A fast, near efficient, randomized-trace based method for fitting stationary Gaussian spatial models to large noisy data sets in the case of a single range-parameter

D.A. Girard

CNRS and University J. Fourier, Grenoble

Abstract

We consider the inference problem of fitting to noisy (gridded) observations, an isotropic zero-mean stationary Gaussian field model which belongs to the Matérn family with known regularity index $\nu \geq 0$, or to the spherical family. For estimating the correlation range and the variance of the field, two simple estimating functions based on the so-called "conditional Gibbs energy mean" (CGEM) and the empirical variance (EV) were recently introduced. This article presents a rather extensive Monte Carlo simulation study for problems with around a thousand observations and settings including large, moderate, and even "small", correlation ranges. It empirically demonstrates that the statistical efficiency of CGEM-EV is quite satisfying provided the signal-to-noise ratio is strong enough or ν is not too large.

1. Introduction

We consider the following classical inference problem: let $Z(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^2$, be a zero mean stationary Gaussian stochastic process whose autocorrelation function is assumed to belong to the isotropic Matérn family. One realization of this process is observed at $n = n_1 \times n_2$ regularly spaced (with step-size δ_1 in abscissa, δ_2 in ordinate) sites $\mathbf{s}_i, i = 1, \cdots, n$, of $[0, (n_1 - 1)\delta_1] \times [0, (n_2 - 1)\delta_2]$, with an additive white noise whose variance is σ_N^2 (this noise can model either suspected homoscedastic measurement errors or an additional nugget effect in Z, see e.g. Zhang and Zimmerman 2007 and references therein). In this article, we restrict ourselves to the case where σ_N is known. Using a standard lexicographic ordering, the observations thus form a vector $\mathbf{y}$ of size n whose law is Gaussian :

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \tau_0^2 R_{\theta_0} + \sigma_\mathrm{N}^2 I_n) \text{ where } \tau_0^2 = \mathrm{E}((Z(\mathbf{s}))^2) \text{ is the process variance,} \tag{1.1}$$

with I_n denoting the identity matrix and R_θ the autocorrelation matrix of the gridded process i.e. the block Toeplitz matrix (with n_1 Toeplitz square blocks, each of dimension n_2) whose coefficients are given by

$$[R_\theta]_{j,k} = \rho_{\nu,\theta}(||\mathbf{s}_j - \mathbf{s}_k||),\ j,k = 1,\cdots,n,$$

with

$$\rho_{\nu,\theta}(x) = \frac{(\theta x)^\nu}{\Gamma(\nu)2^{\nu-1}} K_\nu(\theta x),$$

where K_ν is the modified Bessel function of order $\nu > 0$. When mentioned, we will also consider another well known autocorrelation function, namely the spherical model $\rho_\theta^{\mathrm{S}}(x)$. See e.g. Zhang and Zimmerman 2007 for these definitions.

In this paper, ν, which is often called the regularity (or differentiability) index, is assumed to be known. Recall that $\rho_{1/2,\theta}(x) = \exp(-\theta x)$ is the exponential model, and that very simple expressions also exist for $\nu = 3/2$ and $5/2$: these ν's correspond to models commonly used (see Stein 1999, Rasmussen and Williams 2006). The parameter θ^{-1} is often called the "decorrelation length" or "the range parameter" .

Estimation of the variance and range parameters in autocovariance models is needed for various tasks, for example for establishing confidence band for the autovariance function, for constructing statistically efficient prediction of the process at unobserved location, or for optimally de-noising the observations.

It is often of great interest to be able to "effectively reduce" the number of parameters, especially when computing the likelihood function is costly. Zhang and Zimmerman (2007) recently proposed to use the classical weigthed least square method (not statistically full-efficient but much less costly than maximum likelihood (ML)) to estimate the range parameters (the θ here), next, to plug-in these estimates in the likelihood which is then maximized only with respect to τ^2 and, possibly, with respect to σ_{N}^2 (the solution, say $\hat{\tau}_{\mathrm{ML}}^2(\theta)$, being typically obtained iteratively, e.g. by Fisher scoring, even if σ_{N}^2 (> 0) is known). The idea underlying this method is that, at least for the Matérn family and an "infill asymptotics" point of view, even if θ is fixed at a wrong value θ_1, the product $\hat{\tau}_{\mathrm{ML}}^2(\theta_1)\theta_1^{2\nu}$ still remains an efficient estimator of $\tau_0^2\theta_0^{2\nu}$ which is the so-called microergodic parameter (see Du et al. (2009) for recent results of this type in the case without additive white noise).

The method that we proposed in Girard (2009), firstly reverses the roles of variance and range in the idea of Zhang and Zimmerman (2007): it is based on a very simple estimate for the signal-variance τ_0^2, namely the bias corrected empirical variance $\hat{\tau}_{\mathrm{EV}}^2$, and its associated signal-to-noise (SNR) estimate:

$$\hat{\tau}_{\mathrm{EV}}^2 = n^{-1}\mathbf{y}^T\mathbf{y} - \sigma_{\mathrm{N}}^2 \quad \text{and } \hat{b}_{\mathrm{EV}} := \frac{\hat{\tau}_{\mathrm{EV}}^2}{\sigma_{\mathrm{N}}^2}. \tag{1.2}$$

Secondly the maximization of the likelihood w.r.t. θ is replaced by the following simple estimating equation in θ : solve, with b fixed at $\hat{b}_{\mathrm{EV}}$

$$\mathbf{y}^T A_{b,\theta}\left(I - A_{b,\theta}\right)\mathbf{y} = \sigma_{\mathrm{N}}^2 \mathrm{tr} A_{b,\theta} \quad \text{where} \quad A_{b,\theta} = bR_\theta\left(I + bR_\theta\right)^{-1}. \tag{1.3}$$

This equation in θ is called the "conditional Gibbs energy mean and empirical variance" based estimating equation (CGEM-EV equation) in Girard (2009) which gives details, heuristic justifications and a theoretical result for the one dimensional "time series" analog setting, stating that an asymptotic full-efficiency is reached as ν decreases to 1/2.

We will also consider a variant of the previous uniform grid of $[0,(n_1-1)\delta_1]\times[0,(n_2-1)\delta_2]$ for the locations of the n observations. In this case, a weighted version of the simple average $\mathbf{y}^T\mathbf{y}/n$ in (1.2), which is motivated by a Riemann-sum type discretization of $\int_\Omega Z(\mathbf{s})^2 d\mathbf{s}/\int_\Omega d\mathbf{s}$ where Ω is the domain of the locations (suggested in Girard (2009)) is also considered in place of $\hat{\tau}^2_{\mathrm{EV}}$.

We first give some comments (Section 2) which complement those in Girard (2009). This article presents (in Section 3) a rather extensive Monte Carlo simulation study for problems with around a thousand observations and settings including large, moderate and even "small" correlation ranges. It empirically demonstrates that the statistical efficiency of CGEM-EV, even when using a fast randomized-trace approximation to $\mathrm{tr}A_{b,\theta}$, is quite satisfying provided the signal-to-noise ratio b_0 is strong enough or ν is not too large. If the observation grid is nonuniform, this efficiency may be degraded; however as it could have been expected from a "minimum variance property" (see last comment of Section 2) the Riemann-sum version of $\hat{\tau}^2_{\mathrm{EV}}$ restores this efficiency.

