}\hat{\sigma}^2_{\mathrm{pcl}}$ of $\sigma_0^2\theta_0$ given in (4).

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

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

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

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

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

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

Theorem 1 can be interpreted as follows: The pairwise likelihood criterion can be decomposed as the sum of two functions, that correspond to the two triple sums in (12) in the proof of Theorem 1. The second function (the second triple sum) is a function of θ and σ^2 which maximizer would yield a consistent estimator of $\theta_0\sigma_0^2$. The first function (the first triple sum) is a function of σ^2 and not of θ, and 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 $\hat{\theta}_{pl}$ is able to compensate, so to speak, these non-negligible fluctuations. More precisely, independently of the value of $\hat{\sigma}^2_{pl}$, in the cases (ii) and (iv), because of the good behavior of the second triple sum in (12), the value of $\hat{\theta}_{pl}$ will be so that $\hat{\theta}_{pl}\hat{\sigma}^2_{pl}$ 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.

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 Bevilacqua and Gaetan (2015)). It can indeed be checked that the first triple sum in (12) 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 Zhang and Zimmerman (2005)).

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 maximizes the pairwise likelihood (without restriction) can be computed explicitly. This case (iii) is a case of inconsistent estimation of the microergodic parameter.

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, $\hat{\theta}_{\text{pcl}}\hat{\sigma}^2_{\text{pcl}}$ converges to $\theta_0\sigma_0^2$ almost surely.*

4 Asymptotic normality

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

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

Let us also define

$$\tau_n^2 = \frac{1}{n}Var\left(\sum_{i=1}^{n-1}\sum_{k=1}^{\min\{K,n-i\}} w_k\left(W_{i,i+k}^2 - 1\right)\right). \tag{7}$$

In Lemma 2, we provide 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 $1 \leq k < l \leq n$, $b_{i,j,k,l}$ be defined by $b_{i,j,k,l} = b_{k,l,i,j}$ and, for $i \leq k$,*

$$b_{i,j,k,l} = \begin{cases} 0 & \text{if } j \leq k \\ \frac{(s_j - s_k)^2}{(s_j - s_i)(s_l - s_k)} & \text{if } k \leq j \leq l \\ \frac{s_l - s_k}{s_j - s_i} & \text{if } k \leq l \leq j. \end{cases} \tag{8}$$

We have as $n \to \infty$

$$\tau_n^2 = \left(\frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{\min(n,i+K)} \sum_{k=1}^{n-1} \sum_{l=k+1}^{\min(n,k+K)} w_{j-i} w_{l-k} b_{i,j,k,l} \right) + o(1). \tag{9}$$

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 we have $\liminf_{n\to\infty} \tau_n^2 \geq 2(\sum_{k=1}^K w_k)^2$.*

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

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, we have*

$$\frac{\sqrt{n}(\sum_{k=1}^K w_k)}{\tau_n \sigma_0^2 \theta_0} (\hat{\sigma}_{\text{pl}}^2 \hat{\theta}_{\text{pl}} - \sigma_0^2 \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{10}$$

Furthermore, under the same conditions as in Theorem 2, we have

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

Because of Lemmas 3 and 4, we see that the rate of convergence is $\sqrt{n}$ in Theorem 3. Furthermore, it is shown in Ying (1991) 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, we see that 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. We also observe that 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 Bachoc et al. (2017) when considering a cross validation estimator in the same context.

Remark 1: It can be shown, from Lemma 1 and from Lemma 1 in Ying (1991), 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 (7) that $\tau_n^2 = 2w_1^2 + o(1)$ so that, indeed, the WPMLE and WPCMLE have the same asymptotic variance as the MLE.

In the following lemma, we show 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 $i = 1, ..., n$ and if $2 \leq K < +\infty$, then* $\liminf_{n\to\infty} \tau_n^2 > 2(\sum_{i=1}^K w_k)^2$.

Remark 2: 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 Ying (1991)). 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 3: The log likelihood criterion can be evaluated with a $O(n)$ computational cost (see Lemma 1 in Ying (1991)). 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.

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 $k \geq 2$.

