



HAL
open science

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

Didier A. Girard

► To cite this version:

Didier A. Girard. 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. 2010. hal-00515832v1

HAL Id: hal-00515832

<https://hal.science/hal-00515832v1>

Preprint submitted on 8 Sep 2010 (v1), last revised 14 Dec 2018 (v4)

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

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

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, a 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 $Z(\mathbf{s})$, $\mathbf{s} \in [0, 1] \times [0, 1]$, be a zero mean stationary gaussian stochastic process whose autocorrelation function is assumed to belong to the isotropic Matern family. One realization of this process is observed at $n = n_1^2$ regularly spaced sites $\mathbf{s}_i, i = 1, \dots, n$, of $[0, 1] \times [0, 1]$ with an additive white noise (this noise can model either suspected measurement errors or an intrinsic nugget effect in Z). In this article, we restrict ourselves to the case where the additive white noise has a known variance, that we set equal to 1 to simplify the notation. Using a standard lexicographic ordering, the observations thus form a vector \mathbf{y} of size n whose law is

$$\mathbf{y} \sim N(0, b_0 R_{\theta_0} + I) \text{ where } b_0 = E((Z(\mathbf{s}))^2) \text{ is the process variance,} \quad (1.1)$$

with I denoting the identity matrix and R_θ the autocorrelation matrix of the gridded process i.e. the block Toeplitz, with Toeplitz blocks, matrix whose coefficients are given by

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

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 ν . When mentioned, we will also consider another well known autocorrelation function, namely the spherical model $\rho_\theta^S(x)$. See e.g. Zhang and Zimmerman 2007 for these definitions.

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 very commonly used (see Stein 1999, Rasmussen and Williams 2006). The parameter θ^{-1} is often called the range parameter.

Estimation of parameters in autocovariance models is needed for various tasks, for example for establishing confidence band for the autocovariance 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 weighted least square method (not statistically full-efficient but much less costly than maximum likelihood (ML)) to estimate the range parameters, next, to plug-in these parameters (the θ here) in the likelihood which is then maximized with respect to b (the solution, say $\hat{b}_{\text{ML}}(\theta)$, being typically obtained iteratively by Fisher scoring). The idea underlying this method is that, at least for the infill asymptotics context, even if θ is fixed at a wrong value θ_1 , the product $\hat{b}_{\text{ML}}(\theta_1)\theta_1^{2\nu}$ still remains an efficient estimator of $b_0\theta_0^{2\nu}$ which is called the microergodic parameter (see Du et al. (2009) for recent results of this type).

The method we introduced in Girard (2009), firstly reverses the roles of variance and range: it is based on a very simple estimate for the variance, namely the bias corrected empirical variance, which is simply defined by

$$\hat{b}_{\text{EV}} := n^{-1}\mathbf{y}^T\mathbf{y} - 1. \quad (1.2)$$

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

$$\mathbf{y}^T A_{b,\theta} (I - A_{b,\theta}) \mathbf{y} = \text{tr} A_{b,\theta} \quad \text{where} \quad A_{b,\theta} = bR_\theta (I + bR_\theta)^{-1}. \quad (1.3)$$

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 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 $\text{tr}A_{b,\theta}$, is quite satisfying provided the signal-to-noise b_0 is strong enough or ν is not too large.

2. Some comments on the CGEM-EV estimating equation

The motivation behind this work is that it should be possible to implement CGEM-EV for very large scale problems. Indeed from works in the 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 the preconditioned conjugate gradient approach. 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 $A_{b,\theta}\mathbf{y}$ which does not form explicitly the matrix $A_{b,\theta}$, one is tempted to try a randomized-trace approximation $((1/n_R) \sum_{r=1}^{n_R} \mathbf{w}_r^T A_{b,\theta} \mathbf{w}_r / (\mathbf{w}_r^T \mathbf{w}_r))$ in place of the exact $(1/n)\text{tr}A_{b,\theta}$. The particular application of the preconditioned conjugate gradient approach to problems like those of the following simulation is not the purpose of this article. Indeed in the following, we consider settings with relatively small data size (less than 900) so that using such iterative solvers is not mandatory. Classical “exact” Cholesky or eigenvalues-eigenvectors decompositions are then used in the following, even to implement randomized-trace versions of the CGEM-EV estimating equation. Note that, in following simulation, one 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.

A last comment is in order about the bias corrected empirical variance \hat{b}_{EV} defined in (1.2) as an estimate of the signal-to-noise ratio. Of course, even with a data set of large size, it may happen, especially in case of large correlation range, that the observed \hat{b}_{EV} has a negative value. It is easy to see that the probability of a such pathological event decreases when n increases and becomes very small as soon as the true b_0 is large enough. In fact this entails that CGEM-EV appears to be not suitable for contexts with weak signal-to-noise ratios.

