Efficiently estimating some common geostatistical models by 'energy–variance matching' or its randomized 'conditional–mean' versions

Didier A. Girard

Efficiently estimating some common geostatistical models by 'energy–variance matching' or its randomized 'conditional–mean' versions

Didier A. Girard
CNRS and Univ. Grenoble Alpes, Lab. LJK, Grenoble, France

Abstract

We consider the problem of fitting an isotropic zero-mean stationary Gaussian field model to (possibly noisy) observations, when the model belongs to the Matérn family with known regularity index $\nu > 0$, or to the spherical family. For estimating the correlation range (also called "decorrelation length") and the variance of the field, two simple estimating functions based on the so-called "conditional Gaussian Gibbs-energy mean" (CGEM) and the empirical variance (EV) were recently introduced. This article presents an extensive Monte Carlo simulation study for problems with around a thousand observations and settings including large, moderate, and even "small", correlation ranges. The known observation sites are either on a 2D grid (including a case of "very non-uniform" grid spacings) or randomly uniformly distributed on a simple 2D region. Some experiments for a 256×256 grid with missing values are also analyzed.

This study empirically demonstrates that, for all the (possibly random) uniform designs, the statistical efficiency of CGEM−EV compared to exact maximum likelihood (ML) is globally very satisfactory (except a degradation for the very extremal ranges in some contexts) provided the signal-to-noise ratio (SNR) is strong enough and ν is not too large, this SNR restriction being alleviated as ν decreases. For the "very non-uniform" design, a simple weighting of EV restores this efficiency. In the less favorable cases, the statistical loss remains in fact acceptable : e.g. for the largest considered index ($\nu = 3/2$) and a "not strong enough" SNR, it may happen (in fact only for large ranges) that CGEM−EV almost doubles the mean squared error for the range parameter or for the widely used combination of the two parameters known as microergodic-parameter. Furthermore an important conclusion for computational efficiency is that the use of the natural fast randomized-trace version of CGEM−EV does not significantly degrade this statistical efficiency.

keyword. Gaussian random fields, Kriging, Spatial data analysis, Covariance estimation, Maximum likelihood, Estimating functions, Matérn autocorrelation, Large scale problems, Preconditioned conjugate gradient, Randomized approximation

1 Introduction

We mainly consider the following statistical model which arises e.g. in remote sensing image analysis: 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 popular 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 $\mathbf{s}_k, k = 1, \cdots, n$, of $[0, (n_1 - 1)\delta_1] \times [0, (n_2 - 1)\delta_2]$, with an additive Gaussian 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, e.g. from previous calibration experiments (as it is common when dealing with satellite data, see Tzeng, Huang and Cressie (2005)). Using a standard lexicographic ordering, the observations thus form a vector $\boldsymbol{y}$ of size n whose law is Gaussian :

$$\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, \tau_0^2 \boldsymbol{R}_{\theta_0} + \sigma_{\mathrm{N}}^2 \boldsymbol{I}_n) \tag{1.1}$$

with $\boldsymbol{I}_n$ denoting the identity matrix and $\boldsymbol{R}_\theta$ the autocorrelation matrix of the gridded process i.e. the block Toeplitz matrix (with n_1^2 Toeplitz square blocks, each of size $n_2 \times n_2$) whose coefficients are given by

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

$||\cdot||$ being the Euclidean norm and $\rho_{\nu,\theta}$ the Matérn autocorrelation function

$$\rho_{\nu,\theta}(x) = \frac{(\theta x)^\nu}{\Gamma(\nu) 2^{\nu-1}} K_\nu(\theta x), \quad x > 0, \ \theta > 0,$$

where K_ν is the modified Bessel function of the second kind of order $\nu > 0$. For more details on these widely used autocorrelation functions see Guttorp and Gneiting (2006). Note that

$$\tau_0^2 = \mathbb{E}((Z(\mathbf{s}))^2) \equiv \mathrm{E}(y_k^2) - \sigma_{\mathrm{N}}^2$$

will be called the process (or signal) variance. When mentioned, we will also consider another well known autocorrelation function, namely the spherical model

ρ_θ^{S}. See e.g. Zhang and Zimmerman (2007) for these definitions. Notice that a significant variant of the above uniform grid, that we call "a very nonuniform Cartesian grid", will be analysed with some details. We also study a case of $n = 1000$ observation sites randomly but uniformly distributed on a simple 2D region. And, to illustrate the "scalability" of the proposed method, we will also consider, albeit with less extensive simulations, a much larger 256×256 grid with a few missing regions.

The order ν, often called the regularity (or differentiability) index, is assumed to be known in this paper. Recall that $\rho_{1/2,\theta}(x) = \exp(-\theta x)$ is the very popular exponential model, and that simple expressions also exist for $\rho_{\nu,\theta}(x)$ for $\nu = 3/2$ and $5/2$: these ν's correspond to models also often used (see e.g. Stein (1999), Rasmussen and Williams (2006)). In the Monte-Carlo simulation study of this paper, we only consider three contexts: the order ν will be either $1/6$, $1/2$ or $3/2$.

The parameter θ^{-1} is often called the "decorrelation length" or "the range parameter".

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

It is generally 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 a simple variogram fitting classically known as "the weighted least-squares 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^2_{N} (the solution, say $\hat{\tau}^2_{\mathrm{ML}}(\theta)$, being typically obtained iteratively, e.g. by Fisher scoring, even if σ^2_{N} (> 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 $\hat{\tau}^2_{\mathrm{ML}}(\theta_1)\theta_1^{2\nu}$ still remains an efficient estimator of $\tau_0^2\theta_0^{2\nu}$ which is the so-called microergodic parameter (see Du, Zhang and Mandrekar (2009) and Wang and Loh (2011) for recent results of this type in the case without additive white noise).

The method that is proposed in Girard (2011), firstly reverses the roles of variance and range-parameter 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 following "bias corrected" empirical variance $\hat{\tau}^2_{\mathrm{EV}}$ (we discuss in Section 2.1 that, in the present study, the probability that $n^{-1}\boldsymbol{y}^T\boldsymbol{y} < \sigma^2_{\mathrm{N}}$ will always be very low so that one can tolerate this simple, not "fully complete", definition for $\hat{\tau}_{\mathrm{EV}}$), which in

turn yields $\hat{b}_{\mathrm{EV}}$ as signal-to-noise (SNR) estimate:

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

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 $\hat{b}_{\mathrm{EV}}$

$$\boldsymbol{y}^T\boldsymbol{A}_{b,\theta}\left(\boldsymbol{I}_n - \boldsymbol{A}_{b,\theta}\right)\boldsymbol{y} = \sigma^2_{\mathrm{N}}\ \mathrm{tr}\boldsymbol{A}_{b,\theta} \qquad \text{where} \quad \boldsymbol{A}_{b,\theta} := b\boldsymbol{R}_\theta\left(\boldsymbol{I}_n + b\boldsymbol{R}_\theta\right)^{-1}. \tag{1.3}$$

Note that the equation $\boldsymbol{y}^T\boldsymbol{A}_{b,\theta}{\boldsymbol{R}_\theta}^{-1}\boldsymbol{A}_{b,\theta}\boldsymbol{y} = \sigma^2_{\mathrm{N}}\, b\ \mathrm{tr}\boldsymbol{A}_{b,\theta}$, which is equivalent to (1.3) (this is easily seen after simple algebra), can be numerically more stable in case of large SNR since the "smoother" $\boldsymbol{A}_{b,\theta}$ then comes close to the identity. In fact when σ^2_{N} is very small compared to τ^2_0, one can use simply

$$\hat{\tau}^2_{\mathrm{EV}} := n^{-1}\boldsymbol{y}^T\boldsymbol{y} \quad \text{and} \quad n^{-1}\boldsymbol{y}^T{\boldsymbol{R}_\theta}^{-1}\boldsymbol{y} = \hat{\tau}^2_{\mathrm{EV}} \tag{1.4}$$

in place of (1.2) and (1.3). This equation (1.3) in θ with b given by (1.2) is called the "conditional Gibbs-energy mean and empirical variance"- based estimating equation (CGEM−EV equation) in Girard (2011) (see Girard (2016) for the "$\sigma^2_{\mathrm{N}} = 0$" version (1.4) for which a simple justification in the "infill asymptotics" regime can be deduced from a result of Kaufman and Shaby (2013)) which gives details, heuristic justifications and a large-n-small-δ justification for the one dimensional "time series" analog setting, stating that an asymptotic full-efficiency (as compared to ML) is reached as the sampling step δ decreases to 0. This large-n-small-δ full-efficiency requires that ν stays "close" to 1/2 when it assesses the error in either the range parameter or the variance parameter, but it holds for any ν when it concerns the microergodic parameter.

In the case of a nonuniform grid for the locations of the n observations, simple weighted versions of the average $\boldsymbol{y}^T\boldsymbol{y}/n$ in (1.2) which are motivated by a Riemann-sum type discretization of $\int_\Omega Z^2(\mathbf{s})\, d\mathbf{s}/\int_\Omega d\mathbf{s}$ where Ω is a simple domain containing the data locations, are suggested in Girard (2011). An example of such estimate of τ^2_0 used in place of $\hat{\tau}^2_{\mathrm{EV}}$ (and denoted $\hat{\tau}^2_{\mathrm{wEV}}$) is detailled Section 2.5. Heuristics (mainly a "minimum variance property" given for this example) and the simulation results of Section 3.7 will give some support to this Riemann-sum approximation approach.