Optimization was carried out using the R (R Development Core Team, 2016) function *optim* with the method "L-BFGS-B" (Byrd et al., 1995) 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

$$\frac{\sqrt{n}}{C_{\mathrm{y}}}\left(\frac{\hat{\sigma}^2_{i,\mathrm{y}}\hat{\theta}_{i,\mathrm{y}}}{\sigma_0^2\theta_0} - 1\right), \quad i = 1, \ldots, 5000,$$

where y =MLE, WPMLE and WPCMLE with the associated theoretical values of the standard Gaussian distribution. Here $C^2_{MLE} = 2$ and $C^2_{WPMLE} = C^2_{WPCMLE} = \widehat{\tau}^2_n / (\sum_{k=1}^{K} w_k)^2$ where $\widehat{\tau}^2_n$ is computed using the approximation in (9) *i.e.*,

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

Under the optimal weights setting and following Remark 1 we have $C^2_{WPMLE} = C^2_{WPCMLE} \approx 2$ 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 slightly faster for MLE and that the asymptotic behavior of the WPMLE and WPCMLE is similar. Figure 1 depicts a graphical comparison when $n = 801$.

In a second simulation study we focus only on the asymptotic variance $n^{-1}(\sigma_0^2\theta_0 C_{\mathrm{y}})^2$ 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,

n	Method	5%	25%	50%	75%	95%	mean	var	Kur
51	WPMLE	-1.6186	-0.7006	0.0264	0.8950	2.3767	0.1703	1.5644	1.6348
	WPCMLE	-1.6186	-0.7006	0.0264	0.8950	2.3767	0.1703	1.5644	1.6348
	MLE	-1.5540	-0.7043	-0.0585	0.6733	1.8372	0.0197	1.0891	0.3874
101	WPMLE	-1.5899	-0.6864	0.0278	0.8043	2.0501	0.1075	1.2520	0.2967
	WPCMLE	-1.5899	-0.6864	0.0278	0.8043	2.0501	0.1075	1.2520	0.2967
	MLE	-1.5556	-0.7173	-0.0289	0.6759	1.8059	0.0215	1.0687	0.0327
201	WPMLE	-1.5733	-0.6702	-0.0029	0.7400	1.8353	0.0497	1.0862	0.0933
	WPCMLE	-1.5733	-0.6702	-0.0029	0.7400	1.8353	0.0497	1.0862	0.0933
	MLE	-1.5702	-0.7038	-0.0476	0.6478	1.7007	-0.0034	1.0062	0.0316
401	WPMLE	-1.6121	-0.6767	0.0171	0.7275	1.7956	0.0419	1.0635	-0.0321
	WPCMLE	-1.6121	-0.6767	0.0171	0.7275	1.7956	0.0419	1.0635	-0.0321
	MLE	-1.6194	-0.6909	-0.0254	0.6770	1.7012	0.0048	1.0288	-0.0607
801	WPMLE	-1.6275	-0.6728	0.0099	0.7229	1.7001	0.0303	1.0350	-0.0133
	WPCMLE	-1.6275	-0.6728	0.0099	0.7229	1.7001	0.0303	1.0350	-0.0133
	MLE	-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

