An experimental assessment of the "Gibbs-Energy and Empirical-Variance" estimating equations (via Kalman smoothing) for Matérn processes

Didier A. Girard

Laboratoire Jean Kuntzmann, Grenoble

Mots clefs : Gaussian process, Matérn correlation, ARMA time-series, Maximum likelihood, CGEM-EV estimating equations, Kalman smoothing, R-package **lsta**, R-package **dml**

The problem of estimating the parameters of a stationary Gaussian process whose autocorrelation function belongs to the Matérn class, appears in many contexts (e.g. [1, 2]).

Def. of a Matérn process on $\mathbb{R}$, with "differentiability" parameter ν :

Matérn processes on $\mathbb{R}$ can be easily formulated in terms of the Fourier transform of their autocorrelation function, namely the spectral density over $(-\infty, +\infty)$:

$$f^*_{\nu,b,\theta}(\omega) = b\, g^*_{\nu,\theta}(\omega), \quad \text{with} \quad g^*_{\nu,\theta}(\omega) := \frac{C_\nu\, \theta^{2\nu}}{(\theta^2+\omega^2)^{\frac{1}{2}+\nu}}. \tag{1.1}$$

In this paper the constant $C_\nu = \frac{\Gamma(\nu+\frac{1}{2})}{\sqrt{\pi}\Gamma(\nu)}$ is chosen so that $\int_{-\infty}^{\infty} g^*_{\nu,\theta}(\omega)d\omega = 1$. Thus b is the variance of $Z(t)$ and θ is the so-called "inverse-range parameter" (in fact, it is $\nu^{1/2}/\theta$ which can be interpreted as an effective range or "correlation length" independently of ν, cf Stein (1999, Section 2.10); we will often drop the term "inverse".

The parameter $c = b\,\theta^{2\nu}$ is generally called the "microergodic coefficient".

Assume that n observations of one realization $Z(\cdot)$ of a Matérn process on $\mathbb{R}$ are given.
Also, we simply assume here that the constant mean of the process is zero and the noise is a Gaussian white noise and its level is known, says 1.
(more precisely, the "noisy" measurement of $Z(t)$ for t in the set $\{\delta, 2\,\delta, \cdots, (n-1)\,\delta, n\,\delta = 1\}$ are

$$y_i = Z(i/n) + \epsilon_i, \quad i = 1, \cdots, n \qquad \text{where } \epsilon_i \text{ are i.i.d. } \mathcal{N}(0,1) \qquad)$$

This work compares two estimation methods (of b and θ) for the Matérn subclass which has its "differentiability" parameter ν fixed to an half-integer, often-used values are 1/2, 3/2 or 5/2 (e.g. [3], [1]).

Indeed with such ν's, the observed series then coincide with particular ARMA series observed with noise.

Maximum likelihood (ML) estimation, via a state-space reformulation, is then classical (computing the criterion or its gradient is classicaly obtained via Kalman smoothing):
the well-established *R*-package **dml** [5] is used. Known analytical constrains on the ARMA coefficients (as it is the case here; see the *Mathematica* Demos [6] and [7] for two examples) can be dealt with by **dml**.

Heuristics for CGEM-EV estimating equation (not restricted to ARMA series):

Let

$$\rho(t,\theta)\left(=\frac{\mathbb{E}[x(s+t)\,x(s)]}{\mathbb{E}[x(s)\,x(s)]}\right)=\begin{cases}\exp(-\theta\,|t|) & \nu=1/2\\ (1+\theta\,|t|)\exp(-\theta\,|t|) & \nu=3/2\end{cases}.$$

Let R_θ denote the candidate correlation matrix of $\mathbf{Z}=\{Z(0),Z(\delta),\cdots,Z((n-1)\,\delta),Z(1)\}$, with θ as inverse-range, that is, the $n\times n$ matrix whose element in the i^{th} row and j^{th} column is $[R_\theta]_{i,j}=\rho(|i-j|\,\delta,\theta)$.

NB: It is important to notice that $R(\theta)$ becomes ill-conditioned for large n and θ decreasing (in the mathematica Demo[7] for $\nu=3/2$, the lower-bound $\theta=1.5$ is used for n=8192)

Zhang and Zimmerman (2007) recently proposed to use the classical weigthed least square method (not statistically fully efficient but much less costly than maximum likelihood (ML)) to estimate the range parameters, next, to plug-in these parameters (the θ here) in the likelihood which is then maximized with respect to b (the solution, say $\hat{b}_{\text{ML}}(\theta)$, being either the explicit (1.2) in the no-nugget case, or obtained iteratively by Fisher scoring otherwise). The idea underlying this method is that, at least for the infill asymptotic context (i.e. $\delta=1/n$ and n large), even if θ is fixed at a wrong value θ_1, the product $\hat{b}_{\text{ML}}(\theta_1)\theta_1^{2\nu}$ still remains an efficient estimator of $c_0:=b_0\theta_0^{2\nu}$ (see Du et al. (2009), Wang and Loh (2011) for recent results of this type). As is now classical (Stein (1999)), c_0 will be called the microergodic parameter of the Matern model (1.1). Zhang (2004) showed that a good estimation of c_0 is more important than a joint estimation of b_0 and θ_0 to obtain a good prediction of $Z(\cdot)$ for dense sampling designs.