This article is structured as follows. We first give some comments in Section 2 which supplement those in Girard (2011), notably about the computational gains which could be expected for CGEM−EV as compared to ML or to the randomized-traces version of the two classical likelihood equations which may be thought of as a computationally efficient alternative along the lines of the recent study by Stein, Chen and Anitescu (2013). This article then 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 $\mathrm{tr}\boldsymbol{A}_{b,\theta}$, is globally very satisfactory (there is a noticeable degradation in efficiency only for extremal ranges in some contexts) provided the signal-to-noise ratio b_0 is strong enough and ν is not too large, with SNR's restriction depending on ν. The meaning that we give here of "strong enough" needs to be clarified. For example when ν is known to be 1/6, a SNR as less as 4 (that we call, somewhat arbitrarily, a "rather weak" SNR) is sufficient so that CGEM−EV and ML can be compared (actually they are seen to be quite close to each other), but for $\nu = 3/2$ both method are unusable under such SNR; this is discussed in Section 3 and we draw several conclusions Section 4. When invoking a "fast randomized-trace", the meaning of "fast" in this paper is that one applies the linear operator $\boldsymbol{A}_{b,\theta}$ to a random vector of size $n \times 1$, only a few times (typically $n_{\mathrm{R}} \leq 20$ with the notation introduced Section 2.2). If the observation grid is non regular and very nonuniform, this efficiency may be degraded; for a such simple case, the Riemann-sum version of $\hat{\tau}^2_{\mathrm{EV}}$ is demonstrated to be able to restore this efficiency. However, such a modification is not always required; indeed for the case of a random but uniform design, a good news, and somewhat surprising, is that, for very various range-parameter values, the unweighted version of CGEM−EV is still quite statistically-efficient. For the experiments with the 256×256 incomplete grid, since a complete comparison with the approximate ML discussed Section 2.3 would have been quite difficult, we only refer to the classical Cramer-Rao lower bounds for unbiased estimators of θ_0 or of the micro-ergodic parameter. The comparisons confirm the high level of the statistical efficiency of CGEM−EV.

2 Some comments on randomized CGEM-EV or its weighted EV version, and their computational advantages

2.1. A first comment is in order about the bias corrected empirical variance $\hat{b}_{\mathrm{EV}}$ 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 $\hat{b}_{\mathrm{EV}}$ has a negative value, in case of which (1.3) has no solution. However, it is easy to see that the probability of a such pathological event tends to zero when the observation domain increases infinitly, possibly with an "infill component" (see e.g. Lemma A.1

of Lahiri, Lee and Cressie (2002)); and 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 ratio. A felling about the meaning of "large enough" can be obtained by examining the list of the probabilistic settings considered Section 3, since a negative value for $\hat{b}_{\mathrm{EV}}$ was never observed in the simulation study (1000 replicates for each setting).

2.2. The motivation behind our work is that CGEM−EV, when using fast randomized-trace approximations, could enjoy a computationally efficient implementation for very large scale problems for which the exact ML method has a prohibitive computational cost (CPU time or memory size). 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, Stroud, Stein and Lysen (2016) and the references therein for 2D or 3D problems). Once a fast solver is available to compute $\boldsymbol{A}_{b,\theta}\boldsymbol{y}$, but does not form explicitly the matrix $\boldsymbol{A}_{b,\theta}$, one is tempted, since $\boldsymbol{A}_{b,\theta}$ may be seen as an instance of "influence matrices involved in regularization of linear equations or data smoothing problems" (cf Girard (1989)), to try a randomized-trace approximation (i.e. generate independent $\boldsymbol{w}_r \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n), r = 1, \cdots, n_{\mathrm{R}}$, and use $(1/n_{\mathrm{R}})\sum_{r=1}^{n_{\mathrm{R}}} \boldsymbol{w}_r^T \boldsymbol{A}_{b,\theta}\boldsymbol{w}_r/(\boldsymbol{w}_r^T\boldsymbol{w}_r)$ in place of the exact $(1/n)\mathrm{tr}\boldsymbol{A}_{b,\theta}$) when solving (1.3) by e.g. a bissection search.

In this article we do not attempt to compare the various possible PCG solvers for problems like those of the following simulations. In the following the contexts where *exact* ML estimates are also simulated are restricted to relatively moderate data size (around 1000) so that using such iterative solvers was not required. Classical "exact" Cholesky or eigenvalues-eigenvectors decompositions were then used in these cases, even to simulate randomized-trace versions of the CGEM−EV estimating equation.

Note that, in the following simulation study, we always chose the "reuse option" (as in Girard (1989)) when computing a randomized-trace version of the CGEM−EV criterion at different tried values of θ; it means that the n_{R} simulated $\boldsymbol{w}_r$'s (or $\boldsymbol{u}_r$'s, see Section 2.3) are kept constant during the processing of each simulated data set. Each required numerical root search can then be reliably done by standard routines; and thus the study of the impact of the size n_{R} is easier.

Recall that for lattice observations, exact ML computations require about $n^{5/2}$ operations (Zimmerman, 1989). Thus as long as the number of iterations required in each product $\boldsymbol{A}_{b,\theta}\boldsymbol{y}$ (or $\boldsymbol{A}_{b,\theta}\boldsymbol{w}$) remains reasonable, the computational gain (time

savings) offered by a fast randomized version of CGEM−EV is a factor approximately equal to $n^{3/2}/\log n$. It is well known today that an approximation of ML which also attains such a gain is the now classical tapered Whittle ML (see Guyon (1982) and Dahlhaus and Künsch (1987)). But this approach has "an element of arbitrariness in implementation (specifically, the choice of a taper) in order to cope with the edge effect" as said by Robinson (2008). Furthermore, even with the best taper, the estimation error can still be much less satisfactory (especially in terms of bias) than ML. In the very recent years Stein, Chen and Anitescu (2013) studied randomized-trace versions of the score equations (i.e. those obtained by setting to 0 the gradient of the likelihood function and using randomized-trace approximations). Is is demonstrated by these authors that this can really produce near-efficient estimates at a cost close to $n \log n$ in some settings. It is clear that "the estimate converge to the true score function as the Monte Carlo sample sizes goes (n_{R}) to infinity" as said in Stroud, Stein and Lysen (2016), but how fast is this convergence is still not clear, especially for contexts with large (or even moderate) correlation range. We analyze in details a "typical" example in the following section.

2.3. Computational advantages over the method of randomized score equations.

2.3.1. We consider in this section a median setting of Matérn field with regularity index $\nu = 3/2$. This setting is called "median" in the sense that its parameters are median among those of the extensive study of Section 3. More precisely we chose a range-parameter $\theta_0^{-1}\sqrt{3} = 0.3$ (see Figure 5) and a SNR, $b_0 = 30000$, which is an intermediate value between the strong SNR of Section 3.2 and the "low" SNR of Section 3.4. We simulated 5 realizations of such a random field. More precisely, we generated 5 datasets of size $n_1 \times n_1$ with $n_1 = 48$ and $\delta_1 = \delta_2 = 1/n_1$. This size still permits the use of exact covariance-matrix decomposition and, above all, an accurate examination of the performance of the randomized-trace versions when using large Monte-Carlo sample sizes. Notice that Kaufman and Shaby (2013) consider settings quite similar to this one, even though they use exact observations (while here we add a "small" white noise of relative magnitude of about 0.5%, as compared to the standard deviation of the Matérn field).

For each of these 5 $\boldsymbol{y}$'s, consider the two score equations $S_1(b, \theta) = 0$ and $S_2(b, \theta) = 0$ obtained by setting to 0 the derivative of the likelihood function respectively w.r.t. b and θ. In fact to have more readable plots (and unconstrained optimizations) we instead consider the two arguments $\tilde{b} = \log_{10}(b)$ and $\tilde{\theta} = \log_{10}(\theta)$. Notice that it could be easily checked that $S_1(b, \theta) = 0$ is equivalent to (1.3) (see the heuristical justification of CGEM in Girard (2011) for an interpretation of this property: indeed $S_1(b, \theta) = 0$ is the constraint which "adjusts" $\hat{\theta}_{\mathrm{CGEM}}(b)$ to a given (well chosen) b and complementing (1.3) with (1.2) is our proposal CGEM−EV,

although other estimates of b_0, like its weighted version, could be used).

A now well known (see e.g. Section 6.4 of Stein (1999) or Kaufman and Shaby (2013)) attractive property of the Matérn model (which is not always shared by the spherical model) is that the maximization of the likelihood has typically a unique local solution, which is thus global (although there is not yet a theoretical statement about this, above the case $\nu = 1/2$ in one-dimension (i.e. case of a AR1 series) as far as we know). So it was not surprising to observe that for each $\boldsymbol{y}$, for any fixed $\tilde{\theta}$, arbitrarily chosen in a rather large domain, numerically solving the first score equation w.r.t. $\tilde{b}$ has a unique solution, that we denote by $\log_{10}(\hat{b}_{\mathrm{CML}}(\tilde{\theta}))$ (the C stands for "constrained by $\tilde{\theta}$ fixed"); and for any fixed $\tilde{b}$ in a large domain, numerically solving the second score equation w.r.t. $\tilde{\theta}$ has a unique solution, that we denote by $\log_{10}(\hat{\theta}_{\mathrm{CML}}(\tilde{b}))$. Furthermore, well in agreement with this strict unimodality property, we also consistently observed that the two parameterized curves $\mathcal{C}_1 : \tilde{\theta} \mapsto (\log_{10}(\hat{b}_{\mathrm{CML}}(\tilde{\theta})), \tilde{\theta})$ and $\mathcal{C}_2 : \tilde{b} \mapsto (\tilde{b}, \log_{10}(\hat{\theta}_{\mathrm{CML}}(\tilde{b})))$ have well a unique intersection point which is the ML-estimator.