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

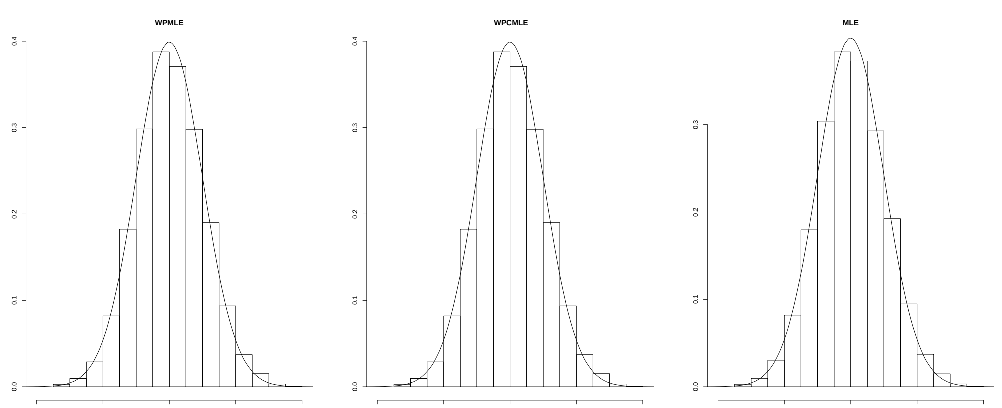

Figure 1: Comparison of the histogram of $\frac{\sqrt{n}}{C_{\mathrm{y}}}\left(\frac{\hat{\sigma}^2_{i,\mathrm{y}}\hat{\theta}_{i,\mathrm{y}}}{\sigma_0^2\theta_0}-1\right)$, $i = 1,\dots,5000$ for y = WPMLE, WPCMLE, MLE (from left to right) with the standard Gaussian distribution when $n = 801$.

		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
	K					
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

Table 2: Asymptotic variance versus sample variance of the WPMLE, WPCMLE and MLE estimation of the microergodic parameter when increasing n and K.

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 $\text{pl}_n(\theta, \sigma^2)$ and of $\text{pcl}_n(\theta, \sigma^2)$, over $\sigma^2 \in (0, +\infty)$, can be expressed explicitly. Hence, from a computational point of view, it is possible to compute explicitly $\min_{\sigma^2 \in (0,+\infty)} \text{pl}_n(\theta, \sigma^2)$ and $\min_{\sigma^2 \in (0,+\infty)} \text{pcl}_n(\theta, \sigma^2)$. These two latter functions can be minimized numerically over θ only, which is computationally beneficial. However, this case corresponds to the case (i) in Theorem 1 (since $c = 0$ and $d = +\infty$) so the WPMLE would be inconsistent in this setting.

6 Concluding remarks

In this paper, we provide, to the best of our knowledge, the first fixed-domain asymptotic properties of weighted pairwise likelihood estimators. We have considered the exponential covariance function in dimension one, which has enabled us to exploit the results of Ying (1991). By means of these results, and of specific new proof techniques, we have obtained the following conclusions: The weighted pairwise maximum likelihood estimator can be inconsistent, if the range of allowed values for the variance is too large. This is is contrast with the increasing-domain asymptotic situation, and allows us to issue a warning for practical use. On the other hand, the weighted pairwise conditional maximum likelihood estimator 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 maximum likelihood estimator, and that depends on the observation point locations and on the weight configurations.

We believe that the above conclusions are representative of the situation for more general families of covariance functions than the exponential one. On the other hand, in this paper, we have also obtained conclusions that are specific to the exponential family of covariance functions. In particular, the maximum likelihood estimator is asymptotically equal to a weighted pairwise conditional maximum likelihood estimator based only on pairs of consecutive observation points. Hence, in future work, we plan to study the fixed-domain asymptotic properties of weighted pairwise (conditional) maximum likelihood estimators, for more general families of covariance functions and $d \geq 1$. The present paper constitutes a first step in this 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 Zhang (2004) and Kaufman and Shaby (2013) for the Matérn model and in Bevilacqua et al. (2018) for the Generalized Wendland model.

Acknowledgments

The research work conducted by Moreno Bevilacqua was supported in part by FONDECYT grant 1160280, Chile and Iniciativa Cientifica Milenio - Minecon Nucleo Milenio MESCD.

A Proofs

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

$$\begin{aligned}
\mathrm{pl}_n(\psi) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{i,j} l_{s_i,s_j}(\psi) \\
&= \sum_{k=1}^{n-1} \sum_{i=1}^{n-1} w_k \mathbf{1}_{i+k\leq n} l_{s_i,s_{i+k}}(\psi) \\
&= \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{n} \mathbf{1}_{1+a+jk+k\leq n} l_{s_{1+a+jk},s_{1+a+jk+k}}(\psi) \\
&= \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_j^{(k,a)},x_{j+1}^{(k,a)}}(\psi).
\end{aligned}$$