The method we propose here, firstly reverses the roles of variance and range, in that it is based on a very simple estimate for the variance, namely the empirical variance in the no-nugget case, and its corrected version for biais otherwise, which is simply defined by

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

Secondly we propose to replace the maximization of the likelihood w.r.t. θ by the simple following estimating equation in θ, in the with-nugget case: solve, with b fixed at $\hat{b}_{\text{EV}}$

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

In the no-nugget case, this equation in θ is simply replaced by $\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}=n\hat{b}_{\text{EV}}$. One may call "Gaussian Gibbs energy" (GE in short) of the underlying discretely sampled process the quantity $(1/n)\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}$ and it is easily seen that $(b/n)\left(\mathbf{y}^T A_{b,\theta}\left(I_n-A_{b,\theta}\right)\mathbf{y}+\text{tr}(I_n-A_{b,\theta})\right)$ is the conditional Gibbs energy mean (CGEM) obtained by taking the expectation of $(1/n)\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}$, conditional on $\mathbf{y}$, for the candidate parameters b,θ. So equation (1.3) in θ will be called the CGEM-EV estimating equation (GEV in the no-nugget case) and we will denote by $\hat{\theta}_{\text{GEV}}$ this new range parameter estimate.

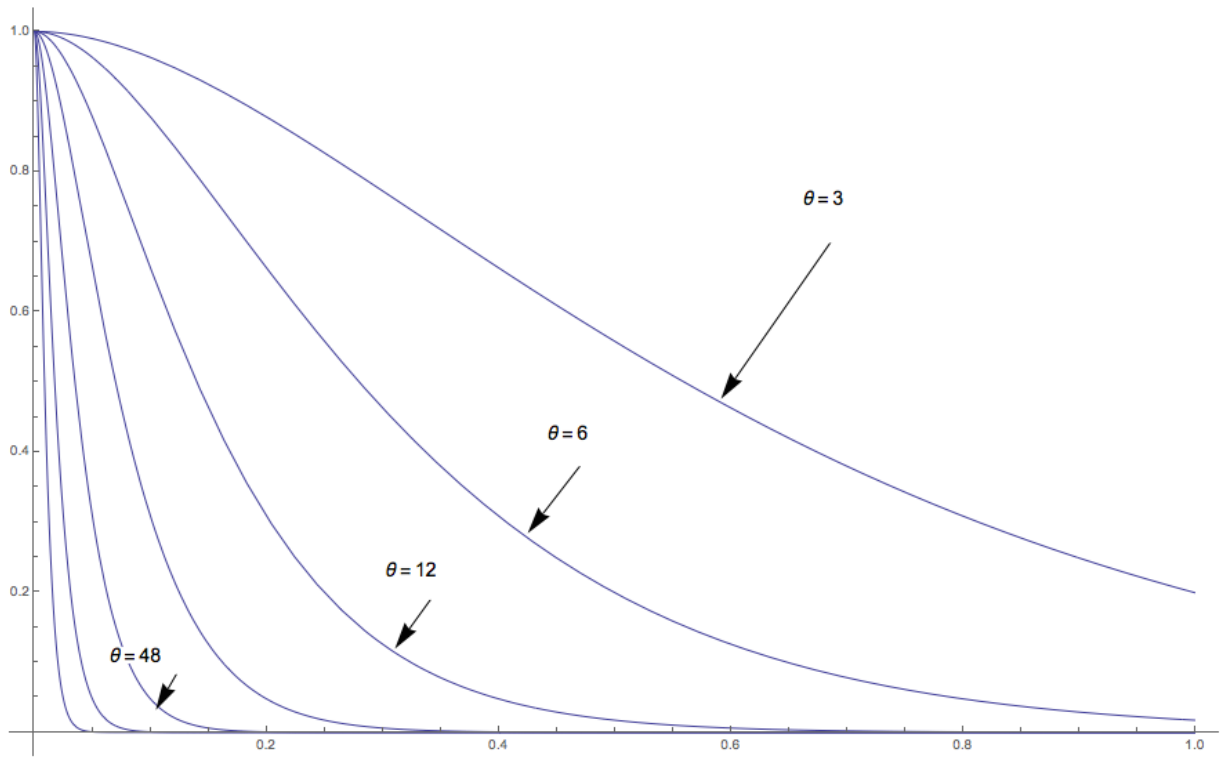
Computation of the quadratic term and the trace-term of (1.3) are direct by-products of a Kalman smoothing.

To solve (1.3) a simple fixed-point algorithm [6, 7, 8, 9] is used here. It proves to be reliable (with fast convergence).