2. Some comments on the CGEM-EV estimating equation

2.1. The motivation behind this work is that it should be possible, and useful, to implement CGEM-EV for very large scale problems for which the exact ML method has a prohibitive computational cost. Indeed from works in the two previous decades, it is now known that a linear system with block Toeplitz - Toeplitz blocks matrix can often be solved with about $n\log(n)$ operations and a memory size of order n, by preconditioned conjugate gradient (PCG) approaches. The number of operations is actually a multiple of $n\log(n)$ which depends on the preconditioning method one employs for each particular application (see Chen, Hurvich and Lu, 2006, for certain time series problems). Once a fast solver is available to compute $A_{b,\theta}\mathbf{y}$, but does not form explicitly the matrix $A_{b,\theta}$, one is tempted to try a randomized-trace approximation (i.e. generate independent $\mathbf{w}_r \sim N(\mathbf{0}, I_n), r = 1, \cdots, n_{\mathrm{R}}$, and use $(1/n_{\mathrm{R}})\sum_{r=1}^{n_{\mathrm{R}}} \mathbf{w}_r^T A_{b,\theta}\mathbf{w}_r/(\mathbf{w}_r^T\mathbf{w}_r)$ in place of the exact $(1/n)\mathrm{tr}A_{b,\theta}$) when solving (1.3) by e.g. a bisection search.

In this article we do not attempt to analyze possible PCG solvers for problems like those of the following simulation. Indeed in the following, where exact ML estimates are also simulated, we consider settings with relatively small data size (less than 900) so that using such iterative solvers was not mandatory. Classical "exact" Cholesky or eigenvalues-eigenvectors decompositions were then used in the following, even to implement randomized-trace versions of the CGEM-EV estimating equation. Note that, in following simulation, we chose to compute the randomized-trace version of the CGEM-EV criterion at different tried values of θ by keeping constants the n_{R} simulated $\mathbf{w}_r$'s during the processing of each simulated data set.

2.2. A first comment is in order about the bias corrected empirical variance $\hat{b}_{\rm EV}$ defined in (1.2) as an estimate of the signal-to-noise ratio. Of course it may happen, especially in case of large correlation range, that the observed $\hat{b}_{\rm EV}$ has a negative value. It is easy to see that the probability of a such pathological event decreases when n increases; it is perhaps more important to observe that, even for “moderate” n, this probability becomes very small as soon as the true b_0 is large enough. Nevertheless, this entails that, in fact, CGEM-EV is not expected to be suitable for contexts with very weak signal-to-noise ratios.

2.3. Although the available theoretical results are restricted to regularly spaced locations, we are tempted to try to extend CGEM-EV to non regular cases. For example, extensions to cases where the n locations form a subset of the nodes of a regular grid (i.e. there are missing data on the grid) should be quite useful since the PCG approach often remains appropriate to efficiently compute the product of any vector of size n by the new $A_{b,\theta}$ matrix (see Fritz et al. 2009 for recent works on this subject).

2.4. With an irregular design for the sites, one may wonder whether weights which would take into account the possible non-uniformity of the design, cannot improve over the equal weights of $\mathbf{y}^T\mathbf{y}$.

To fix ideas, and for future reference, let us now describe the particular design which has been chosen for the simulation experiment analyzed in the following Section 3.7. Let $\mathcal{S}_1$ be the union of two juxtaposed uniform one-dimensional grids:

$$\mathcal{S}_1 = \{x_1, \cdots, x_{n_1}\} = \{0.02, 0.03, \cdots, 0.20\} \cup \{0.25, 0.30, \cdots, 0.95\},$$

thus $n_1 = 27$ and put

$$\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_1. \tag{2.1}$$

The locations of the observations are thus assumed to be the points of the Cartesian product $\mathcal{S}_1 \times \mathcal{S}_1$. With the choice of (2.1), the random field Z is much more densely observed in the subregion $[0, 0.20] \times [0, 0.20]$. This setting resembles the one already used by Zhang (2004, Fig.1), except that our grid is Cartesian. The reason of our choice is simply that there then exists a very simple (and commonly used) Riemann-sum approximation, precisely, with the so-called mid-points defined by $x_{i+1/2} = (x_{i+1} + x_i)/2, i = 1, \cdots, n_1 - 1$ and $x_{1/2} = 0$, $x_{n_1+1/2} = 1$:

$$\int_{[0,1]^2} Z^2(\mathbf{s})d\mathbf{s} \approx \sum_{i=1}^{n_1}\sum_{j=1}^{n_1} w_{i,j} Z^2(x_i, x_j) \text{ with } w_{i,j} = (x_{i+1/2} - x_{i-1/2})(x_{j+1/2} - x_{j-1/2}). \tag{2.2}$$

For such a Cartesian setting, denoting $y_{i,j}$ the observation of location (x_i, x_j), the weigthed version of $\hat{\tau}^2_{\rm EV}$ will be defined (noting that $\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} w_{i,j} = 1$ here) by

$$\hat{\tau}^2_{\rm wEV} := \sum_{i=1}^{n_1}\sum_{j=1}^{n_1} w_{i,j}(y^2_{i,j} - \sigma^2_{\rm N}). \tag{2.3}$$

The associated version of CGEM-EV (i.e. solve (1.3) with b fixed at $\hat{b}_{\text{wEV}} := \frac{\hat{\tau}^2_{\text{wEV}}}{\sigma^2_{\text{N}}}$) will be denoted by CGEM-wEV.

Let us now attempt to justify the choice of such weights in the case (2.1) and to suggest extensions to more general designs. Firstly, because we are dealing with (highly) correlated data, it is intuitive (at least in the case of no additive white noise) that the ordinary spatial average $\mathbf{y}^T\mathbf{y}/n$ should be modified so that, for example, two observations which are at very close sites (hereby two likely very close observations) be replaced by a single observation: the variance of the spatial average should then decrease. More generally, we should reduce the weights assigned to observations whose associated locations are in a "cluster". To give a more "quantitative" insight, we come back to the equispaced case and we recall a rather remarkable property of the equal weights in this case. Preliminarily, notice that the problem of estimating τ_0^2 can be formulated as the estimation of the mean of the stationary process Z^2, so we can refer to the related literature. One of the results of Shin and Song (2000) (which generalizes a well known analogous result in one-dimension) says that, for the estimation of $\tau_0^2 + \sigma^2_{\text{N}}$ in the model (1.1), under some "integrability" and "invertibility" conditions on the correlation, the uniform weighting (yielding the ordinary least square estimate or OLSE, linear in the squared observations) is asymptotically as efficient as the optimal weighting (yielding the best linear unbiased estimate or BLUE) which would require the knowledge of θ_0 (there, the asymptotic frame is an "increasing domain" regime where $(n_1, n_2) \to (\infty, \infty)$ for fixed (δ_1, δ_2)). In one-dimension, it is well known that, even with a "long memory" correlation, the OLSE of the mean often has good properties compared to the BLUE; see, for example, Yajima (1991) and references therein.