3. Monte-Carlo simulation study

Even if the known theoretical justification (Girard 2009) only concerns settings where the correlation between observations at neighboring sites is close to 1, the simulation study 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 0.02 or less. 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.

In contrast to the case without additive white noise (Zhan and Zimmerman, 2007) the magnitude of the SNR (here the process variance b_0) has a strong impact on the results, at least from the inference point of view taken here. We present the obtained simulation results, firstly in the case of either essentially no additive white noise or “weak” noise, next, for settings with “moderate” noise, lastly, with “rather strong” noise.

Concerning the question of existence and unicity of the root, the CGEM-EV method was quite good for the settings considered here: in “almost” all the considered cases one observed a single root in the initial search interval we used (this interval was chosen quite large in all settings, typically $[0.05, 100.]$ for θ). This simulation study demonstrates that, in fact, the “almost” term we use above, is only due to settings where the 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) has no root.

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 occurred. So we chose to rather maintain a model with additive white noise. We have

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 very close to $\sqrt{2/n} = 0.047$ for any θ_0^{-1} between 0.02 and 3.

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 negligible loss). The loss in efficiency is rather small for $\theta_0^{-1} = 3$ and reasonable for $\theta_0^{-1} = 0.02$. The magnitude, in log-scale, of the range is, as expected, much less easily estimable, especially for large θ_0^{-1} . Nevertheless CGEM-EV performs also nearly as well as ML: neither the bias of $\text{Log}_{10}(\hat{\theta}_{\text{ML}})$ nor its standard deviation are significantly increased.

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} . The attractive property of the CGEM-EV estimation is that its inefficiency is rather small, 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 $\text{Log}_{10}(\theta_0)$ the performance of CGEM-EV, as compared to ML, is 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.

For $\nu = 1/2$ (i.e. exponential model) and $b_0 = 10^3$ the results concerning the estima-

tion 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 $\text{Log}_{10}(\theta_0)$ 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.

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. 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 approximation.

We finally 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 signal-to-noise 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.

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).

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.

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.

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.

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

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 \pm sd	mean \pm sd	ineff ^{1/2}
	summary for the errors $\text{Log}_{10}(\frac{\hat{\theta}}{\theta_0})$		
0.02	0.00 \pm 0.07	0.01 \pm 0.09	1.29
0.05	0.00 \pm 0.11	0.00 \pm 0.11	1.04
0.09	0.00 \pm 0.16	0.01 \pm 0.17	1.04
0.125	0.01 \pm 0.20	0.02 \pm 0.21	1.05
0.2	0.02 \pm 0.27	0.04 \pm 0.28	1.05
0.3	0.04 \pm 0.34	0.06 \pm 0.36	1.06
0.5	0.09 \pm 0.42	0.11 \pm 0.46	1.09
0.7	0.13 \pm 0.49	0.14 \pm 0.53	1.09
1.	0.17 \pm 0.55	0.18 \pm 0.60	1.09
2.	0.26 \pm 0.69	0.27 \pm 0.74	1.07
3.	0.32 \pm 0.77	0.33 \pm 0.82	1.05
	summary for the ratios $\frac{\hat{b}\hat{\theta}^{2\nu}}{b_0\theta_0^{2\nu}}$		
0.02	1.00 \pm 0.06	1.01 \pm 0.07	1.27
0.05	1.00 \pm 0.05	1.00 \pm 0.05	1.01
0.09	1.00 \pm 0.05	1.00 \pm 0.05	1.01
0.125	1.00 \pm 0.05	1.00 \pm 0.05	1.00
0.2	1.00 \pm 0.05	1.00 \pm 0.05	1.00
0.3	1.00 \pm 0.05	1.00 \pm 0.05	1.00
0.5	1.00 \pm 0.05	1.00 \pm 0.05	1.01
0.7	1.00 \pm 0.05	1.00 \pm 0.05	1.00
1.	1.00 \pm 0.05	1.00 \pm 0.05	1.00
2.	1.00 \pm 0.05	1.00 \pm 0.05	1.01
3.	1.00 \pm 0.05	1.00 \pm 0.05	1.07