However it is clear in Figure 1 that, for each one of the 5 $\boldsymbol{y}$'s, the corresponding two curves $\mathcal{C}_1, \mathcal{C}_2$ (which were computed on a grid of candidate $\tilde{\theta}$ and a grid of $\tilde{b}$ respectively, and plotted, both with a same color for each $\boldsymbol{y}$, $\mathcal{C}_1$ being the continuous curve, $\mathcal{C}_2$ being in dotted style) appear to coincide along a whole half-line with implicit equation "$\tilde{b} + 2\nu\tilde{\theta} = \text{constant}$", and that this constant seems rather invariant from one $\boldsymbol{y}$ to another one. Actually, these two observations are in perfect agreement with the well known fact that "the likelihood function can have long ridges along which it is nearly constant" (p. 173 of Stein (1999)) and the further results by Zhang (2004) andDu, Zhang and Mandrekar (2009) and on the estimation of the micro-ergodic parameter $b_0\theta_0^{2\nu}$. And thus the 5 unique intersection points can hardly be distinguished on this half-line in Figure 1. A simple way to remedy this is to replace the $\tilde{b}, \tilde{\theta}$ plot by using the affine transformation $(\tilde{b}, \tilde{\theta}) \mapsto (\tilde{b}, \tilde{b} + 2\nu\tilde{\theta})$, as we will do for studying the impact of the randomized-trace approximation on the estimation of the micro-ergodic parameter. However, before to do this, let us discuss the estimation of θ_0 and b_0. The reasonable spread of the 5 replicates of $\log_{10}(\hat{\theta}_{\mathrm{ML}})$ (horizontal dashed arrows) in Figure 1 is in agreement with the extensive simulations of Kaufman and Shaby (2013) where it is concluded that "It is apparent in all cases that the data does indeed contain information about the range parameter". Now, recall that the construction of $\hat{\theta}_{\mathrm{CGEM-EV}}$ simply consists in intersecting the horizontal line $\tilde{b} = \log_{10}(\hat{b}_{\mathrm{EV}})$ with $\mathcal{C}_1$ in a $\tilde{b}, \tilde{\theta}$ plot. The first good news demonstrated in Figure 1 is that the $\log_{10}(\hat{\theta}_{\mathrm{CGEM-EV}})$'s (horizontal continuous arrows deduced from the 5 vertical continuous arrows) have a variability around $\log_{10}(\theta_0)$ very similar to that of the *exact* ML estimates. And a similar conclusion holds for the related estimates of $\log_{10}(b_0)$ (vertical dashed arrows for ML and vertical continuous arrows for EV). Nevertheless the extensive study of Sec-

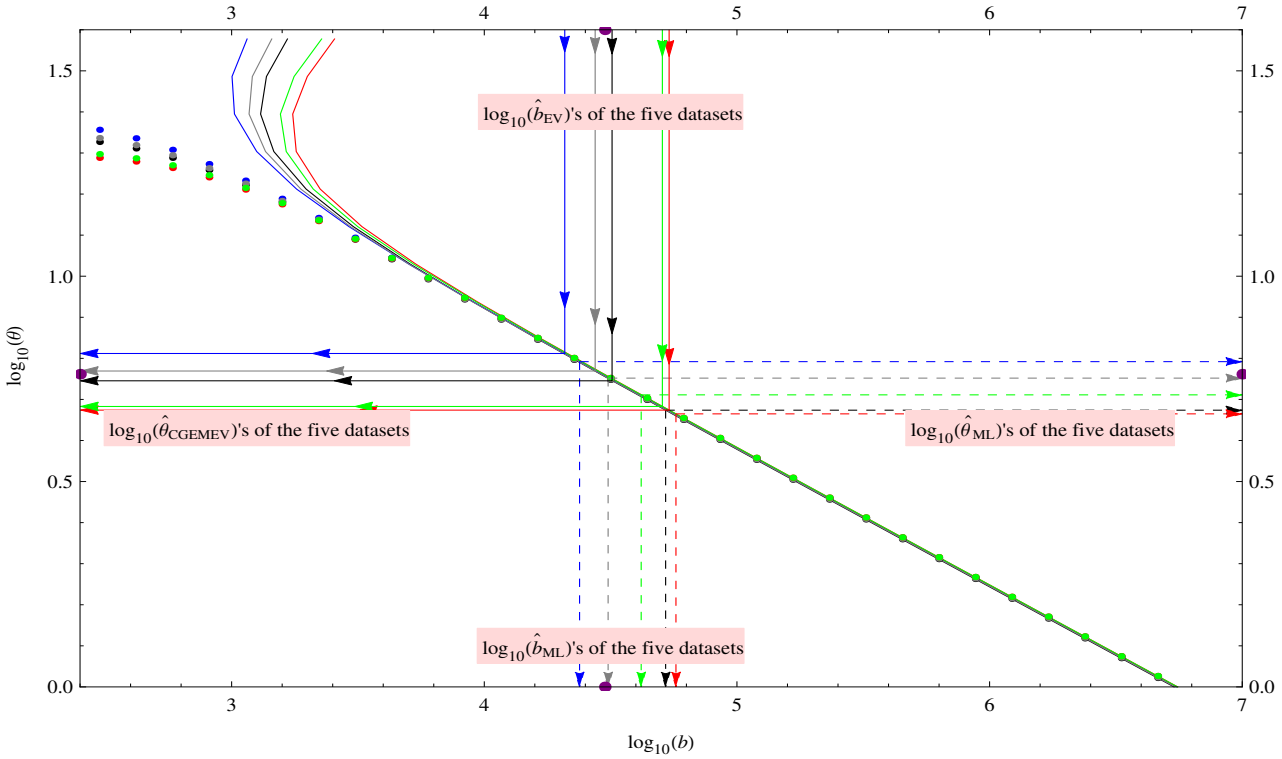

Figure 1: $n = 48 \times 48$. Results for 5 datasets from a Matérn model ρ_{ν,θ_0} with $\nu = 3/2$ and median range (see Figure 5 for this autocorrelation) and a SNR = 30000. For example for the fourth $\boldsymbol{y}$, 2 curves, 2 horizontal half-lines and 2 vertical half-lines are all plotted in green: the curve $\mathcal{C}_1$ (in continuous style) implicitly defined by the first score equation, the curve $\mathcal{C}_2$ (in dotted style) defined by the second score equation, the vertical continuous line is $\hat{b}_{\mathrm{EV}}$, the horizontal continuous line is $\hat{\theta}_{\mathrm{CGEM-EV}}$, similarly in dashed style for $\hat{b}_{\mathrm{ML}}$ and $\hat{\theta}_{\mathrm{ML}}$. The true parameter values are marked by (purple) half-circles on the frame of the figure

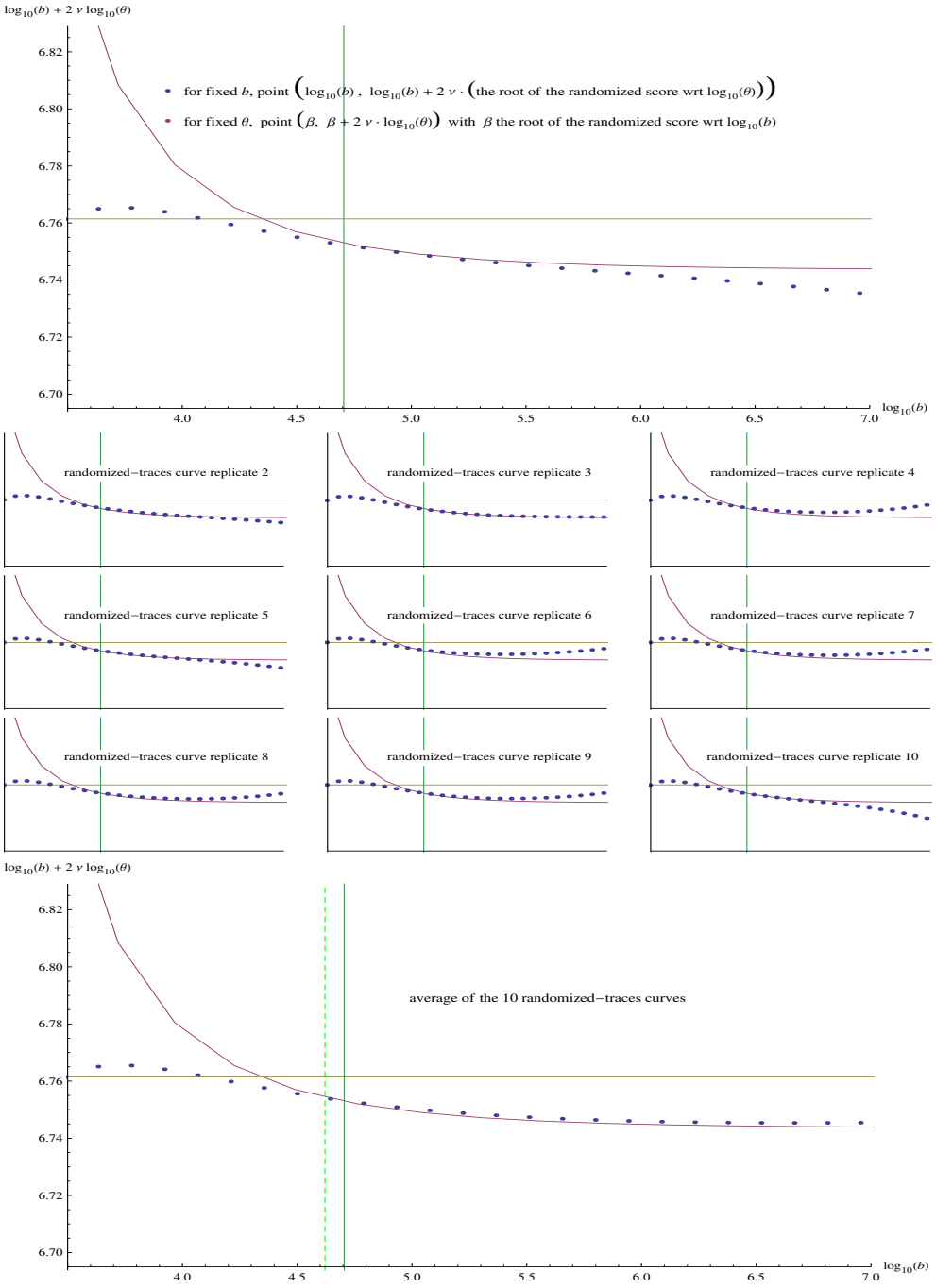

Figure 2: Setting identical to Figure 1. Results for the fourth $\boldsymbol{y}$. The first ten plots are the results of using ten different seeds in the computation of the randomized-trace (using the Hadamard-matrix based dependent sampling, with $n_{\mathrm{R}} = 256$) version of $\mathcal{C}_1$ and $\mathcal{C}_2$. The 11th plot (bottom) is equivalent to using $n_{\mathrm{R}} = 2560$. The vertical dashed line indicates the exact $\hat{b}_{\mathrm{ML}}$, the vertical continuous line is $\hat{b}_{\mathrm{EV}}$. The horizontal line is the true $b_0\theta_0^3$

tion 3 (and theoretical results as $n \to \infty$) state that CGEM−EV is not rigorously fully-efficient as compared to ML.