Now we return to a general design, except that we restrict ourselves to the one-dimensional case and we assume $\sigma_{\text{N}} = 0$; a Matérn process is observed at $0 = t_{1,n} < \cdots < t_{n,n} = T_n$. It is clear from the above, that a desirable property of the weights would be that the weighted sum perform nearly as well as (or possibly better than) the time-average $T_n^{-1} \int_0^{T_n} Z(t)^2 dt$ which is the continuous-time version of "the asymptotically efficient OLSE" mentioned above. Of course a natural approach to do this is to introduce a cubature rule (like a simple Rieman sum) since an extensive numerical analysis literature can furnish tools to control the (realized) integration error as

$$\delta_n := \max_{i=1,\cdots,n-1} |t_{i+1,n} - t_{i,n}| \to 0$$

in the case of a bounded $[0, T_n]$. However it is insightful here to consider the case $T_n \to \infty$ since the OLS estimate is then a consistent estimate. Now an interesting propery of the Rieman sum which corresponds to the integration of the "broken line" interpolation, is the following proposition which is an easy consequence of one of the results of Elogne et al. (2009). Preliminarily, let us recall (see for example Section 3 of the Appendix of Hannan 1970) that under regularity and integrability conditions on the squared correlation function (recall that Z being assumed centered and Gaussian, the autocorrelation function of Z^2 is simply ρ^2_{ν,θ_0}) which are clearly satisfied for the Matérn family, that

$$T_n \mathrm{E}\left(\left(\frac{1}{T_n}\int_0^{T_n} Z(t)^2 dt - \tau_0^2\right)^2\right) \to 2\tau_0^4 \int_{-\infty}^{\infty} \rho_{\nu,\theta_0}^2(t)dt \quad \text{as } T_n \to \infty.$$

Proposition 1. If Z has a Matérn autocorrelation function with $\nu > 1$, and

$$\hat{\tau}_{\mathrm{wEV}}^2 := \frac{1}{T_n}\sum_{i=1}^{n}(t_{i+1/2,n} - t_{i-1/2,n})Z(t_{i,n})^2$$

with the mid-points defined by $t_{i+1/2,n} = (t_{i+1,n}+t_{i,n})/2, i = 1, \cdots, n-1$ and $t_{1/2,n} = 0$, $t_{n+1/2,n} = T_n$, then

$$T_n \mathrm{E}\left(\left(\hat{\tau}_{\mathrm{wEV}}^2 - \tau_0^2\right)^2\right) < 2\tau_0^4 \int_{-\infty}^{\infty} \rho_{\nu,\theta_0}^2(t)dt + O(\delta_n^{\nu})$$

where the O term is uniform in T_n as $n \to \infty$.

Proof To apply Theorem 2 of Elogne et al. (2009), it suffices to check that $X := Z^2$ is mean square differentiable and satisfies: there exist κ such that, for all abscissae s and t

$$\left(\mathrm{E}\left(\left|X'(s) - X'(t)\right|^2\right)\right)^{1/2} \leq \kappa|s-t|^{\gamma},$$

where $\gamma = \nu - 1 > 0$. It is a relatively easy exercise to show this from the relation between the above expectation and the behavior of the second derivative, near zero, of the autocorrelation function of X (see e.g. Chapter 2 of Stein 1999).

The condition $\nu > 1$ is restrictive. We believe that one should be able to relax it and furthermore to establish bounds better than $O(\delta_n^{\nu})$, as is done in Elogne et al. (2008) but for an estimate that is less simple (in short Z in place of Z^2 is interpolated and the estimate is defined as the integral of the squared interpolant; see their Corrolary 1). An improved control of the accuracy of $\hat{\tau}_{\mathrm{wEV}}^2$ might be useful to establish asymptotic properties of the resulting CGEM-wEV estimates of the range and the microergodic parameters.

3. Monte-Carlo simulation study

In this study, the domain on which the locations of the observations are regularly (except in Section 3.7) spaced, is a square and the grid is Cartesian, thus $n_1 = n_2 = \sqrt{n}$ and $\delta_1 = \delta_2 =: \delta$. Of course, multiplying both δ and the range θ_0^{-1} by a same constant, does not change the simulated observations. Thus we set $\delta = 1/\sqrt{n}$ everywhere (except in Section 3.7) so that the simulation settings be easily comparable with those of published studies.

Even though the known theoretical justification (Girard 2009) is given only for the case of very strong correlation between observations at neighboring sites, the simulation study that we present here, was done with not only "moderate" and "large" correlation ranges chosen for the true range, but also rather "small" correlation ranges. Somewhat arbitrarily we call a "small correlation range", a range for which θ_0^{-1} is about 0.05 or less, and call a "large range" one for which θ_0^{-1} is about 0.7 or more. If $\nu = 1/6$ then the considered θ_0^{-1} varies in $\{0.02, 0.05, 0.09, 0.125, 0.2, 0.3, 0.5, 0.7, 1., 2., 3.\}$. Otherwise we used slight variants of this set.

Recall that, in the case of no additive white noise (i.e. $\sigma_{\rm N} = 0$ in (1.1)), the actual value of τ_0 has no influence on the relative accuracy of the ML estimates (e.g. Zhang 2004). Here if we assume $\sigma_{\rm N} > 0$, it is then easy to see from (1.2) and (1.3) that, if both the observations $\mathbf{y}$ and the given $\sigma_{\rm N}$ are multiplied by a same constant c, then the resulting $\hat{\tau}_{\rm EV}$ will be multiplied by this constant and, $\hat{b}_{\rm EV}$ being thus unchanged, the new estimating equation will have the same root(s). And such an invariance can also be easily seen for the ML method. Thus, denoting by b_0 the SNR

$$b_0 := \frac{\tau_0^2}{\sigma_{\rm N}^2},$$

it is only through b_0 that τ_0 and $\sigma_{\rm N}$ influence the respective performance of CGEM-EV vs ML. Since we essentially consider cases with $\sigma_{\rm N} > 0$ we almost always use in the following b_0 instead of τ_0^2 as variance-parameter (equivalently, the following study, except in Section 3.8, is a study of equivalent settings obtained by normalizing the observations vector so that the noise level be 1). We present the obtained simulation results, firstly in the case of either essentially no additive white noise or "weak" noise ($b_0 = 10^{12}$), next, for settings with "moderate" noise, thirdly, with "rather strong" noise ($b_0 = 4$).

For $b_0 = 10^{12}$ the data could have been considered as exact data and the estimation methods (both ML and CGEM-EV) could have been simplified. However very ill-conditioned matrix inversions (especially in the case ν large) would then have appeared. So we still chose a model with additive white noise (we come back to this point in Section 3.8).

The first question is of course the one of the existence of a root for the CGEM-EV estimating equation in θ and its unicity. The following simulation results exhibit a quite satisfying behavior of the CGEM-EV from this point of view: in "almost" all the considered cases one observed a single root in a search interval (typically $[0.05, 100.]$ for θ) that might be considered by many readers as a "quite large" interval, while the numerical search was a rather exhaustive grid search (typically 700 values for θ equispaced in logarithmic scale). In fact, the "almost" term we use above, is only due to settings where the true correlation range θ_0^{-1} is quite small. Indeed, it is only for such ranges that it happened for a few percent of the replicates (see the results marked with "*" in the following Tables) that the CGEM-EV estimating equation (or its randomized-trace version) had no root.