The proof of (6) is the same, after observing that

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

□

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

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

Proof. The lemma is a special case of Lemma A.3 in Lopez-Lopera et al. (2017). □

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

$$\begin{aligned}
\mathrm{pl}_n(\psi) &= \sum_{k=1}^{n-1} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_j^{(k,a)},x_{j+1}^{(k,a)}}(\psi) \\
&= \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} \left[\log(\sigma^2) + \frac{1}{\sigma^2} Z\left(x_j^{(k,a)}\right)^2 \right] \\
&\quad + \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)}|x_j^{(k,a)}}(\psi). \qquad (12)
\end{aligned}$$

We observe that, for any $k = 1, ..., K$ and $a = 0, ..., k-1$ the points $x_0^{(k,a)}, ..., x_{n_{k,a}}^{(k,a)}$, with $n_{k,a} = \lfloor \frac{n-1-a-k}{k} \rfloor = (n/k)(1+o(1))$ satisfy the conditions of the setting of Ying (1991) (they have increasing two-by-two distinct values and are restricted to $[0,1]$). Furthermore, the set $\{k = 1, ..., K, a = 0, ..., k-1\}$ is finite with cardinality not depending on n. Hence it can be seen that the proof of Theorem 1 in Ying (1991) (see in particular the offline equation at the top of Page 289 and Lemma 4) leads to

$$\sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\psi) - \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\bar{\psi})$$
$$= \left[\sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} (n/k) \left(\frac{\theta_0 \sigma_0^2}{\theta \sigma^2} - \frac{\theta_0 \sigma_0^2}{\bar{\theta} \bar{\sigma}^2} + \log(\theta\sigma^2) - \log(\bar{\theta}\bar{\sigma}^2) \right) \right] + r_{1,\psi,\bar{\psi}},$$

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

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

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

$$\begin{aligned} \mathrm{pl}_n(\psi) - \mathrm{pl}_n(\bar{\psi}) = n \left(\sum_{k=1}^{K} w_k \right) \Bigg[& \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} - \frac{\theta_0 \sigma_0^2}{\bar{\theta} \bar{\sigma}^2} + \log(\theta\sigma^2) - \log(\bar{\theta}\bar{\sigma}^2) \\ & + \log(\sigma^2) + \frac{1}{\sigma^2} \left\{ \frac{1}{n} \sum_{i=1}^{n} Z(s_i)^2 \right\} - \log(\bar{\sigma}^2) \\ & - \frac{1}{\bar{\sigma}^2} \left\{ \frac{1}{n} \sum_{i=1}^{n} Z(s_i)^2 \right\} \Bigg] + r_{3,\psi,\bar{\psi}}, \end{aligned}$$

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

$$\mathrm{pl}_n(\psi) - \mathrm{pl}_n(\bar{\psi}) = n \left(\sum_{k=1}^{K} w_k \right) \left[g(\theta\sigma^2) - g(\bar{\theta}\bar{\sigma}^2) + g_S(\sigma^2) - g_S(\bar{\sigma}^2) \right] + r_{3,\psi,\bar{\psi}}. \quad (13)$$

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 (13)

$$\begin{aligned}
&\frac{1}{n\sum_{k=1}^K w_k} \liminf_{n\to\infty} \left(\inf_{\psi \in J; |\theta\sigma^2 - \theta_0\sigma_0^2| \leq \epsilon} \mathrm{pl}_n(\psi) - \mathrm{pl}_n(\psi_1) \right) \\
&\geq C_{inf} + \log(c) + \frac{1}{d'} S - C_{sup} - \log(d) - \frac{1}{d} S \\
&= \left(\frac{1}{d'} - \frac{1}{d} \right) S + C_{inf} - C_{sup} + \log(c) - \log(d).
\end{aligned}$$

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

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

Hence, with probability $\mathcal{P} > 0$ we have, for fixed $\epsilon > 0$

$$\frac{1}{n\sum_{k=1}^K w_k} \liminf_{n\to\infty} \left(\inf_{\psi \in J; |\theta\sigma^2 - \theta_0\sigma_0^2| \leq \epsilon} \mathrm{pl}_n(\psi) - \mathrm{pl}_n(\psi_1) \right) > 0. \tag{14}$$

This implies that $\hat{\theta}_{\mathrm{pl}}\hat{\sigma}_{\mathrm{pl}}^2$ 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 (13) 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 $\sigma^2 \geq c'$