Now we are going to study the impact of using randomized-traces instead of exact traces in both methods. Our first finding is that, for this "median" setting, choosing $n_{\mathrm{R}} = 1$ is sufficient for CGEM−EV while even a n_{R} of several hundreds may still produce a monte-carlo error in the randomized score approach which eventually leads to an estimate of θ_0 much worse than CGEM−EV. We only present the analysis of one example of data set and $n_{\mathrm{R}} = 256$, but a similar behavior, namely an issue of "serious unstability w.r.t. the Monte Carlo noise", of "the" root of the pair of randomized scores was consistently observed on other datasets similarly replicated. To be fair, we used in this analysis the modification of the classical randomized trace estimator (recalled in Section 2.2) which is advocated in Stein, Chen and Anitescu (2013) when $n_{\mathrm{R}} > 1$. These authors replace each $\boldsymbol{w}_r$ by a vector $\boldsymbol{u}_r$ of iid centered Bernouilli variables and, above all, use n_{R} vectors $\boldsymbol{u}_r$'s drawn in a dependent way following a sampling constructed from the classical $n_{\mathrm{R}} \times n_{\mathrm{R}}$ Hadamard matrix. They indeed show that, for n_{R} fixed, this dependent sampling can be a significant improvement over the independent sampling of the vectors $\boldsymbol{u}_r$'s. The dataset $\boldsymbol{y}$ used in the following is the one which gave the curves in green in Figure 1. We chose this one because the exact ML is seen to actually improve over exact CGEM−EV (this is not always the case since the loss in efficiency is rather low). In Figure 2, all the plots are $(\tilde{b}, \tilde{b} + 2\nu\tilde{\theta})$-plots which are produced from this dataset. The first 10 plots are the results of using 10 different seeds in the computations of the randomized-traces. Now the mentioned serious "unstability issue" is clearly seen as the combination of the intrinsic "ill"-conditioning in the solve of the exact score equations (even on a such $(\tilde{b}, \tilde{b} + 2\nu\tilde{\theta})$-plot, the exact $\mathcal{C}_1$ and $\mathcal{C}_2$ would still be nearly coincident over a large domain of $\tilde{b}$) and of the perturbation of the scores due to using a finite n_{R}: this combination creates a dramatic perturbation of the root. Indeed in 2 cases among these 10 (the first and the last ones), the 2 curves seem to become nearly tangential but they actually have no intersection; with the 2nd and the 5th seeds there are clearly at least 2 roots and for the seeds 8, 9 and particularly 3, the 2 curves really become tangential and a unique root is hardly distinguishable in such a plot. Finally we averaged the 10 randomized-traces functions previously obtained : this gave the 11th pair of randomized scores (which may be thought of as a randomized approximation of the scores using $n_{\mathrm{R}} = 2560$) displayed in the bottom panel of Figure 2; now a unique intersection point clearly appears, however it is not really an improvement over CGEM−EV; in fact the obtained $\hat{b}$ appears to be much closer to $\hat{b}_{\mathrm{EV}}$ than the exact $\hat{b}_{\mathrm{ML}}$ (vertical line in dashed green).

Notice that it is also seen in Figure 2 that the "perturbation of the scores due to using a finite n_{R}" mentioned above actually concerns essentially only the curve

$\mathcal{C}_2$. In fact we observed that even with $n_{\mathrm{R}} = 1$ the curves "randomized $\mathcal{C}_1$" are not visually perturbed in plots like Figure 2 by the randomization error whereas these plots have a resolution sufficiently accurate to display the variability (from one $\boldsymbol{y}$ to another) of the estimates of the micro-ergodic parameter $\tilde{b}_0 + 2\nu\tilde{\theta}_0$. The extensive Monte-Carlo of Section 3 will confirm that this "finding" concerning CGEM−EV with $n_R = 1$ typically holds, although using $n_R = 20$ may be sometimes useful, in the sense that it provides a significantly better statistical accuracy, for settings where the true equivalent range is much greater than this median value 0.3 and the SNR is weaker.

2.3.2. A natural question is now: since increasing n_{R} in the randomized-scores approach would eventually converge toward the exact ML, whereas CGEM−EV cannot exactly reach the statistical accuracy of ML, what is the computational gain offered by our proposal. Beforehand, one have to attack the "no or several root(s)" issue that is demonstrated above.

Of course a sensible remedy to attack the "no or several root(s)" issue, is to re-start the whole numerical solve (that is, an unconstrained root-finding in $\mathbb{R}^2$) of the two randomized score functions, using a new set of n_{R} pseudo-random $\boldsymbol{u}$'s at each re-start, until a reasonable root be reached. This approach will be called a randomized-scores solve using "possibly re-started" (PRS randomized-scores or PRS-RS for short) algorithm. Actually the algorithm used here was the one from the R-package nleqslv by Hasselman (2016) with classical options; precisely : method = "Broyden", global="dbldog". Notice that for CGEM-EV we used the classical uniroot R-function since a scalar-root solver is sufficient. Note that in the case $\sigma_{\mathrm{N}} = 0$ the 2 score functions, even randomized, can be concentrated in one function (precisely, as is well known, the unkwown b can be eliminated) and the simpler uniroot can also be applied. However this does not eliminate the "no or several root(s)" issue. More importantly, neglecting the presence of a measurement error, even small, can have a strong negative impact on the ML appoach, even exactly implemented; this is well demonstrated by Girard (2014) for a particular case ($\nu = 1/2$). And indeed, a simulation study (not reported here) for the above probabilistic setting where the 0.5% noise was not taken in account also demonstrated such a negative impact (essentially a rather strong bias is observed).

Concerning the possible strategies to implement randomized score equations, we observed that, in fact, it is generally advantageous to apply, to a given $\boldsymbol{y}$, a such PRS-RS solve algorithm several times (the number of applications will be denoted by m), again with new n_{R} pseudo-random $\boldsymbol{u}$'s each time, and to retain the median of the so-obtained m roots (actually the median of the θ-component of the m roots), instead of increasing n_{R} by a factor of m; for example we have used this approach with m equal to 3, 5 or 7. Indeed this takes rather well in account

that, when there exists too roots one of them is often an "outlier". We call this estimator the "median after m PRS-RS solves" estimator (M−PRS−RS estimator for short). As another natural "robustifying" strategy, it could have been proposed to replace the mean of the n_{R} primary trace-estimates by their median, but this could cause discontinuities in the score functions, whereas algorithms like nleqslv are more reliable when applied to smooth functions.

A precise "computational gain" question may now be formulated: when the evaluation, for a trial value of (b, θ), of either the difference of the two sides of the randomized-trace version of the CGEM−EV estimating equation (1.3) or the 2 randomized scores, is assumed to have a computational cost proportional to $1+n_R$ (the proportionality constant being, typically, the very large cost of solving one $n \times n$ linear system), how many computational gain is there in CGEM−EV (where n_{R} is fixed to 1), compared with a M−PRS−RS estimator which would be the fastest among the M−PRS−RS estimators which have a similar statistical accuracy than the one of CGEM−EV?

Recall that this proportionality assumption is rather well fulfilled when an iterative linear solver, like PCG, is used, even if many of the invoked linear systems only differ by their second member.

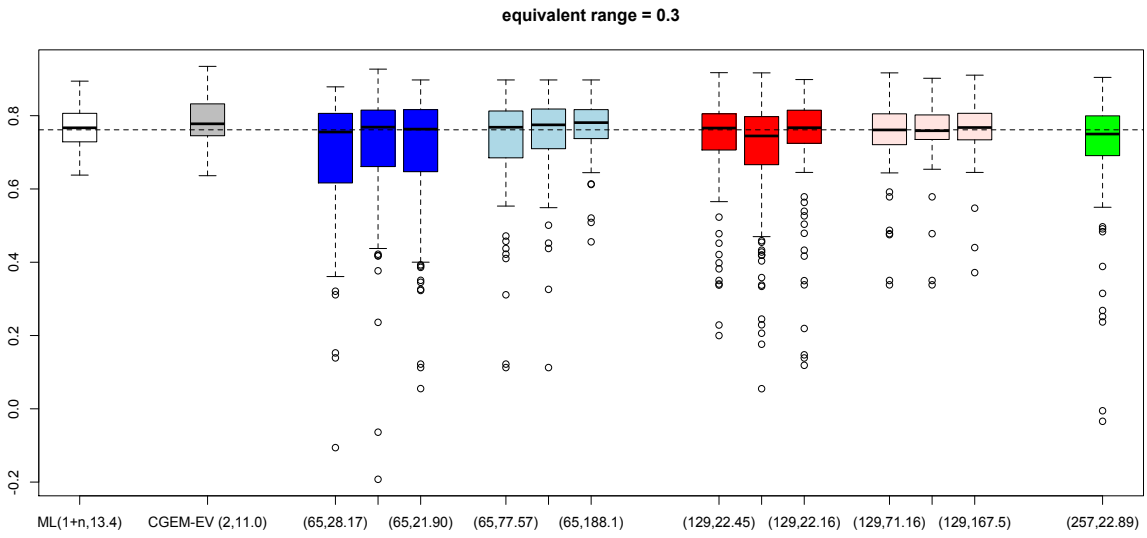

Figure 3: Setting identical to Fig.1. Box-plots, in white and grey, of the $\log_{10}(\hat{\theta}_{\mathrm{X}})$'s for method X = ML, CGEM−EV, next in blue for X = M−PRS−RS with $(n_{\mathrm{R}}, m) = (64,1), (64,1), (64,1), (64,3), (64,5), (64,7)$, next in red with $(n_{\mathrm{R}}, m) = (128,1), (128,1), (128,1), (128,3), (128,5), (128,7)$, next in green with $(n_{\mathrm{R}}, m) = (256,1)$