Each displayed result is a summary over 1000 replicates. Recall that, if a random variable is normally distributed, its observed standard deviation over 1000 replicates is an estimate of its true standard deviation with a relative accuracy of $\sqrt{1/2000} \approx 2\%$.

Note that in the following statistical summaries, we use a logarithmic transformation for the estimates of θ_0 because it has consistently been observed that this is necessary to produce "nearly" normal distribution (at least, the empirical distributions of $\log_{10}(\hat{\theta}_{\mathrm{ML}})$ or of $\log_{10}(\hat{\theta}_{\mathrm{CGEM-EV}})$, are generally much more symmetric than the ones of $\hat{\theta}_{\mathrm{ML}}$ or $\hat{\theta}_{\mathrm{CGEM-EV}}$). Note that this was not necessary for the considered estimates of $\tau_0\theta_0^{2\nu}$. The term "inefficiency" of a particular estimator, for example the randomized-trace version of the CGEM-EV (denoted randCGEM-EV) estimator of $\log_{10}(\theta_0)$, means here, as usual, the ratio of the observed mean squared error over 1000 replicates (denoted MSE) of this estimate to the MSE (same replicates) of the corresponding ML estimator. The columns labelled "ineff$^{1/2}$" display the square root of such observed inefficiencies.

3.1. For $\nu = 1/6$ and $b_0 = 10^{12}$ we clearly see in Table 1 that the relative accuracy of ML estimation for the microergodic parameter $b_0\theta_0^{2\nu}$ is extremely close to $\sqrt{2/n} = 0.047$ for any θ_0^{-1} between 0.09 and 3 (and it sligthly departs from 0.047 for the two smaller correlation ranges). The attractive property of the CGEM-EV estimation of $b_0\theta_0^{2\nu}$ is that it is practically as efficient as ML for θ_0^{-1} between 0.05 and 2 (with scarcely perceptible loss). Furthermore, the loss in efficiency is rather small for $\theta_0^{-1} = 3$ and reasonable for $\theta_0^{-1} = 0.02$. Concerning the range-parameter, its true magnitude is, as expected from the recent literature (especially Zhang (2004)), much less easily estimable, especially for large θ_0^{-1}. Nevertheless CGEM-EV performs also nearly as well as ML: neither the bias of $\log_{10}(\hat{\theta})$ nor its standard deviation are significantly increased by using CGEM-EV instead of ML.

3.2. For $\nu = 3/2$ and $b_0 = 10^{12}$ we again see in Table 2 that the relative accuracy of ML estimation for the microergodic parameter $b_0\theta_0^{2\nu}$ is almost always very close to $\sqrt{2/n}$ with a noticeable small degradation for small θ_0^{-1} and a "big" departure from $\sqrt{2/n}$ only for $\theta_0^{-1} = 0.02$. The attractive property of the CGEM-EV estimation is that its efficiency is still quite good, although a departure from 1 of is now perceptible: CGEM-EV root-inefficiency relative to ML (that is the ratio of the two resulting root mean square errors) is always between 1.05 and 1.10, except for $\theta_0^{-1} = 0.02$. For this small range setting, a small degradation is noticeable. Concerning the estimation of $\log_{10}(\theta_0)$ the performance of CGEM-EV, as compared to ML, is also not as excellent as in Table 1. This seems to be in agreement with the theoretical result of Girard (2009) which states that full-efficiency is obtained as ν decreases.

3.3. For $\nu = 1/2$ (i.e. exponential model) and $b_0 = 10^3$ the results concerning the estimation of $b_0\theta_0^{2\nu}$ in Table 3 are rather similar to the ones in Table 1 except for $\theta_0^{-1} = 0.02$ for which there were 48 replicates among the 1000 for which the CGEM-EV estimating equation had no root. The results concerning the estimation of $\log_{10}(\theta_0)$ are intermediate between the corresponding results in Table 1 and Table 2.

A spherical model which can be thought "similar" to the previous one is considered in Table 4 (except that n is here 20×20). Thus ν is chosen equal to 1/2 (as discussed in Zhang and Zimmerman, 2007). The results are very similar to those of Table 3 although the efficiency of CGEM-EV is a little bit degraded.

3.4. For $\nu = 3/2$ and $b_0 = 10^3$ (and $n = 30 \times 30$) we first see in Table 5 that the previously observed accuracy of ML estimation for the microergodic parameter is much decreased when the range θ_0^{-1} increases. Some theoretical studies have established (essentially in one-dimensional setting) that, in the infill asymptotics framework, the relative accuracy of $\sqrt{2/n}$ which holds in the case of exact data (Du, Zhang and Mandrekar 2009) is lost as soon as the observations are contaminated by a white noise. These experiments, compared with those of section 3.3, show that a quite moderate noise has a rather strong impact on the attainable accuracies when $\nu = 3/2$ whileas its impact is weak for the exponential or the spherical models. Now an important result is that the efficiency of CGEM-EV for the microergodic parameter remains quite good when the true range θ_0^{-1} remains smaller than 0.7; otherwise is not as satisfying as in all the previous settings.

By comparing Table 2 and Table 5, we conclude that the signal-to-noise may have a large impact on the efficiency of CGEM-EV. This is not in complete agreement with the theoretical result of Girard (2009, second part of Theorem 4.1); but recall that this result describes a limit for a particular asymptotic regime.

A second important result seen in Table 5 is that the replacement of the exact traces in the CGEM-EV estimating equation by their randomized version, does not degrade the performance of CGEM-EV provided at least about 20 replicates are used for each randomized trace approximation.

3.5. We next presents cases with $b_0 = 10$ or even $b_0 = 4$ in Table 6, Table 7 and Table 8. These 3 tables concern respectively $\nu = 1/2$, $\nu = 1/6$ and the spherical model (akin to $\nu = 1/2$). The displayed results show that for such ν the SNR has a much weaker impact than for $\nu = 3/2$. On the subject of the CGEM-EV efficiency relatively to ML, the results are quite similar to the corresponding previous tables for large b_0 (resp. Table 3, Table 1 and Table 4). But as expected the relative accuracy of the ML estimate of the microergodic parameter now deviates from the theoretical $\sqrt{2/n}$ especially for the large θ_0^{-1}'s.

Thus in the case of "rather strong" noise, the CEGEM-EV approach is still quite efficient provided ν is not too large.

3.6. On the subject of what is sacrificed in using randomized-traces instead of exact traces, the previously discussed Table 6, Table 7 and Table 8 also demonstrate that the CGEM-EV efficiency is practically never degraded with $n_R = 20$. In fact, even using $n_R = 1$ induces a negligible degradation in the settings of Tables 1, 2, 3, 4, 7, 8 (note that the columns corresponding the randomized CGEM-EV are not displayed in Tables 1, 2, 3, 4 because they would have been equal, for this display using 3 digits, to the columns of the exact CGEM-EV). However increasing n_R (from1 to 20) does improve the efficiency of CGEM-EV in the case of large correlation range and weak SNR (see the last line , i.e. $\theta_0^{-1} = 3$, of Table 5, Table 6 or Table 7).