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

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

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

when $\epsilon \leq \nu$, 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 $\epsilon \leq c'/4$, $\epsilon \leq \nu$, $S \leq c'/4$ and $S \leq d/2$,

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

From Lemma 6, we can show that with probability $\mathcal{P} > 0$, not depending on n, we have $\sup_{t \in [0,1]} Z^2(t) \leq \min(c'/4, d/2)$. Hence, we can choose $\epsilon > 0$, not depending on n so that, from (13), with probability $\mathcal{P} > 0$

$$\frac{1}{n \sum_{k=1}^K w_k} \liminf_{n \to \infty} \left(\inf_{\psi \in J; |\theta\sigma^2 - \theta_0\sigma_0^2| \leq \epsilon} \mathrm{pl}_n(\psi) - \mathrm{pl}_n((\theta, \sigma^2 - \epsilon)) \right) > 0. \tag{15}$$

This implies that $\hat{\theta}_{\mathrm{pl}}\hat{\sigma}^2_{\mathrm{pl}}$ 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 (14) and (15) also each imply the case (iii).

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

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

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 (12), from the last display of the proof of Theorem 1 in Ying (1991) and from similar arguments as before, we have for $\epsilon > 0$, almost surely, with $\tilde{\psi} = (\tilde{\theta}, \tilde{\sigma}^2) \in J$ with $\tilde{\theta}\tilde{\sigma}^2 = \theta_0 \sigma_0^2$,

$$
\begin{aligned}
&\frac{1}{n\sum_{k=1}^K w_k} \liminf_{n\to\infty} \inf_{\substack{|\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \geq \epsilon \\ (\theta,\sigma^2)\in J}} \left(\mathrm{pl}_n(\psi) - \mathrm{pl}_n(\tilde{\psi}) \right) \\
&\geq \inf_{\substack{|\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \geq \epsilon \\ (\theta,\sigma^2)\in J}} \left(g(\theta\sigma^2) - g(\theta_0\sigma_0^2) + g_S(\sigma^2) - g_S(\tilde{\sigma}^2) \right) \\
&\geq \inf_{\substack{|\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \geq \epsilon \\ (\theta,\sigma^2)\in J}} \left(g(\theta\sigma^2) - g(\theta_0\sigma_0^2) \right) \\
&> 0.
\end{aligned}
$$

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 Ying (1991), we have, with the notation and arguments in (12),

$$
\sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\psi) - \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\tilde{\psi}) \geq \frac{n}{\theta} \left(\frac{\theta_0 \sigma_0^2}{2d} \right) + \frac{1}{\theta} r_{4,\psi,\tilde{\psi}}, \tag{16}
$$

where $\sup_{\theta \leq \rho, c \leq \sigma^2 \leq d} |r_{4,\psi,\tilde{\psi}}| = o(n)$ 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 Ying (1991), we have

$$
\sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\psi) - \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\tilde{\psi}) \geq \nu n + r_{5,\psi,\tilde{\psi}}, \tag{17}
$$

where $\sup_{\theta \geq B, c \leq \sigma^2 \leq d} |r_{5,\psi,\tilde{\psi}}| = o(n)$ 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 (13),

$$
\begin{aligned}
\sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\psi) - \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-1-a-k}{k} \rfloor} l_{x_{j+1}^{(k,a)} | x_j^{(k,a)}}(\tilde{\psi}) \qquad (18) \\
= n \left(g(\theta\sigma^2) - g(\theta_0\sigma_0^2) \right) + r_{6,\psi,\tilde{\psi}} \quad ,
\end{aligned}
$$

where $\sup_{\rho\leq\theta\leq B, c\leq\sigma^2\leq d}|r_{6,\psi,\tilde{\psi}}| = o(n)$ almost surely. We have for $\epsilon > 0$ $\inf_{|\theta\sigma^2-\theta_0\sigma_0^2|\geq\epsilon} g(\theta\sigma^2) - g(\theta_0\sigma_0^2) > 0$. Hence, there exists a constant $A_{inf} > 0$, not depending on n, so that, from (16), (17), (18) and (12)

$$\begin{aligned}
\inf_{\substack{\theta\in(0,\infty),\sigma^2\in[c,d]\\|\theta\sigma^2-\tilde{\theta}\tilde{\sigma}^2|\geq\epsilon}} \mathrm{pl}_n(\psi) - \mathrm{pl}_n(\tilde{\psi}) \geq & n\left(\sum_{k=1}^K w_k\right)\left(A_{inf} + g_S(\sigma^2) - g_S(\tilde{\sigma}^2)\right) + r_{7,\psi,\tilde{\psi}}, \\
\geq & n\left(\sum_{k=1}^K w_k\right) A_{inf} + r_{7,\psi,\tilde{\psi}},
\end{aligned}$$