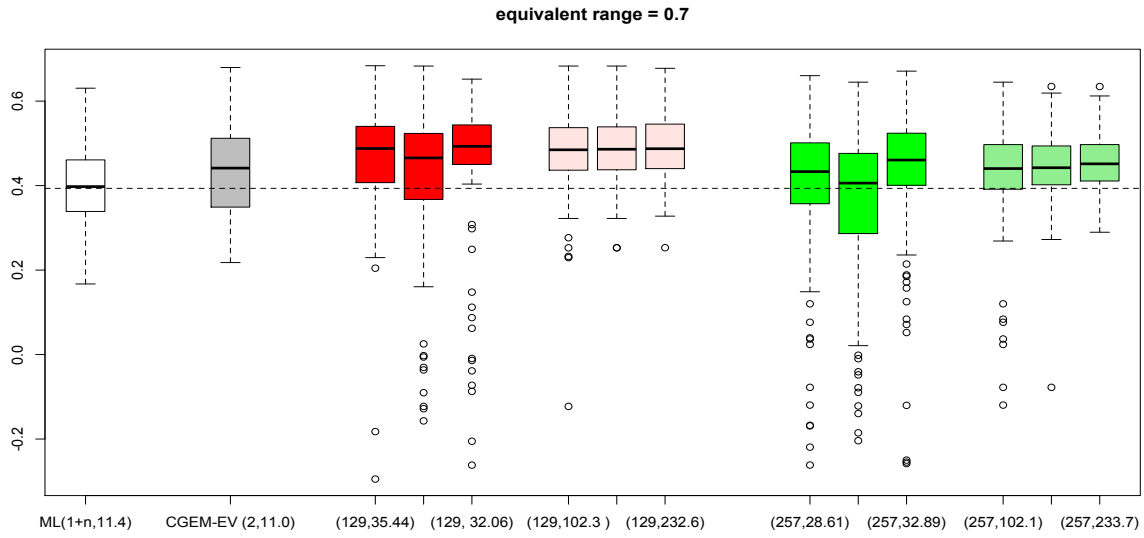

Figure 4: Idem Figure 3 except that $\theta_0^{-1}\sqrt{3} = 0.7$. The 6 box-plots in red (respectively in green) are for X = M−PRS−RS and $(n_{\mathrm{R}}, m) = (128,1),(128,1),(128,1),(128,3),(128,5),(128,7)$ (respectively $(n_{\mathrm{R}}, m) = (256,1),(256,1),(256,1),(256,3),(256,5),(256,7)$)

In this paper we only assess this gain in the probabilistic setting studied above in this Section, firstly with $\theta_0^{-1}\sqrt{3} = 0.3$, then with different θ_0, by a Monte-Carlo simulation study of size 100 for each setting. Here, instead of determining the mentioned "fastest M−PRS−RS" (which would require an exhaustive examination of all the couples (n_{R}, m)), we simply tried $n_{\mathrm{R}} = 64$, next 128, possibly next 256; and we considered $m = 1$ (no robustification) then $m = 3, 5, 7$.

The results are sumarized Figure 3 and Figure 4, with range-parameter $\theta_0^{-1}\sqrt{3} = 0.3$ and $\theta_0^{-1}\sqrt{3} = 0.7$ respectively (in each figure, the dashed line indicates $\log_{10}(\theta_0)$). From the mentioned "proportionality assumption" and defining the unit of computational costs as the cost of solving one $n \times n$ linear system, the cost can thus be resumed as the product of $1+n_{\mathrm{R}}$ times the average (over the 100 $\boldsymbol{y}$'s) of the number of function evaluations required either by uniroot or nleqslv. Note that for exact ML, $1+n_{\mathrm{R}}$ can be replaced by $1+n$ since the n vectors of the canonical basis of $\mathbb{R}^n$, in place of the $\boldsymbol{w}$'s, give the exact trace, in the absence of a fastest algorithm. For each estimator, these 2 numbers are displayed at the abscissae of the corresponding box-plot. Let us first discuss Figure 3. We observe (and this was not apparent in Figure 1) that the $\log_{10}(\hat{\theta}_{\mathrm{CGEM-EV}})$'s (second box, in grey) are slightly less good than the the *exact* ML estimates. An important result in Figure 3 is that, among the colored box-plots, the sixth (corresponding to $(n_{\mathrm{R}}, m) = (64, 7)$) appears to

have a spread similar to the grey box-plot and only the box-plots associated to $(n_{\mathrm{R}}, m) = (128, 5)$ or $(n_{\mathrm{R}}, m) = (128, 7)$ are also similar, the other ones being more (sometimes much more) spreaded. By examining the abscissae of the grey box-plot and the one of this sixth colored box-plot, we deduce that CGEM−EV thus offers a computational gain of about $(65 \times 188.1)/(2 \times 11.0)$, that is about 5.5 hundreds.

For the case $\theta_0^{-1}\sqrt{3} = 0.7$, we see Figure 4 that the degradation (in both standard deviation and in bias) of $\log_{10}(\hat{\theta}_{\mathrm{CGEM-EV}})$ as compared to exact ML, is now larger than in the previous setting. Let us remark that a non-negligible bias is also apparent in the M−PRS−RS estimtors, even with $(n_{\mathrm{R}}, m) = (256, 7)$ (we presently have no clear explanation of this bias). Also, we had to consider greater values for $n_{\mathrm{R}} \times m$, as compared to the case $\theta_0^{-1}\sqrt{3} = 0.3$, so that this bias decreases. Now, an important finding of this study, is that CGEM−EV, compared to similarly accurate M−PRS−RS, again offers a computational gain of about several hundreds (this can deduced by an examination of the couples ($1 + n_{\mathrm{R}}$, number of function evaluations required by nleqslv) displayed at the x-axe of Figure 4).

We have also done simulation studies of probabilistic settings with smaller correlation range. A general tendency which has been found from these studies, is that it becomes easier to spot the so-called "bad roots". Thus with a more easily tuned PRS strategy, a n_{R} as low as 16 or 32 can now suffice to guaranty a statistical accuracy near the one of ML (this is agreement with the results of Stein, Chen and Anitescu (2013)). On the other hand, CGEM-EV also becomes closer to exact ML as the range decreases (the extensive simulation studies of Section 3 make this more precise). Thus the computational gains remain, rougthly, at least of a factor ten.

2.4. Although the available theoretical results for CGEM−EV are todays restricted to regularly spaced locations, we are of course 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 "$\boldsymbol{A}_{b,\theta}$" matrix (see Fritz, Neuweiler and Nowak (2009) and Stroud, Stein and Lysen (2016) for recent works on this subject). One finding of our Monte-Carlo experiments is that the unmodified CGEM−EV still works very well when the missing data are simply those located inside a small number of simple regions, like 5 disks (see Section 3.9).

2.5. For the cases of "very" non-uniform locations for the observations, we suggest in this section that weights which take this into account can really improve upon

the equal weights of $\boldsymbol{y}^T\boldsymbol{y}$. 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:

$$\mathcal{S}_1 := \{x_1, \cdots, x_{n_1}\} = \{0.02, 0.03, \cdots, 0.20\} \cup \{0.25, 0.35, \cdots, 0.95\},$$

thus $n_1 = 19 + 8$ 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 weighted version of $\hat{\tau}^2_{\mathrm{EV}}$ will be thus defined (noticing that $\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} w_{i,j} = 1$ here) by

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

The associated version of CGEM−EV (i.e. solve (1.3) with b fixed at $\hat{b}_{\mathrm{wEV}} := \hat{\tau}^2_{\mathrm{wEV}}/\sigma^2_{\mathrm{N}}$) will be then 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 the spatial sampling of a (at least) continuous (in a mean square sense) field Z, it is intuitive (at least in the case of weak additive white noise) that the ordinary spatial average $\boldsymbol{y}^T\boldsymbol{y}/n$ should be modified so that 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 first 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^2_0 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 statements of Shin and Song (2000) (which generalizes a well known analogous result in one-dimension) says that, for the estimation of $\tau_0^2 + \sigma_{\mathrm{N}}^2$ 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 black) 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 black; 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 $\sigma_{\mathrm{N}} = 0$; 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 is that the weighted sum should perform nearly as well as (or possibly better than) the time-average ${T_n}^{-1} \int_0^{T_n} Z^2(t)\, dt$ 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 when

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

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, Perrin and Thomas-Agnan (2008). 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_{ν,θ_0}^2) which are clearly satisfied for the Matérn family, we have

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

Proposition 1. *If Z is a stationary one-dimensional centered Gaussian process with Matérn autocovariance $\tau_0^2\, \rho_{\nu,\theta_0}(t)$ with $1 < \nu < 2$, and if*

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

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*

$$\mathbb{E}\left(\left|\hat{\tau}^2_{\mathrm{wEV}} - \tau_0^2\right|^2\right) \leq {T_n}^{-1} 2\tau_0^4 \int_{-\infty}^{\infty} \rho^2_{\nu,\theta_0}(t)\, dt + O(\delta_n^\nu)$$

when $\delta_n \to 0$ *and* $T_n \to \infty$ *as* $n \to \infty$.

Proof To apply Theorem 2 of Elogne, Perrin and Thomas-Agnan (2008), it suffices to check that $X := Z^2$ is mean square differentiable (let X' its m.s. derivative) and satisfies: there exists κ such that, for all abscissae s and t,

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

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, establish bounds better than $O(\delta_n^\nu)$, especially for $\nu \geq 2$, as is done in Elogne, Perrin and Thomas-Agnan (2008) for an estimate that is less simple (briefly said, Z in place of Z^2 is interpolated and the estimate is defined as the integral of the squared interpolant; see their Corollary 1). An improved control of the accuracy of $\hat{\tau}^2_{\mathrm{wEV}}$ might also be useful to establish asymptotic properties of the resulting CGEM−wEV estimates of the range and the microergodic parameters. This may deserve further study.

3 Monte-Carlo simulation study

In this study, the domain on which the observations are located at the vertices of a regular grid (except in Sections 3.7-3.8), is a square (except in Sections 3.8-3.9 where missing regions are considered).Thus in Sections 3.1-3.6 and Section 3.10, $n_1 = n_2 = \sqrt{n}$ 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 Sections 3.7-3.9) so that the simulation settings be easily comparable with those of previously published studies. Even though the known theoretical justification (Girard, 2011) 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. To be more precise, if $\nu = 1/2$ then θ_0^{-1} varies in

$\{0.02, 0.05, 0.09, 0.125, 0.2, 0.3, 0.5, 0.7, 1., 1.5, 3.\}$. Otherwise we used slight variants of this set of values for $\theta_0^{-1}/2$ (resp. $\theta_0^{-1}\sqrt{3}$) when $\nu = 1/6$ (resp. $\nu = 3/2$). Figure 2 is a plot of the spherical autocorrelation model and the three considered Matérn autocorrelation models corresponding to the median of these chosen values, that is, 0.3. Somewhat arbitrarily we call a "very small (resp. large) correlation range", a range ten times smaller (resp. greater) than 0.3.