3.7. We now present obtained simulation results in the case of the non-uniform Cartesian grid (2.1) with the exponential model and $b_0 = 10^4$, in Table 9. The displayed results clearly show that the CGEM-EV efficiency relatively to ML is degraded as compared to the uniform grid case in Table 3, but the Riemann-sum based weighted version CGEM-wEV restores it quite well, both for the range parameter and the microergodic parameter. Note that the setting here essentially differs from the setting analyzed in Section 3.3 only by the deformation of the grid (however the noise-level is slightly different). It is interesting to observe that the accuracies obtained by CGEM-wEV are very similar to those obtained in the equispaced case (compare the penultimate column of Table 3 with the penultimate column of Table 9).

3.8. We finally consider a setting for which CGEM-EV is compared with the so-called "Hybrid method", that is, the method proposed by Zhang and Zimmerman (2007) briefly described in the Introduction. The setting is in fact exactly the first of those considered by these authors in their simulation study. The correlation model was the exponential model and four values of range parameter were chosen (we added a fifth value, that is $\theta_0^{-1} = 1.5$). There was no additive white noise in the data. Since, in this paper, we maintain a model (1.1) with a given $\sigma_{\rm N} > 0$ in our estimates (thus a misspecified model), the problem of choosing $\sigma_{\rm N}$ arises. We have tried three values for $\sigma_{\rm N}^2$: two moderately small values (.002 and .0005) and a very small one (10^{-8}). Firstly, this simulation study with "exact" data demonstrates that CGEM-EV is quite insensitive to $\sigma_{\rm N}$ over several orders of magnitudes: indeed the results dispayed in the corresponding 3 columns of Table 10 are hardly distinguishable, the only exception is the case of large range ($\theta_0^{-1} = 1.5$) where the largest of the three $\sigma_{\rm N}^2$'s yields a noticeable degradation of the performance of CGEM-EV. Thus a simple rule which should prevent such degradation is the following: if one suspects a strong correlation in the true model (and one knows that there is no additive white noise), the (miss)specified $\sigma_{\rm N}$ must be chosen "small" enough.

Secondly, the summaries in Table 10, where the column labelled "Hyb" is in fact a copy of the summaries displayed in Zhang and Zimmerman (2007, Table 1), demonstrates that the Hybrid estimates are clearly less efficient than the CGEM-EV estimates (and especially for the estimation of τ_0^2) in this setting.

4. Conclusion and further questions

A rather extensive simulation study was performed for Matérn random fields with $\nu \in \{1/6, 1/2, 3/2\}$. A side remark is that the magnitude of the SNR has a strong impact on the results (even those of ML), at least from the inference point of view taken here.

Firstly, concerning the question of existence and unicity of the root, the CGEM-EV method proved to be satisfying for all the settings considered here. More importantly for the usefulness of this approach, these experiments demonstrate that the CGEM-EV variance and range-parameter estimates, and, above all, the resulting microergodic-parameter estimate, are nearly as efficient as the ML estimates for many various settings provided ν is not too large and the SNR is not too weak. (Notice that the precise meaning of "not too weak" depends on ν since the results are still good for a SNR of 4 when $\nu = 1/6$.)

Otherwise this efficiency may be degraded especially for the cases with large range-parameter. We do not know yet whether this rather deceiving behavior would be rectified for larger n. Anyway, in such "unfavorable" settings the CGEM-EV variance and range-parameter estimates might nevertheless be a useful starting point for a classical one-Newton-step based on the linearized likelihood equations. This may deserve a deeper investigation.

As is usual for any point-estimation method, it would be useful that this method be supplemented by accuracy estimates, for example to build confidence band for the underlying correlation function. For all the contexts where numerically solving the CGEM-EV equation is reasonably fast, it is tempting to consider "parametric bootstrap"-type confidence bands. Further works are necessary to develop and assess such methods.

This work mainly concerned designs (for the locations of the observations) which coincide with a uniform grid or with a simple variant of a uniform grid. It is clear that for the considered variant (i.e. the experiment whose results are reported in Table 9), the weighted version CGEM-wEV, based on (2.3), produced a rather impressive improvement of CGEM-EV. The idea of using weights based on a cubature rule which approximates the integral of Z^2 is thus promising since it can guide us to extend CGEM-wEV to other designs. The choice of a cubature rule appropriate to any irregular designs which would combine computational efficiency and (asymptotic?) statistical efficiency, is not simple and it deserves thus further study.

The final experiment (reported in Table 10) demonstrates that, in the case of no additive white noise, CGEM-EV can be much more efficient than the hybrid method proposed by Zhang and Zimmerman, and is very robust with respect to the somewhat arbitrary choice of a "small" σ_{N} that CGEM-EV requires here for the noise level. However in the case of a non-negligible additive white noise, we made no comparison with the results Zhang and Zimmerman (2007) since the present version of CGEM-EV requires that σ_{N} be known. It is clear that it would be useful to extend CGEM-EV to the case of unknown noise level (and to possibly heteroscedastic measurement errors).

REFERENCES

Chen, W., Hurvich, C.M., and Lu, Y., 2006. On the correlation matrix of the discrete Fourier transform and the fast solution of large Toeplitz systems for long-memory time series. J. Amer. Statist. Assoc. 101(474), 812822.

Du, J., Zhang, H., Mandrekar, V., 2009. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. Ann. Statist. 37, 3330-3361.

Elogne, S., Perrin, O., Thomas-Agnan, C. 2008. Non parametric estimation of smooth stationary covariance functions by interpolation methods. Statistical Inference for Stochastic Processes, 11, 177-205.

Fritz, J., Neuweiler, I., Nowak, W., 2009. Application of FFT-based Algorithms for Large-Scale Universal Kriging Problems. Mathematical Geosciences, 41, 509-533.

Girard, D.A., 2009. Asymptotic near-efficiency of a "Gibbs-energy" estimating function approach for fitting Matérn covariance models to a dense (noisy) series. Preprint. http://hal.archives-ouvertes.fr/hal-00413693/en/

Hannan E.J., 1970. Multiple time series. New-York London Sydney Toronto : John Wiley and Sons, Inc.

Kessler, M., 2000. Simple and explicit estimating functions for a discretely observed diffusion process. Scand. J. Statist. 27, 65-82.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press.

Shin, D.W., Song, S.,H., 2000. Asymptotic efficiency of the OLSE for polynomial regression models with spatially correlated errors Statistics & Probability Letters, 47, 1 - 10.

Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer.

Yajima, Y., 1991. Asymptotic properties of the LSE in a regression model with long-memory stationary errors. Ann. Statist. 19 158177

Zhang, H., 2004. Inconsistent estimation and asymptotically equivalent interpolation in model-based geostatistics. J. Amer. Statist. Assoc. 99, 250-261.

Zhang, H., Zimmerman, D.L., 2007. Hybrid estimation of semivariogram parameters. Mathematical Geology 39 (2), 247-260.