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

Proof of Theorem 2. Let $n_{k,a} = \lfloor\frac{n-1-a-k}{k}\rfloor$, $R_\psi^{(k,a)} = [\sigma^2 e^{-\theta|x_i^{(k,a)}-x_j^{(k,a)}|}]_{0\leq i,j\leq n_{k,a}+1}$ and $Z^{(k,a)} = (x_0^{(k,a)}, ..., x_{n_{k,a}+1}^{(k,a)})^\top$. 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 Ying (1991)),

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

Hence, we conclude the proof in the case $J = [a,b] \times [c,d]$ or $J = (0,\infty) \times [c,d]$ from (16), (17) and (18). In the case $J = [a,b] \times (0,\infty)$, we can express $\hat{\sigma}^2(\theta) = \mathrm{argmin}_{\sigma^2\in(0,\infty)}\mathrm{pcl}_n(\theta,\sigma^2)$ explicitly, and conclude using similar techniques as in Ying (1991) and similar arguments as in (12). We skip the details. $\square$

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

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

One can show, by computing explicitely the covariances $\mathrm{cov}(W_{i,j}, W_{k,l})$, by distinguishing the different cases in (8), and by considering Taylor expansions, that we have for $1 \leq i < j \leq \min(n, i+K)$ and $1 \leq k < l \leq \min(n, k+K)$,

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

where

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

In the multiple sum

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

one can see that for any $a = 2, ..., n$, $\Delta_a = s_a - s_{a-1}$ 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 $n \geq K+1$,

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

Let $k, l \in \{1, ..., K\}$ be fixed. Without loss of generality, assume that $l \leq k$.

We have, for $i = 1, ..., n-k$

$$\begin{aligned}\sum_{j=1}^{n-l} b_{i,i+k,j,j+l} \geq & \sum_{j=i}^{i+k-l} b_{i,i+k,j,j+l} \\ = & \sum_{j=i}^{i+k-l} \frac{s_{j+l} - s_j}{s_{i+k} - s_i} \\ \geq & \frac{s_{i+l} - s_i}{s_{i+k} - s_i} + \sum_{j=i+1}^{i+k-l} \frac{s_{j+l} - s_{j+l-1}}{s_{i+k} - s_i} \\ = & \frac{s_{i+k} - s_i}{s_{i+k} - s_i} \\ = & 1.\end{aligned}$$

Hence $\sum_{i=1}^{n-k} \sum_{j=1}^{n-l} b_{i,i+k,j,j+l} \geq n-k$. This concludes the proof. □

Proof of Lemma 4. The lemma follows from Lemma 2, because, with the notation there, each $b_{i,j,k,l}$ is in $[0,1]$ and, for any fixed i, there are less than $2K^3$ values of (j,k,l) in (9) for which $b_{i,j,k,l}$ is non-zero. □

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 (10), since the proof of (11) is similar.

Let

$$\omega(\psi) = \frac{\partial}{\partial \theta} \mathrm{pl}_n(\psi). \tag{19}$$

From (5), we can write

$$\omega(\psi) = \sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor \frac{n-a-1-k}{k} \rfloor} \frac{\partial}{\partial \theta} l_{s_{jk+a+1}, s_{(j+1)k+a+1}}(\psi). \tag{20}$$