Recall that, in the case of no additive white noise (i.e. $\sigma_{\mathrm{N}} = 0$ 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 $\sigma_{\mathrm{N}} > 0$, it is then easy to see from (1.2) and (1.3) that, if both the observations $\boldsymbol{y}$ and the given σ_{N} are multiplied by a same constant, then the resulting $\hat{\tau}_{\mathrm{EV}}$ will be multiplied by this constant and, $\hat{b}_{\mathrm{EV}}$ being thus unchanged, the new estimating equation (1.3) 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

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

it is only through b_0 that τ_0 and σ_{N} influence the respective performance of CGEM−EV vs ML. Since we essentially consider cases with $\sigma_{\mathrm{N}} > 0$ we almost always use in the following b_0 instead of τ_0^2 as variance-parameter (equivalently, the following study, except in Section 3.10, 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 "very weak" noise ($b_0 = 10^{12}$), next, for settings with "moderate" noise, thirdly, with "rather strong" noise ($b_0 = 4$). We come back to the case "very weak noise" in the settings discussed Sections 3.7-3.10.

For all the experiments with $b_0 = 10^{12}$ the data might have been considered as exact data. In such a case, the estimation approaches (both ML and CGEM−EV) have a simpler form; see Girard (2011) for a brief discussion of the version (called Gibbs-energy estimating equation) of the equation (1.3) for the case without additive noise. However very ill-conditioned matrix inversions (especially in the case $\nu = 3/2$ and θ_0^{-1} large) would then have appeared. So we still chose a model with additive white noise (we come back to this point in Section 3.10).

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 satisfactory behavior of the CGEM−EV from this point of view: in "almost" all the considered cases we 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 which are equispaced in logarithmic scale). In fact, the "almost" term we use above, is only due to settings where the true corre-

lation 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 (except for the case, Section 3.9, of the much larger size for the datasets) 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 well approximated by $\sqrt{1/2000} \approx 2\%$.

Note that in the following statistical summaries, we use a logarithmic transformation for the estimates of θ_0 because it has often 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 such a transformation was not necessary for the considered estimates of $\tau_0^2\theta_0^{2\nu}$. 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.

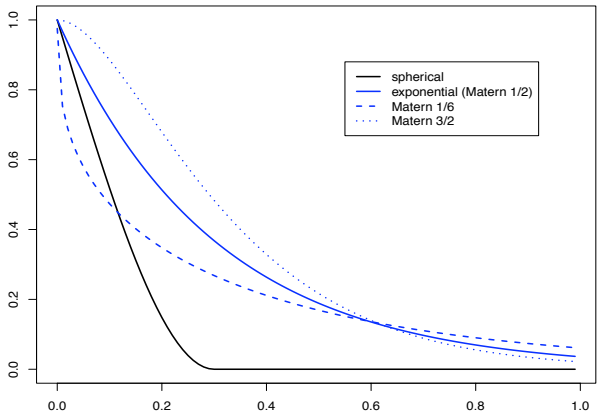

Figure 5: Spherical autocorrelation with range parameter $\theta^{-1} = 0.3$ and the three considered Matérn autocorrelation models $\rho_{\nu,\theta}$ with $\nu = 1/2$ (resp. $\nu = 1/6$ and $\nu = 3/2$), here with θ^{-1} (resp. $\theta^{-1}/2$ and $\theta^{-1}\sqrt{3}$) equal to 0.3

3.1. For $\nu = 1/6$ and $b_0 = 10^{12}$ we first observe in Table 1 that the relative accuracy of ML estimation for the microergodic parameter $b_0\theta_0^{2\nu}$ is extremely close to $\sqrt{2/n} = 0.047$ (here $n = 30\times30$) for any correlation range ($\theta_0^{-1}/2$) between 0.09

and 3, and it significantly (albeit moderately) departs from 0.047 only for the "very small" correlation range 0.02. Now, 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 $\theta_0^{-1}/2$ between 0.05 and 2 (with scarcely perceptible loss). Furthermore, the loss in efficiency is rather small for $\theta_0^{-1}/2 = 3$ and reasonable for $\theta_0^{-1}/2 = 0.02$.

Concerning the range-parameter, its true magnitude is, as expected from the recent literature (especially Zhang (2004)), much less easily estimable than $b_0\theta_0^{2\nu}$, 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 for the microergodic parameter $b_0\theta_0^{2\nu}$ is still very close to $\sqrt{2/n}$ but, this time, only for correlation ranges $(\theta_0^{-1}\sqrt{3})$ greater than, say, 0.2. Otherwise the attained accuracy decreases with θ_0^{-1}, e.g. the standard deviation is approximately two times $\sqrt{2/n}$ when the considered range is 0.04.

The attractive property of the CGEM−EV estimation is that its efficiency is still quite good, although a small departure from 1 is now perceptible: CGEM−EV root-inefficiency relative to ML is always between 1.05 and 1.10, except for $\theta_0^{-1}\sqrt{3} = 0.02$. For this "very small range" setting, a small degradation in efficiency is noticeable, together with a (very) small probability that the estimating equation has no solution.

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 is well in agreement with the theoretical result of Girard (2011) which states that full-efficiency for θ_0 is obtained as ν decreases. Notice that all the observed square root inefficiencies are nevertheless bounded by 1.54 and smoothly decrease toward 1.10 as the correlation range decreases toward 0.04 (we presently have no explanation for this decrease in both the variance and the bias) and thus CGEM−EV may be of interest also to users who would only target $\log_{10}(\theta_0)$.

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 (with a relative accuracy for both estimates of the microergodic parameter rather close to $\sqrt{2/n} = 0.052$ although a slight biais is now present in CGEM−EV) 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 that, in the infill asymptotics framework, the relative accuracy of $\sqrt{2/n}$ which holds in the case of exact data (see Du, Zhang and Mandrekar (2009) and Wang and Loh (2011)) is lost as soon as the observations are contaminated by a white noise (see Chen, Simpson and Ying (2000)). These experiments, compared with those of Section 3.2, show that decreasing the SNR from 10^{12} to 10^3 has a rather strong impact on the attainable accuracies when $\nu = 3/2$ whileas we have seen in Section 3.3 that, for $\nu = 1/2$, a SNR of 10^3 is large enough to attain accuracies near $\sqrt{2/n}$ (this relatively weak impact of the SNR for $\nu = 1/2$ and for $\nu = 1/6$ will be confirmed in the next section). Note that we have also performed experiments with $\nu = 3/2$ and $b_0 = 10$: even the microergodic parameter was then very difficult to estimate by ML in case of large correlation range, so we do not report the details of the comparison of ML and CGEM−EV here.

Now an important result from these simulations is that the efficiency of CGEM−EV for the microergodic parameter remains quite good when the true range $\theta_0^{-1}\sqrt{3}$ is less than, say, 0.7; otherwise CGEM−EV is not as satisfactory as in all the previous settings (notice, however, that the worst inefficiency is only 1.73^2 and corresponds to the largest θ_0^{-1}).

By comparing Table 5 and Table 2, we conclude that the signal-to-noise may have a noticeable impact on the efficiency of CGEM−EV. This is not in complete agreement with the theoretical results of Girard (2011, Section 4); but let us remind that these results only describe 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 (notice that the degradation is nevertheless "moderate" with only $n_{\mathrm{R}} = 1$).

3.5. We next present cases with $b_0 = 10$ or even $b_0 = 4$ in Table 6, Table 7 and Table 8. These three 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 even in case of "rather strong" noise, the CEGEM-EV approach appears to be very efficient (resp. rather efficient) to estimate the microergodic parameter (resp. the range parameter) for Matérn model with ν not too large or for the spherical model.

3.6. On the subject of what is sacrificed by using randomized-traces instead of exact traces, Table 5 and the previously discussed Table 6, Table 7 and Table 8 clearly demonstrate that the CGEM−EV efficiency is practically never degraded with $n_{\mathrm{R}} = 20$. In fact, even using $n_{\mathrm{R}} = 1$ induces a negligible degradation in the settings of Tables 1, 2, 3, 4, 7 and a moderate degradation for Table 8. Notice 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. This also holds for all the following contexts of Tables 9, 10, 11, 12 which all consider quite strong (even infinite) SNR. However it is important to observe that increasing n_{R} from 1 to 20 is justified in some contexts since this does increase the efficiency of CGEM−EV for the microergodic parameter in the case of large correlation range and weak SNR (see the lines corresponding to a range greater than .5, of Table 5, Table 6 and, although to a lesser extend, of Table 8).

3.7. We now present obtained simulation results in the case of the "very" 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 (using definition (2.2) for the weights) 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. Consider now another specific type of irregular design: one where the sites randomly but uniformly fill the domain $[0,1] \times [0,1]$, with possibly a few simple missing regions. The design actually used, with $n = 1000$, in this Section, is pictured Figure 6. Since developing a weighted version is not a trivial task, one must first assess the performance of the un-weigthed version. Simulations were made for the exponential model and $b_0 = 10^{12}$. The results are displayed in Table 10. It is rather surprising that, in contrast to the previous Section, the un-weigthed version is very nearly as efficient as ML. In fact the result are very similar to the ones displayed in Table 3. Other random uniform designs of this type were analyzed: a similar performance was consistently observed.