Table 1: $n = 30 \times 30$. Simulation summary (mean, standard deviation of ML estimates, CGEM-EV estimates and respective MSE inefficiency) for Matern model with $\nu = 1/6$ and $b_0 = 10^{12}$

	ML	CGEM-EV	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$		
0.02	0.00± 0.07	0.01± 0.09	1.29
0.05	0.00± 0.11	0.00± 0.11	1.04
0.09	0.00± 0.16	0.01± 0.17	1.04
0.125	0.01± 0.20	0.02± 0.21	1.05
0.2	0.02± 0.27	0.04± 0.28	1.05
0.3	0.04± 0.34	0.06± 0.36	1.06
0.5	0.09± 0.42	0.11± 0.46	1.09
0.7	0.13± 0.49	0.14± 0.53	1.09
1.	0.17± 0.55	0.18± 0.60	1.09
2.	0.26± 0.69	0.27± 0.74	1.07
3.	0.32± 0.77	0.33± 0.82	1.05
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/\left(b_0\theta_0^{2\nu}\right)$		
0.02	1.003± 0.057	1.007± 0.072	1.273
0.05	1.003± 0.049	1.003± 0.050	1.015
0.09	1.003± 0.047	1.004± 0.048	1.012
0.125	1.003± 0.047	1.004± 0.047	1.004
0.2	1.004± 0.046	1.004± 0.046	1.003
0.3	1.003± 0.046	1.005± 0.046	1.002
0.5	1.003± 0.046	1.005± 0.046	1.010
0.7	1.003± 0.046	1.004± 0.046	1.004
1.	1.003± 0.046	1.004± 0.046	1.005
2.	1.003± 0.046	1.004± 0.046	1.008
3.	1.003± 0.046	1.005± 0.049	1.068

Table 2: $n = 30 \times 30$. Simulation summary (mean, standard deviation of ML estimates, CGEM-EV estimates and respective MSE inefficiency) for Matern model with $\nu = 3/2$ and $b_0 = 10^{12}$. Results with * are averages after removal of 1 "outlier" among the 1000 replicates

	ML	CGEM-EV	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$		
0.02	0.00± 0.03	0.00* ± 0.03*	1.13*
0.04	0.00± 0.02	0.00± 0.02	1.10
0.09	0.00± 0.03	0.00± 0.03	1.17
0.125	0.00± 0.03	0.00± 0.04	1.21
0.2	0.00± 0.05	0.01± 0.06	1.25
0.3	0.00± 0.06	0.02± 0.07	1.32
0.5	0.01± 0.08	0.04± 0.10	1.38
0.7	0.02± 0.09	0.06± 0.12	1.47
1.	0.02± 0.11	0.08± 0.15	1.50
2.	0.04± 0.14	0.12± 0.20	1.53
3.	0.05± 0.16	0.14± 0.22	1.54
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/(b_0\theta_0^{2\nu})$		
0.02	1.017± 0.166	1.029* ± 0.189*	1.149*
0.04	1.005± 0.089	1.007± 0.096	1.082
0.09	1.003± 0.060	1.006± 0.066	1.096
0.125	1.002± 0.055	1.007± 0.061	1.113
0.2	1.003± 0.050	1.009± 0.054	1.080
0.3	1.002± 0.048	1.010± 0.051	1.083
0.5	1.003± 0.047	1.010± 0.050	1.080
0.7	1.003± 0.046	1.010± 0.049	1.069
1.	1.002± 0.046	1.010± 0.049	1.079
2.	1.002± 0.046	1.008± 0.048	1.049
3.	1.002± 0.046	1.008± 0.048	1.065

Table 3: $n = 27 \times 27$. Simulation summary (mean, standard deviation of ML estimates, CGEM-EV estimates and respective MSE inefficiency) for Exponential model with $b_0 = 10^3$. Results with * are averages after removal of 48 "outliers" among the 1000 replicates

	ML	CGEM-EV	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$		
0.02	0.00± 0.04	0.01* ± 0.05*	1.18*
0.05	0.00± 0.05	0.00± 0.05	1.07
0.09	0.01± 0.07	0.01± 0.07	1.10
0.125	0.01± 0.08	0.01± 0.09	1.12
0.2	0.01± 0.12	0.02± 0.13	1.12
0.3	0.02± 0.15	0.04± 0.17	1.18
0.5	0.04± 0.21	0.07± 0.23	1.16
0.7	0.06± 0.25	0.10± 0.28	1.16
1.	0.08± 0.30	0.13± 0.32	1.14
1.5	0.12± 0.33	0.17± 0.38	1.17
3.	0.20± 0.43	0.25± 0.47	1.12
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/\left(b_0\theta_0^{2\nu}\right)$		
0.02	1.006± 0.100	1.015* ± 0.114*	1.154*
0.05	1.002± 0.065	1.003± 0.066	1.023
0.09	1.001± 0.059	1.002± 0.060	1.015
0.125	1.000± 0.057	1.003± 0.058	1.024
0.2	1.000± 0.055	1.003± 0.056	1.020
0.3	1.000± 0.055	1.003± 0.056	1.018
0.5	1.000± 0.055	1.003± 0.055	1.013
0.7	1.000± 0.055	1.003± 0.055	1.019
1.	1.000± 0.055	1.003± 0.056	1.014
1.5	1.000± 0.055	1.002± 0.056	1.012
3.	0.999± 0.058	1.002± 0.059	1.024

Table 4: $n = 20 \times 20$. Simulation summary (mean, standard deviation of ML estimates, CGEM-EV estimates and respective MSE inefficiency) for Spherical model with $b_0 = 10^3$

	ML	CGEM-EV	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$		
0.2	-0.00± 0.03	0.01± 0.06	2.09
0.3	-0.01± 0.06	0.02± 0.09	1.68
0.5	-0.01± 0.08	0.03± 0.14	1.82
0.7	0.01± 0.10	0.05± 0.18	1.87
1.	0.07± 0.16	0.07± 0.23	1.41
1.5	0.11± 0.24	0.11± 0.29	1.17
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/\left(b_0\theta_0^{2\nu}\right)$		
0.2	1.00± 0.07	1.02± 0.08	1.11
0.3	1.00± 0.07	1.03± 0.08	1.19
0.5	1.00± 0.07	1.02± 0.08	1.11
0.7	1.00± 0.07	1.02± 0.07	1.07
1.	1.00± 0.07	1.01± 0.07	1.04
1.5	1.00± 0.07	1.01± 0.07	1.04

Table 5: $n = 30 \times 30$. Simulation summary (mean, standard deviation of ML estimates, CGEM-EV estimates -with exact or randomized traces- and respective MSE inefficiency) for Matern model with $\nu = 3/2$ and $b_0 = 10^3$