From (3.11) in Ying (1991) , we can write, with $\sup_{\psi \in [a,b] \times [c,d]} |r_{i,\psi}| = O_p(1)$

for $i = 1, 2, 3$,

$$\begin{aligned}
\omega(\psi) &= -\sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \frac{\sigma_0^2\theta_0}{\sigma^2\theta^2} \sum_{j=0}^{\lfloor\frac{n-a-1-k}{k}\rfloor} \left[W^2_{jk+a+1,(j+1)k+a+1} + \left(\lfloor \frac{n-a-1-k}{k} \rfloor \right) \frac{1}{\theta} \right] \\
&\quad + r_{1,\psi} \\
&= \left[\sum_{k=1}^{K} w_k \sum_{a=0}^{k-1} \sum_{j=0}^{\lfloor\frac{n-a-1-k}{k}\rfloor} \left(\frac{-\sigma_0^2\theta_0}{\sigma^2\theta^2} W^2_{jk+a+1,(j+1)k+a+1} + \frac{1}{\theta} \right) \right] + r_{2,\psi} \\
&= \left[\sum_{i=1}^{n} \sum_{k=1}^{\min\{K,n-i\}} w_k \left(-\frac{\sigma_0^2\theta_0}{\sigma^2\theta^2} W^2_{i,i+k} + \frac{1}{\theta} \right) \right] + r_{3,\psi}. \qquad (21)
\end{aligned}$$

Now $\hat{\sigma}^2 \in [a, b]$ and by Theorem 1, $\hat{\sigma}^2_{\text{pl}}\hat{\theta}_{\text{pl}} \xrightarrow{a.s} \theta_0\sigma_0^2$ and in view of (21), we obtain

$$O_p(1) = \sum_{i=1}^{n} \sum_{k=1}^{\min\{K,n-i\}} w_k \left(-\sigma_0^2\theta_0 W^2_{i,i+k} + \hat{\sigma}^2_{\text{pl}}\hat{\theta}_{\text{pl}} \right) \qquad (22)$$

and

$$O_p(1) = \sum_{i=1}^{n} \sum_{k=1}^{\min\{K,n-i\}} w_k \left[-\sigma_0^2\theta_0 \left(W^2_{i,i+k} - 1 \right) \right] + \sum_{i=1}^{n} \sum_{k=1}^{\min\{K,n-i\}} w_k (\hat{\sigma}^2_{\text{pl}}\hat{\theta}_{\text{pl}} - \sigma_0^2\theta_0).$$

Then

$$\sqrt{n} \left(\sum_{k=1}^{K} w_k \right) (1 + o_p(1)) \left(\hat{\sigma}^2_{\text{pl}}\hat{\theta}_{\text{pl}} - \sigma_0^2\theta_0 \right)$$

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

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

$$\begin{aligned}
\sqrt{n} \left(\sum_{k=1}^{K} w_k \right) \left(\hat{\sigma}^2_{\text{pl}}\hat{\theta}_{\text{pl}} - \sigma_0^2\theta_0 \right) &= \frac{\sigma_0^2\theta_0}{\sqrt{n}} \left(\sum_{i=1}^{n} \sum_{k=1}^{\min\{K,n-i\}} w_k T_{(i-1)K+k} \right) \\
&\quad + o_p(1). \qquad (23)
\end{aligned}$$

Since T_i is independent of T_j for $|i-j| \geq K(K+1)$ we can apply Theorem 2.1 in Neumann (2013) for weakly dependent variables and we can establish a central limit theorem for $\frac{\sqrt{n}(\sum_{k=1}^K w_k)}{\tau_n \sigma_0^2 \theta_0} \left(\hat{\sigma}_{\text{pl}}^2 \hat{\theta}_{\text{pl}} - \sigma_0^2 \theta_0 \right)$.

Let us assume

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

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

$$\frac{\sqrt{n}(\sum_{k=1}^K w_k)}{\tau_{K,n'} \sigma_0^2 \theta_0} \left(\hat{\sigma}_{\text{pl}}^2 \hat{\theta}_{\text{pl}} - \sigma_0^2 \theta_0 \right) \overset{\mathcal{D}}{\nrightarrow} \mathcal{N}(0,1). \tag{24}$$

The triangular array (T_i) satisfies the conditions of Theorem 2.1 in Neumann (2013), thus we obtain

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

Hence from (23) and Slutsky's lemma we are in contradiction with (24). This concludes the proof. □

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 (8), since $s_i = i/n$ for all $n \in \mathbb{N}$ and $i = 1, ..., n-k$

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

Hence

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

□

References

Abt, M. and Welch, W. (1998). Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes. *The Canadian Journal of Statistics*, 26:127–137.