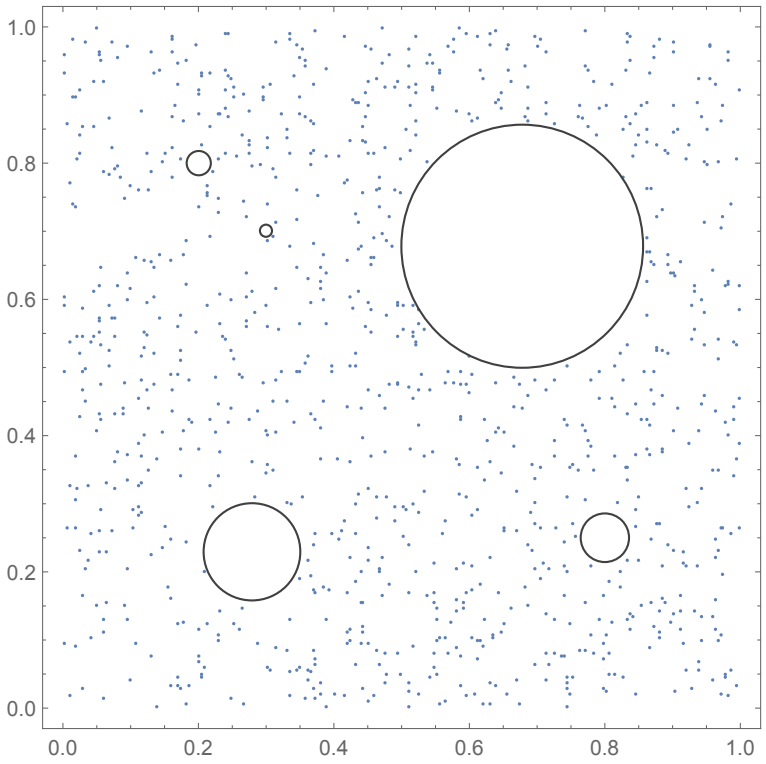

Figure 6: The used random design for the observation locations ($n = 1000$) drawn from an uniform distribution over the relative complement of five disks in $[0, 1] \times [0, 1]$

3.9. We have also considered a much larger data size, precisely $n = 57592$, and a strong SNR ($b_0 = 10^{12}$). The sites are a regular 256×256 lattice in $[0,1] \times [0,1]$ with five missing disks, those of the previous simulation. Notice that the linear solver used here was the PCG preconditioned by a classical factored sparse approximate inverse (FSAI) (see Girard (2015) for more details). Now only 200 replicates were analyzed for each setting. They were simulated using the R-package fields (Nychka, Furrer and Sain, 2009). ML was not implemented. However, for the case when the data can be assumed un-noisy, it is immediate to obtain the (Frechet-Darmois-)Cramer-Rao lower bound for the unbiased estimators of the micro-ergodic parameter in a more favorable situation: the one where θ_0 would be known. Indeed the square root of this lower bound is simply $\sqrt{2/n}\, c_0 = 0.00589\, c_0$. In fact, when $\sigma_{\mathrm{N}} > 0$, the Fisher information for c_0 when θ_0 is known, is easily shown to be $\mathrm{tr}\boldsymbol{A}^2_{b_0,\theta_0}/(2c_0^2)$. We also computed a 2-digit approximation of this by averaging a sufficient number of randomized-trace approximations for each of the 5 true ranges (θ_0^{-1}) we tried: at this accuracy we have not observed a departure from the no-noise value $n/(2c_0^2)$ (so the last column of Table 11 displays 0.0058 without repeating it). Now an important observation in Table 11 is that for all the correlation ranges considered, the standard errors of the CGEM$-$EV estimator of c_0 clearly attain this bound, and thus the statistical efficiency of CGEM-EV is very satisfactory, at least with respect to the estimation of this parameter.

Concerning the estimation of θ_0, we invested some computer time to also compute a 2-digit approximation of a Cramer-Rao type lower bound for this parameter. To avoid the inversion of a 2×2 ("badly") approximated (possibly ill-conditioned) information matrix, we assume this time that c_0 were known (the Cramer-Rao lower bound so obtained will necessarily be a lower bound for the case b_0 and θ_0 both unknown); we also have to compute a certain trace (see e.g. Gaetan and Guyon (2010) for deriving its expression) and for this we were required to average several thousand of primary randomized-trace estimates. Now the second important observation in Table 11 is that for all the correlation ranges considered, the standard deviations of the CGEM$-$EV estimator of θ_0 are clearly approximately equal to or lower than this bound, and in fact the column of squared biases (which become relatively more and more important as θ_0^{-1} increases) would have to be added to the squared standard deviations to obtain a column of values similar to the CR lower bounds. Thus the statistical efficiency of CGEM-EV is quite satisfactory also for the estimation of θ_0.

3.10. We finally consider a setting for which CGEM$-$EV is compared with the "Hybrid method", that is, the method proposed by Zhang and Zimmerman (2007) 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, precisely $\theta_0^{-1} = 1.5$). There was no additive white noise in the data. Since, in this paper, we maintain a model (1.1) with a given $\sigma_N > 0$ in our estimates (thus a "slightly" misspecified model), the problem of choosing σ_N arises. We have tried three values for σ_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 σ_N over several orders of magnitudes: indeed the results dispayed in the corresponding 3 columns of Table 12 are hardly distinguishable, the only exception is the case of large range ($\theta_0^{-1} = 1.5$) where the largest of the three σ_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 σ_N must be chosen "small" enough, keeping an eye on possible ill-conditioning in matrix inversions which may then require a "not too small" σ_N.

Secondly, the summaries in Table 12, 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 estimators are clearly less efficient than the CGEM−EV estimators (and especially for the estimation of τ_0^2) in this setting.

Thirdly, we also applied to the simulated data sets of this setting, the "estimation" method which could be suggested by the infill asymptotic results mentioned in the Introduction: fix an arbitrary range θ_1 and use $\widehat{\tau^2\theta} := \hat{\tau}^2_{\mathrm{ML}}(\theta_1)\theta_1^{2\nu}$ as estimate of $\tau_0^2\theta_0$. We chose here $\theta_1^{-1} = 0.4$. The summaries of the produced estimates are also displayed Table 12 in the column "fixed θ_1- ML", similarly as the other columns (except that this method does not provide estimates of τ_0^2.) As expected, this method is more efficient than ML (although not to a great extent) in the case $\theta_0^{-1} = 0.4$. And one observes that it is still slightly more efficient in the case $\theta_0^{-1} = 0.3$. Now it is useful to note that it is however less efficient than both ML and CGEM−EV for all the other cases, the loss in efficiency (mainly due to bias) being quite large for $\theta_0^{-1} = 0.1$, even when compared to the Hybrid method. We also tried this "fixed θ_1- ML" with the value 0.1 for θ_1^{-1} : as seen in Table 12, its global performance was even worse than with the previous choice $\theta_1^{-1} = 0.4$.

4 Conclusion and discussion

A rather extensive simulation study was performed for Matérn random fields with $\nu \in \{1/6, 1/2, 3/2\}$ observed on a dense grid of $[0, 1]^2$. A side remark is that, when both ν and the correlation range are "large", 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. Such a strong impact was not observed for the spherical autocorrelation model, for which simulation results rather similar to those for the exponential autocorrelation model ($\nu = 1/2$) are reported.

Firstly, concerning the question of existence and unicity of the root, the CGEM−EV method proved to be rather satisfactory for all the settings considered here.

Secondly, and perhaps more importantly for the usefulness of this approach, these experiments demonstrate that the CGEM−EV variance and range-parameter estimators, and, above all, the resulting estimate of the microergodic-parameter ($b_0\theta_0^{2\nu}$, or $\tau_0^2\theta_0^{2\nu}$ in the without-noise case), are nearly as efficient as the ML estimators for many various settings provided ν is not too large or the SNR is not too weak. Notice that the precise meaning of "not too weak" depends on ν since the CGEM−EV efficiency is still very good for a SNR of 4 when $\nu = 1/6$.

Otherwise this efficiency may be degraded especially for the cases with very large range-parameter. We do not know yet whether this somewhat disappointing behavior could have been rectified by using, say, a data size 10 times larger. Anyway, in such "unfavorable" settings, since the mentioned degradation remains moderate (indeed, the worst observed value for the inefficiency was 1.73^2 in the Matérn case, and 2.09^2 in the case of spherical autocorrelation), 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.

In all the considered settings, the replacement of the exact traces by their randomized version, does not significantly increase the inefficiency of CGEM−EV provided at least about 20 replicates are used for each randomized trace approximation. In fact by using only a single replicate (i.e. $n_{\mathrm{R}} = 1$) one observes an increase of inefficiency (compared to exact CGEM-EV) which is always "moderate", this degradation even being negligible in the case of strong enough SNR. This particularly good performance deserves, of course, a theoretical justification.

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 (2.2), produced a rather impressive improvement on CGEM−EV. A first theoretical property for this weighting scheme is also stated (Proposition 1) for one-dimensional Matérn

models. 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 straightforward and it deserves thus further study. However, notice that the study of refined weights can be useless in some contexts. Indeed a good news from our study is that the unweighted version still works very well when the irregular observation locations are drawn from a uniform distribution over a simple region. Note also that a possible alternative to these fast estimators $\hat{\tau}^2_{\mathrm{wEV}}$, might be considered, based on the remarkable fact that τ_0^2 may be efficiently estimated by the well established "tapered likelihood" approach (Kaufman, Schervish and Nychka, 2008) even when a very severe covariance tapering is used, thus yielding very significant computational savings (whereas the accuracy of the tapered likelihood estimator of the inverse-range θ_0 is often substantially degraded by choosing a severe tapering): this fact is indeed well depicted by Bevilacqua and Gaetan (2015) in their numerical study (see their tables 2, 3 and 4 where τ_0^2 is denoted σ^2, and especially the bottom-left panel of their figure 2).

The final experiment (reported in Table 12) 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 (2007), and is very robust with respect to the somewhat arbitrary choice of a "small" noise level (σ_{N}) in CGEM−EV. However in the case of a non-negligible additive white noise, we made no comparison with the Hybrid method 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).

Acknowledgements Thanks are due to the two reviewers for suggestions which led to several improvements in the paper, and to Remy Drouilhet for his invaluable help in producing the R-codes.

References

Bevilacqua, M., Gaetan, C., 2015. Comparing Composite Likelihood Methods Based on Pairs for Spatial Gaussian Random Fields. Stat. Comput. 25, 877-892.

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), 812-822.

Chen, H.S., Simpson, D.G. and Ying, Z., 2000. Infill asymptotics for a stochastic process model with measurement error. *Statistica Sinica* 10, 141-156.