	ML	CGEM-EV		randCGEM-EV			
				$n_R = 1$		$n_R = 20$	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$						
0.04	0.00± 0.02	0.00± 0.02	1.11	0.00± 0.02	1.11	0.00± 0.02	1.11
0.09	0.00± 0.03	0.00± 0.03	1.16	0.00± 0.03	1.16	0.00± 0.03	1.16
0.125	-0.00± 0.03	0.00± 0.04	1.22	0.00± 0.04	1.21	0.00± 0.04	1.22
0.2	0.00± 0.05	0.01± 0.06	1.25	0.01± 0.06	1.25	0.01± 0.06	1.25
0.3	0.00± 0.06	0.02± 0.08	1.33	0.02± 0.08	1.33	0.02± 0.08	1.33
0.5	0.01± 0.08	0.04± 0.11	1.38	0.04± 0.11	1.39	0.04± 0.11	1.38
0.7	0.02± 0.10	0.06± 0.13	1.47	0.06± 0.13	1.49	0.06± 0.13	1.48
1.	0.02± 0.12	0.08± 0.16	1.50	0.08± 0.16	1.52	0.08± 0.16	1.51
2.	0.05± 0.16	0.13± 0.22	1.52	0.13± 0.22	1.55	0.13± 0.22	1.52
3.	0.06± 0.18	0.16± 0.25	1.53	0.16± 0.26	1.56	0.16± 0.25	1.53
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/\left(b_0\theta_0^{2\nu}\right)$						
0.04	1.00± 0.09	1.01± 0.10	1.09	1.01± 0.10	1.09	1.01± 0.10	1.09
0.09	1.00± 0.06	1.01± 0.07	1.09	1.01± 0.07	1.09	1.01± 0.07	1.09
0.125	1.00± 0.06	1.01± 0.07	1.11	1.01± 0.07	1.10	1.01± 0.07	1.11
0.2	1.00± 0.06	1.01± 0.07	1.09	1.01± 0.07	1.10	1.01± 0.07	1.09
0.3	1.00± 0.07	1.02± 0.08	1.11	1.02± 0.08	1.16	1.02± 0.08	1.11
0.5	1.01± 0.10	1.03± 0.11	1.13	1.03± 0.13	1.32	1.03± 0.11	1.15
0.7	1.01± 0.12	1.05± 0.14	1.22	1.05± 0.17	1.48	1.05± 0.14	1.24
1.	1.01± 0.15	1.07± 0.19	1.31	1.08± 0.24	1.65	1.07± 0.19	1.34
2.	1.03± 0.23	1.14± 0.32	1.47	1.18± 0.42	1.94	1.14± 0.32	1.49
3.	1.05± 0.30	1.22± 0.47	1.73	1.29± 0.64	2.32	1.22± 0.48	1.77

Table 6: $n = 27 \times 27$. Simulation summary (mean, standard deviation of ML estimates, CGEM-EV estimates -with exact or randomized traces- and respective MSE inefficiency) for Exponential model and $b_0 = 10$. Results with * are averages after removal of 66 "outliers" among the 1000 replicates

	ML	CGEM-EV		randCGEM-EV			
				$n_R = 1$		$n_R = 20$	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$						
0.02	0.00± 0.05	0.00* ± 0.05*	1.18*	0.00* ± 0.05*	1.18*	0.00* ± 0.05*	1.18*
0.05	0.00± 0.05	0.00± 0.05	1.08	0.00± 0.05	1.08	0.00± 0.05	1.08
0.09	0.01± 0.07	0.01± 0.08	1.10	0.01± 0.08	1.11	0.01± 0.08	1.11
0.125	0.01± 0.09	0.01± 0.10	1.12	0.01± 0.10	1.13	0.01± 0.10	1.12
0.2	0.01± 0.13	0.02± 0.14	1.12	0.02± 0.14	1.12	0.02± 0.14	1.12
0.3	0.02± 0.16	0.04± 0.18	1.16	0.04± 0.18	1.17	0.04± 0.18	1.16
0.5	0.04± 0.22	0.07± 0.25	1.15	0.07± 0.25	1.15	0.07± 0.25	1.15
0.7	0.06± 0.26	0.10± 0.29	1.17	0.10± 0.29	1.17	0.10± 0.29	1.17
1.	0.09± 0.31	0.13± 0.34	1.13	0.13± 0.34	1.14	0.13± 0.34	1.13
1.5	0.12± 0.35	0.18± 0.40	1.17	0.18± 0.40	1.18	0.18± 0.40	1.17
3.	0.20± 0.45	0.25± 0.50	1.13	0.25± 0.50	1.14	0.25± 0.50	1.13
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/(b_0\theta_0^{2\nu})$						
0.02	1.01± 0.11	1.01* ± 0.13*	1.17*	1.01* ± 0.12*	1.16*	1.01* ± 0.13*	1.17*
0.05	1.00± 0.08	1.00± 0.08	1.04	1.00± 0.08	1.04	1.00± 0.08	1.05
0.09	1.00± 0.08	1.00± 0.08	1.03	1.00± 0.08	1.04	1.00± 0.08	1.03
0.125	1.00± 0.08	1.00± 0.08	1.04	1.01± 0.08	1.06	1.01± 0.08	1.05
0.2	1.00± 0.09	1.01± 0.09	1.04	1.01± 0.09	1.07	1.01± 0.09	1.04
0.3	1.00± 0.10	1.01± 0.10	1.04	1.01± 0.11	1.09	1.01± 0.10	1.04
0.5	1.00± 0.12	1.01± 0.12	1.03	1.01± 0.13	1.13	1.01± 0.12	1.04
0.7	1.00± 0.13	1.02± 0.14	1.04	1.02± 0.15	1.18	1.02± 0.14	1.05
1.	1.01± 0.15	1.02± 0.15	1.04	1.02± 0.18	1.21	1.02± 0.16	1.05
1.5	1.01± 0.17	1.03± 0.18	1.06	1.03± 0.22	1.27	1.03± 0.18	1.07
3.	1.01± 0.22	1.03± 0.24	1.08	1.05± 0.30	1.40	1.04± 0.24	1.10

Table 7: $n = 30 \times 30$. Simulation summary (mean, standard deviation of ML estimates, CGEM-EV estimates -with exact or randomized traces- and respective MSE inefficiency) for Matern model with $\nu = 1/6$ and $b_0 = 4$. Results with * are averages after removal of 67 "outliers" among the 1000 replicates

	ML	CGEM-EV		randCGEM-EV			
				$n_R = 1$		$n_R = 20$	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$						
0.02	-0.00± 0.08	-0.01* ± 0.09*	1.11*	-0.01* ± 0.09*	1.19*	-0.01* ± 0.09*	1.12*
0.05	0.00± 0.11	0.00± 0.12	1.06	0.00± 0.12	1.07	0.00± 0.12	1.06
0.09	0.00± 0.17	0.01± 0.17	1.03	0.01± 0.17	1.03	0.01± 0.17	1.03
0.125	0.01± 0.20	0.02± 0.21	1.06	0.02± 0.21	1.06	0.02± 0.21	1.06
0.2	0.02± 0.27	0.04± 0.29	1.05	0.04± 0.30	1.11	0.04± 0.29	1.05
0.3	0.04± 0.34	0.06± 0.36	1.06	0.06± 0.37	1.10	0.07± 0.36	1.06
0.5	0.09± 0.43	0.11± 0.47	1.11	0.11± 0.47	1.12	0.11± 0.47	1.11
0.7	0.13± 0.49	0.14± 0.54	1.09	0.14± 0.54	1.10	0.14± 0.54	1.09
1.	0.17± 0.56	0.19± 0.61	1.08	0.19± 0.61	1.09	0.19± 0.61	1.08
2.	0.26± 0.70	0.28± 0.75	1.07	0.28± 0.75	1.07	0.28± 0.75	1.07
3.	0.32± 0.78	0.34± 0.82	1.06	0.33± 0.83	1.06	0.33± 0.83	1.06
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/(b_0\theta_0^{2\nu})$						
0.02	1.00± 0.07	1.00* ± 0.08*	1.05*	1.00* ± 0.08*	1.10*	1.00* ± 0.08*	1.05*
0.05	1.00± 0.07	1.00± 0.07	1.00	1.01± 0.07	1.03	1.01± 0.07	1.03
0.09	1.00± 0.07	1.01± 0.07	1.00	1.01± 0.07	1.02	1.01± 0.07	1.01
0.125	1.01± 0.07	1.01± 0.07	1.00	1.01± 0.07	1.01	1.01± 0.07	1.01
0.2	1.01± 0.07	1.01± 0.07	1.01	1.01± 0.07	1.02	1.01± 0.07	1.01
0.3	1.01± 0.08	1.01± 0.08	1.00	1.01± 0.08	1.02	1.01± 0.08	1.01
0.5	1.01± 0.08	1.01± 0.08	1.01	1.01± 0.08	1.03	1.01± 0.08	1.01
0.7	1.01± 0.08	1.01± 0.08	1.01	1.01± 0.09	1.03	1.01± 0.08	1.01
1.	1.01± 0.09	1.01± 0.09	1.01	1.01± 0.09	1.03	1.01± 0.09	1.01
2.	1.01± 0.10	1.01± 0.10	1.01	1.01± 0.10	1.05	1.01± 0.10	1.02
3.	1.01± 0.10	1.01± 0.11	1.02	1.01± 0.11	1.10	1.01± 0.11	1.06