Bachoc, F. (2013). Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model mispecification. *Computational Statistics and Data Analysis*, 66:55–69.

Bachoc, F. (2014). Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis*, 125:1–35.

Bachoc, F. (2018). Asymptotic analysis of covariance parameter estimation for gaussian processes in the misspecified case. *Bernoulli*, 24(2):1531–1575.

Bachoc, F., Lagnoux, A., and Nguyen, T. M. N. (2017). Cross-validation estimation of covariance parameters under fixed-domain asymptotics. *Journal of Multivariate Analysis*, 160:42–67.

Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. (2018). Estimation and prediction using generalized wendland functions under fixed domain asymptotics. *The Annals of Statistics*. Forthcoming.

Bevilacqua, M. and Gaetan, C. (2015). Comparing composite likelihood methods based on pairs for spatial gaussian random fields. *Statistics and Computing*, 25:877–892.

Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012). Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *Journal of the American Statistical Association*, 107:268–280.

Byrd, R. H., Lu, P., Nocedal, J., , and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16:11901208.

Cao, Y. and Fleet, D. J. (2014). 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*. arXiv preprint arXiv:1410.7827.

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

Chen, H., Simpson, D., and Ying, Z. (2000). Infill asymptotics for a stochastic process model with measurement error. *Statistica Sinica*, 10:141–156.

Cressie, N. (1993). *Statistics for spatial data*. J. Wiley.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

Deisenroth, M. P. and Ng, J. W. (2015). Distributed Gaussian processes. *Proceedings of the 32nd International Conference on Machine Learning, Lille, France. JMLR: W&CP volume 37*.

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

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 111:1502–1512.

Heagerty, P. and Lele, S. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, 93:1099 – 1111.

Hensman, J. and Fusi, N. (2013). Gaussian processes for big data. *Uncertainty in Artificial Intelligence*, pages 282–290.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800.

Ibragimov, I. A. and Rozanov, Y. A. (1978). *Gaussian Random Processes*. Springer-Verlag New York.

Joe, H. and Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, 100:670–685.

Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555.

Kaufman, C. G. and Shaby, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, 100:473–484.

Lopez-Lopera, A. F., Bachoc, F. ., Durrande, N., and Roustant, O. (2017). Finite-dimensional gaussian approximation with linear inequality constraints. *arxiv.org/abs/1710.07453*.

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–146.

Neumann, M. (2013). A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, 17:120–134.

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields, Theory and applications*. Chapman & Hall.

Rullière, D., Durrande, N., Bachoc, F., and Chevalier, C. (2018). Nested kriging predictions for datasets with a large number of observations. *Statistics and Computing*, 28(4):849–867.

Shaby, B. A. and Ruppert, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, 21(2):433–452.

Stein, M. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics*, 16:55–63.

Stein, M. (1990a). Bounds on the efficiency of linear predictions using an incorrect covariance function. *The Annals of Statistics*, 18:1116–1138.

Stein, M. (1990b). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *The Annals of Statistics*, 18:850–872.

Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer Series in Statistics. Springer-Verlag New York.

Stein, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics*, 8:1–19.

Tresp, V. (2000). A bayesian committee machine. *Neural Computation*, 12(11):2719–2741.

van der Vaart, A. W. (1996). Maximum likelihood estimation under a spatial sampling scheme. *The Annals of Statistics*, 24(5):2049–2057.

van Stein, B., Wang, H., Kowalczyk, W., Bäck, T., and Emmerich, M. (2015). Optimally weighted cluster kriging for big data regression. In *International Symposium on Intelligent Data Analysis*, pages 310–321. Springer.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42.

Varin, C. and Vidoni, P. (2006). Pairwise likelihood inference for ordinal categorical time series. *Computational Statistics and Data Analysis*, 51:2365–2373.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate Analysis*, 36:280–296.

Ying, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *The Annals of Statistics*, 21:1567–1590.

Zhang, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261.

Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92:921–936.