Dahlhaus, R., Künsch, H., 1987. Edge effects and efficient parameter estimation for stationary random elds. Biometrika 74, 877-882.

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. Stat. Inference Stoch. 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.

Gaetan, C., Guyon, X., 2010. Spatial Statistics and Modeling. Series: Springer Series in Statistics.

Girard, D.A., 1989. A fast 'Monte-Carlo cross-validation' procedure for large least squares problems with noisy data. Num. Math., 56, 1-23.

Girard, D.A., 2011. Asymptotic near-efficiency of the "Gibbs-energy and empirical variance" estimating functions for fitting Matérn models to a dense (noisy) series. Preprint (version 2) http://hal.archives-ouvertes.fr/hal-00413693/en/

Girard, D.A., 2014. Estimating a Centered Ornstein-Uhlenbeck Process under Measurement Errors. Wolfram Demonstrations Project. http://demonstrations.wolfram.com/EstimatingACenteredOrnsteinUhlenbeckProcessUnderMeasurementE/

Girard, D.A., 2015. Estimating a centered Matérn (1) process: Three alternatives to maximum likelihood via conjugate gradient linear solvers. Wolfram Demonstrations Project. http://demonstrations.wolfram.com/EstimatingACenteredMatern1ProcessThreeAlternativesToMaximu

Girard, D.A., 2016. Asymptotic near-efficiency of the "Gibbs-energy and empirical-variance" estimating functions for fitting Matérn models – I: Densely sampled processes. Statistics and Probability Letters, 110, 191-197.

Guttorp, P., Gneiting, T. 2006. Studies in the history of probability and statistics XLIX: On the Matérn correlation family. Biometrika, 93, 989-995.

Guyon, X., 1982. Parameter estimation for a stationary process on a d-dimensional lattice. Biometrika 69, 95-106.

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

Hasselman, B., 2016. Solve Systems of Nonlinear Equations. R-package version 3.03, https://CRAN.R-project.org/package=nleqslv

Kaufman, C., Schervish, M.J., Nychka, D. 2008. Covariance tapering for likelihood-based estimation in large spatial data sets. J. Amer. Statist. Assoc. 103, 1545-1555.

Kaufman, C., Shaby, B., 2013. The role of the range parameter for estimation and prediction in geostatistics. Biometrika 100, 473-484.

Lahiri, S.N., Lee, Y., Cressie, N., 2002. On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. J. Stat. Planning Infer., 103, 65-85.

Nychka, D., Furrer, R. and Sain S., 2009. Fields: Tools for spatial data. R package. http://cran.r-project.org/web/packages/fields/

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

Robinson, P.M., 2008. Correlation testing in time series, spatial and cross-sectional data. Journal of Econometrics, 147, 5-16.

Robinson, P. M. and Vidal Sanz, J., 2006. Modified Whittle estimation of multilateral models on a lattice, Journal of Multivariate Analysis , 97, 1090-1120.

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.

Stein, M.L., Chen, J. and Anitescu, M., 2013. Stochastic approximations of score functions for Gaussian processes. Annals of Applied Statist. 7 1162-1191

Stroud, J.R., Stein M.L., and Shaun Lysen S., 2016. Bayesian and maximum likelihood estimation for Gaussian processes on an incomplete lattice. Journal of Computational and Graphical Statistics.

Tzeng S.L., Huang H.C., and Cressie N., 2005. A fast, optimal spatial-prediction method for massive datasets. Journal of the American Statistical Association, 100:472, 1343-1357

Wang, D., Loh, W.-L., 2011. On fixed-domain asymptotics and covariance tapering in Gaussian random field models. Electronic Journal of Statistics. 5, 238-269.

Yajima, Y., 1991. Asymptotic properties of the LSE in a regression model with long-memory stationary errors. Ann. Statist. 19 158-177

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.

Zimmerman, D. L., 1989. Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models. Journal of Statistical Computation and Simulation, 32, 1-15.

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	
$\theta_0^{-1}/2$	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 $\hat{b}\hat{\theta}^{2\nu}/(b_0\theta_0^{2\nu})$		
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 Matérn 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	
$\theta_0^{-1}\sqrt{3}$	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 $\hat{b}\hat{\theta}/(b_0\theta_0)$		
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 $\hat{b}\hat{\theta}/(b_0\theta_0)$		
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 Matérn model with $\nu = 3/2$ and $b_0 = 10^3$

	ML	CGEM-EV		randCGEM-EV			
				$n_R = 1$		$n_R = 20$	
$\theta_0^{-1}\sqrt{3}$	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 $\hat{b}\hat{\theta}^{2\nu}/(b_0\theta_0^{2\nu})$						
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 $\hat{b}\hat{\theta}/(b_0\theta_0)$						
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 Matérn 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			
				$n_R = 1$		$n_R = 20$	
$\theta_0^{-1}/2$	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 $\hat{b}\hat{\theta}/(b_0\theta_0)$						
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 (using (2.2)) estimates, and respective MSE inefficiency) for exponential model and $b_0 = 10^4$

	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 $\hat{b}\hat{\theta}/(b_0\theta_0)$				
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: $n = 1000$: Randomly but uniformly distributed sites on $[0,1] \times [0,1]$ with five missing disks (design pictured Fig.3). Simulation summary (mean, standard deviation of ML, CGEM-EV estimates, and respective MSE inefficiency) for exponential model and $b_0 = 10^{12}$

	ML	CGEM-EV	
θ_0^{-1}	mean ± sd	mean ± sd	ineff$^{1/2}$
	summary for the errors $\log_{10}(\hat{\theta}/\theta_0)$		
0.05	0.00± 0.05	0.00± 0.05	1.09
0.09	0.00± 0.07	0.01± 0.07	1.09
0.2	0.01± 0.12	0.02± 0.13	1.08
0.3	0.02± 0.16	0.03± 0.17	1.09
0.5	0.03± 0.22	0.06± 0.23	1.10
0.7	0.05± 0.26	0.09± 0.28	1.11
1.	0.08± 0.30	0.13± 0.32	1.12
1.5	0.11± 0.34	0.17± 0.37	1.14
	summary for the ratios $\hat{b}\hat{\theta}/(b_0\theta_0)$		
0.05	1.002± 0.051	1.004± 0.053	1.039
0.09	1.002± 0.048	1.004± 0.049	1.030
0.2	1.002± 0.046	1.004± 0.046	1.019
0.3	1.002± 0.045	1.004± 0.046	1.015
0.5	1.002± 0.045	1.004± 0.045	1.013
0.7	1.002± 0.045	1.004± 0.045	1.012
1.	1.002± 0.045	1.004± 0.045	1.010
1.5	1.002± 0.045	1.003± 0.045	1.009

Table 11: $n = 57592$. Observation locations on a 256×256 grid with missing values on 5 disks, the ones pictured Fig.3. Simulation summary (mean, standard deviation) of CGEM-EV estimates for the exponential model and $b_0 = 10^{12}$. The CR bounds are the classical Cramer-Rao lower bounds for the variances of regular unbiased estimators

	CGEM-EV			
θ_0^{-1}	mean ± sd of the errors $\log_{10}(\hat{\theta}/\theta_0)$	(CR bound)$^{1/2}$ for $\log_{10}(\theta_0)$, case c_0 known	mean ± sd of the ratios $\hat{b}\hat{\theta}/(b_0\theta_0)$	(CR boun for $c_0 =$ case θ_0 k
0.1	0.01± 0.07	0.065	1.000± 0.0059	0.005
0.2	0.02± 0.13	0.13	0.997± 0.0056	”
0.3	0.04± 0.17	0.17	0.997± 0.0056	”
0.7	0.09± 0.28	0.29	0.998± 0.0057	”
1.5	0.16± 0.36	0.41	0.998± 0.0055	”

Table 12: $n = 20 \times 20$. 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}, fixed θ_1 -ML estimates) for Exponential model with $\tau_0^2 = 2$ and without additive white noise

	ML	Hyb	fixed θ_1 -ML		CGEM-EV		
			$\theta_1^{-1} = 0.1$	$\theta_1^{-1} = 0.4$	$\sigma_{\mathrm{N}}^2 = .002$	$\sigma_{\mathrm{N}}^2 = .0005$	$\sigma_{\mathrm{N}}^2 = 10^{-8}$
θ_0^{-1}	mean of the errors $\hat{\tau}^2 - \tau_0^2$						
0.1	0.01	0.02	NA	NA	0.0091	–	0.0111
0.2	0.03	0.2	NA	NA	0.0029	–	0.0049
0.3	0.04	0.51	NA	NA	0.0101	–	0.0121
0.4	0.05	0.38	NA	NA	0.0150	0.0165	0.0170
1.5	0.12	–	NA	NA	0.0328	0.0343	0.0348
	MSE of $\hat{\tau}^2$						
0.1	0.11	0.26	NA	NA	0.1274	–	0.1274
0.2	0.35	1.84	NA	NA	0.3822	–	0.3822
0.3	0.64	5.08	NA	NA	0.7424	–	0.7425
0.4	0.87	6.19	NA	NA	1.1396	1.1396	1.1396
1.5	3.68	–	NA	NA	4.2818	4.2819	4.2819
	mean of the errors $\widehat{\tau^2\theta} - \tau_0^2\theta_0$						
0.1	0.0209	0.28	-0.093	-2.113	-0.0026	–	0.0601
0.2	-0.0047	0.12	1.371	-0.380	-0.0272	–	0.0263
0.3	-0.0001	0.08	1.489	-0.107	-0.0317	–	0.0207
0.4	-0.0017	0.09	1.486	-0.024	-0.0367	0.0023	0.0163
1.5	0.0005	–	1.334	0.051	-0.0446	-0.0075	0.0051
	MSE of $\widehat{\tau^2\theta}$						
0.1	2.5494	3.71	1.910	6.029	2.6038	–	2.6070
0.2	0.5561	0.75	2.627	0.602	0.5779	–	0.5788
0.3	0.2346	0.31	2.761	0.224	0.2463	–	0.2455
0.4	0.1299	0.18	2.747	0.123	0.1362	0.1350	0.1353
1.5	0.0085	–	3.242	0.0145	0.0106	0.0087	0.0087