Table 8: $n = 20 \times 20$. Simulation summary (mean, standard deviation of ML estimates, CGEM-EV estimates -with exact or randomized traces- and respective MSE inefficiency) for Spherical model and $b_0 = 10$

	ML	CGEM-EV		randCGEM-EV			
				$n_R = 1$		$n_R = 20$	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$						
0.2	-0.00± 0.04	0.01± 0.06	1.58	0.01± 0.06	1.59	0.01± 0.06	1.57
0.3	-0.01± 0.07	0.02± 0.09	1.31	0.02± 0.09	1.33	0.02± 0.09	1.31
0.5	-0.01± 0.09	0.03± 0.14	1.54	0.03± 0.14	1.55	0.03± 0.14	1.54
0.7	0.01± 0.12	0.05± 0.18	1.59	0.05± 0.18	1.61	0.05± 0.18	1.59
1.	0.07± 0.17	0.07± 0.23	1.30	0.07± 0.23	1.32	0.07± 0.23	1.30
1.5	0.10± 0.24	0.11± 0.29	1.19	0.11± 0.29	1.20	0.11± 0.29	1.19
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/\left(b_0\theta_0^{2\nu}\right)$						
0.2	1.00± 0.09	1.01± 0.09	1.05	1.01± 0.10	1.07	1.01± 0.09	1.05
0.3	1.00± 0.10	1.02± 0.11	1.08	1.02± 0.11	1.10	1.02± 0.11	1.08
0.5	0.99± 0.12	1.02± 0.13	1.07	1.03± 0.13	1.12	1.02± 0.13	1.07
0.7	0.99± 0.13	1.02± 0.14	1.06	1.02± 0.15	1.13	1.02± 0.14	1.06
1.	1.00± 0.15	1.02± 0.16	1.04	1.02± 0.17	1.12	1.02± 0.16	1.04
1.5	1.00± 0.18	1.02± 0.18	1.05	1.02± 0.21	1.18	1.02± 0.19	1.07

Table 9: Nonuniform grid with $n = 27 \times 27$. Simulation summary (mean, standard deviation of ML, CGEM-EV and CGEM-wEV estimates, and respective MSE inefficiency) for exponential model and $b_0 = 10^4$ (CGEM-wEV uses (2.1) in place of $\hat{b}_{\mathrm{EV}}$)

	ML	CGEM-EV		CGEM-wEV	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$				
0.05	0.00± 0.06	0.01± 0.10	1.71	0.00± 0.07	1.18
0.09	0.01± 0.07	0.02± 0.13	1.82	0.01± 0.08	1.10
0.2	0.02± 0.12	0.05± 0.20	1.64	0.02± 0.13	1.10
0.3	0.02± 0.16	0.07± 0.23	1.49	0.03± 0.18	1.11
0.5	0.03± 0.22	0.10± 0.28	1.33	0.06± 0.24	1.11
0.7	0.05± 0.26	0.13± 0.32	1.30	0.09± 0.29	1.14
1.	0.09± 0.30	0.16± 0.36	1.27	0.12± 0.33	1.14
1.5	0.12± 0.34	0.20± 0.41	1.27	0.16± 0.38	1.16
	summary for the ratios $\hat{b}\hat{\theta}^{2\nu}/(b_0\theta_0^{2\nu})$				
0.05	0.999± 0.058	1.009± 0.068	1.171	1.000± 0.061	1.047
0.09	0.999± 0.056	1.011± 0.064	1.167	1.001± 0.057	1.019
0.2	0.999± 0.054	1.010± 0.059	1.104	1.001± 0.055	1.015
0.3	0.999± 0.054	1.008± 0.058	1.079	1.001± 0.054	1.009
0.5	0.999± 0.054	1.006± 0.056	1.054	1.001± 0.054	1.010
0.7	0.999± 0.054	1.005± 0.056	1.046	1.002± 0.054	1.009
1.	0.999± 0.054	1.004± 0.055	1.031	1.001± 0.054	1.006
1.5	0.999± 0.054	1.004± 0.055	1.027	1.001± 0.054	1.011

Table 10: $n = 20 \times 20$. Same setting as in Zhang and Zimmerman (2007, Table 1). Simulation summary (mean and mean squared error (MSE) of ML estimates, Hyb, CGEM-EV estimates -with misspecified σ_{N}) for Exponential model with $\tau_0^2 = 2$ and without additive white noise

	ML	Hyb	CGEM-EV		
			$\sigma_{\mathrm{N}}^2 = .002$	$\sigma_{\mathrm{N}}^2 = .0005$	$\sigma_{\mathrm{N}}^2 = 10^{-8}$
θ_0^{-1}	mean of the errors $\hat{\tau}^2 - \tau_0^2$				
0.1	0.01	0.02	0.0091		0.0111
0.2	0.03	0.2	0.0029		0.0049
0.3	0.04	0.51	0.0101		0.0121
0.4	0.05	0.38	0.0150	0.0165	0.0170
1.5	0.12		0.0328	0.0343	0.0348
	MSE of $\hat{\tau}^2$				
0.1	0.11	0.26	0.1274		0.1274
0.2	0.35	1.84	0.3822		0.3822
0.3	0.64	5.08	0.7424		0.7425
0.4	0.87	6.19	1.1396	1.1396	1.1396
1.5	3.68		4.2818	4.2819	4.2819
	mean of the errors $\hat{\tau}^2\hat{\theta} - \tau_0^2\theta_0$				
0.1	0.0209	0.28	-0.0026		0.0601
0.2	-0.0047	0.12	-0.0272		0.0263
0.3	-0.0001	0.08	-0.0317		0.0207
0.4	-0.0017	0.09	-0.0367	0.0023	0.0163
1.5	0.0005		-0.0446	-0.0075	0.0051
	MSE of $\hat{\tau}^2\hat{\theta}$				
0.1	2.5494	3.71	2.6038		2.6070
0.2	0.5561	0.75	0.5779		0.5788
0.3	0.2346	0.31	0.2463		0.2455
0.4	0.1299	0.18	0.1362	0.1350	0.1353
1.5	0.0085		0.0106	0.0087	0.0087