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 \pm sd	mean \pm sd	ineff ^{1/2}
	summary for the errors $\text{Log}_{10}(\frac{\hat{\theta}}{\theta_0})$		
0.02	0.00 \pm 0.03	0.00* \pm 0.03*	1.13*
0.04	0.00 \pm 0.02	0.00 \pm 0.02	1.10
0.09	0.00 \pm 0.03	0.00 \pm 0.03	1.17
0.125	0.00 \pm 0.03	0.00 \pm 0.04	1.21
0.2	0.00 \pm 0.05	0.01 \pm 0.06	1.25
0.3	0.00 \pm 0.06	0.02 \pm 0.07	1.32
0.5	0.01 \pm 0.08	0.04 \pm 0.10	1.38
0.7	0.02 \pm 0.09	0.06 \pm 0.12	1.47
1.	0.02 \pm 0.11	0.08 \pm 0.15	1.50
2.	0.04 \pm 0.14	0.12 \pm 0.20	1.53
3.	0.05 \pm 0.16	0.14 \pm 0.22	1.54
	summary for the ratios $\frac{\hat{b}\hat{\theta}^{2\nu}}{b_0\theta_0^{2\nu}}$		
0.02	1.02 \pm 0.17	1.03* \pm 0.19*	1.15*
0.04	1.00 \pm 0.09	1.01 \pm 0.10	1.08
0.09	1.00 \pm 0.06	1.01 \pm 0.07	1.10
0.125	1.00 \pm 0.05	1.01 \pm 0.06	1.11
0.2	1.00 \pm 0.05	1.01 \pm 0.05	1.08
0.3	1.00 \pm 0.05	1.01 \pm 0.05	1.08
0.5	1.00 \pm 0.05	1.01 \pm 0.05	1.08
0.7	1.00 \pm 0.05	1.01 \pm 0.05	1.07
1.	1.00 \pm 0.05	1.01 \pm 0.05	1.08
2.	1.00 \pm 0.05	1.01 \pm 0.05	1.05
3.	1.00 \pm 0.05	1.01 \pm 0.05	1.07

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.

θ_0^{-1}	ML	CGEM-EV	
	mean \pm sd	mean \pm sd	ineff ^{1/2}
	summary for the errors $\text{Log}_{10}(\frac{\hat{\theta}}{\theta_0})$		
0.02	0.00 \pm 0.04	0.01* \pm 0.05*	1.18*
0.05	0.00 \pm 0.05	0.00 \pm 0.05	1.07
0.09	0.01 \pm 0.07	0.01 \pm 0.07	1.10
0.125	0.01 \pm 0.08	0.01 \pm 0.09	1.12
0.2	0.01 \pm 0.12	0.02 \pm 0.13	1.12
0.3	0.02 \pm 0.15	0.04 \pm 0.17	1.18
0.5	0.04 \pm 0.21	0.07 \pm 0.23	1.16
0.7	0.06 \pm 0.25	0.10 \pm 0.28	1.16
1.	0.08 \pm 0.30	0.13 \pm 0.32	1.14
1.5	0.12 \pm 0.33	0.17 \pm 0.38	1.17
3.	0.20 \pm 0.43	0.25 \pm 0.47	1.12
	summary for the ratios $\frac{\hat{b}\hat{\theta}^{2\nu}}{b_0\theta_0^{2\nu}}$		
0.02	1.01 \pm 0.10	1.01* \pm 0.11*	1.15*
0.05	1.00 \pm 0.06	1.00 \pm 0.07	1.02
0.09	1.00 \pm 0.06	1.00 \pm 0.06	1.02
0.125	1.00 \pm 0.06	1.00 \pm 0.06	1.02
0.2	1.00 \pm 0.06	1.00 \pm 0.06	1.02
0.3	1.00 \pm 0.05	1.00 \pm 0.06	1.02
0.5	1.00 \pm 0.05	1.00 \pm 0.06	1.01
0.7	1.00 \pm 0.05	1.00 \pm 0.06	1.02
1.	1.00 \pm 0.05	1.00 \pm 0.06	1.01
1.5	1.00 \pm 0.06	1.00 \pm 0.06	1.01
3.	1.00 \pm 0.06	1.00 \pm 0.06	1.02

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 \pm sd	mean \pm sd	ineff ^{1/2}
	summary for the errors $\text{Log}_{10}(\frac{\hat{\theta}}{\theta_0})$		
0.2	-0.00 \pm 0.03	0.01 \pm 0.06	2.09
0.3	-0.01 \pm 0.06	0.02 \pm 0.09	1.68
0.5	-0.01 \pm 0.08	0.03 \pm 0.14	1.82
0.7	0.01 \pm 0.10	0.05 \pm 0.18	1.87
1.	0.07 \pm 0.16	0.07 \pm 0.23	1.41
1.5	0.11 \pm 0.24	0.11 \pm 0.29	1.17
	summary for the ratios $\frac{\hat{b}\hat{\theta}^{2\nu}}{b_0\theta_0^{2\nu}}$		
0.2	1.00 \pm 0.07	1.02 \pm 0.08	1.11
0.3	1.00 \pm 0.07	1.03 \pm 0.08	1.19
0.5	1.00 \pm 0.07	1.02 \pm 0.08	1.11
0.7	1.00 \pm 0.07	1.02 \pm 0.07	1.07
1.	1.00 \pm 0.07	1.01 \pm 0.07	1.04
1.5	1.00 \pm 0.07	1.01 \pm 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 $b_0 = 10^3$

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

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

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

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