A rather "extensive" Monte-carlo assessment of CGEM-EV has been made, for the cases ν =1/2 and ν =3/2

Design for the following histograms (1000 replicates for each setting):

- □ $\nu = 3/2$
- □ time-series length n = 800, 5000, 20 000
- □ signal-to-noise ratios b_0 $\left(= \frac{\mathrm{var}(Z)}{\mathrm{var}(\epsilon)}\right)$ chosen among: 20^2, 100^2, 1000^2, $10\,000^2$
- □ For each b_0 , range-parameter θ_0 = 3, 6, 12, 24, 48, 96, 192

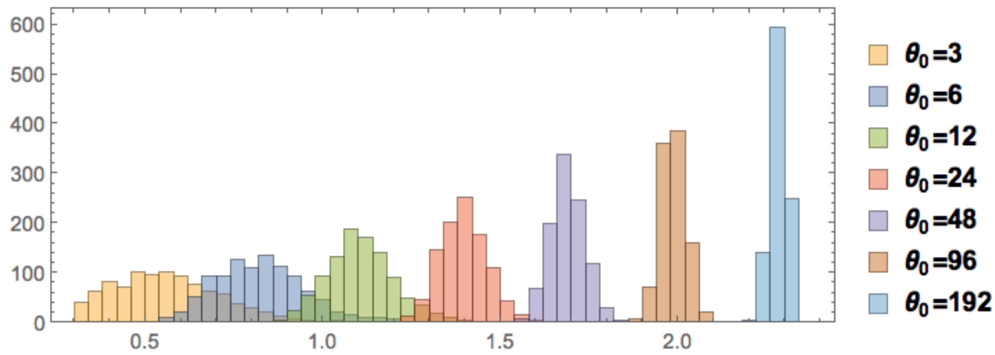


7 autocorrelation functions "Matérn-(3/2) " for θ in {3, 6, 12, 24, 48, 96, 192}

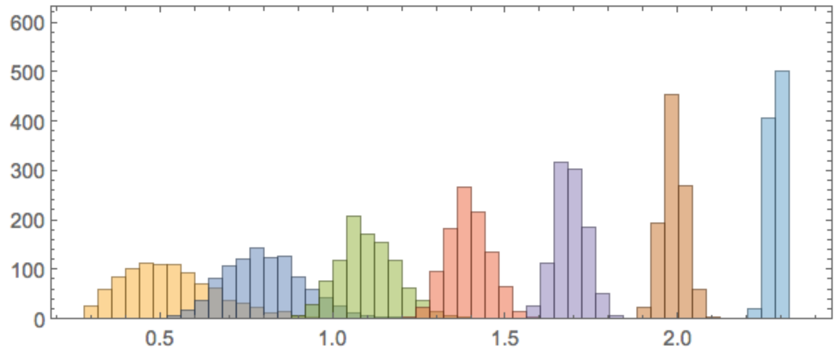
histograms of $\log_{10}(\hat{\theta})$ (1000 time-series replicates), time-series length n=5000

$$b_0 \ \left(= \frac{\mathrm{var}(Z)}{\mathrm{var}(\epsilon)} \right) \ = 1000^2$$

CGEM - EV

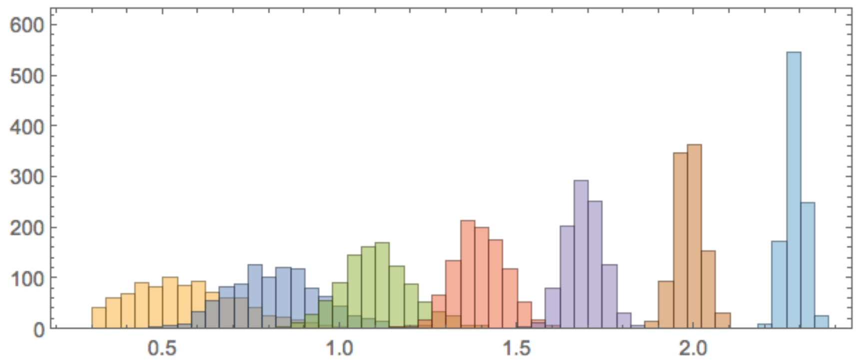


ML (CGEMEV as initial values)

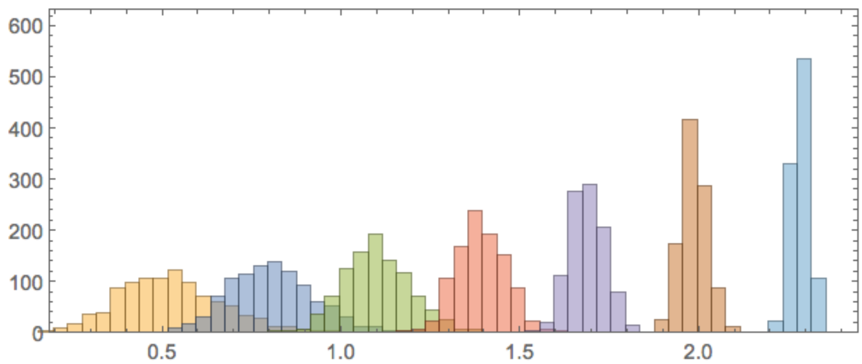


$$b_0 = 20^2$$

CGEM - EV

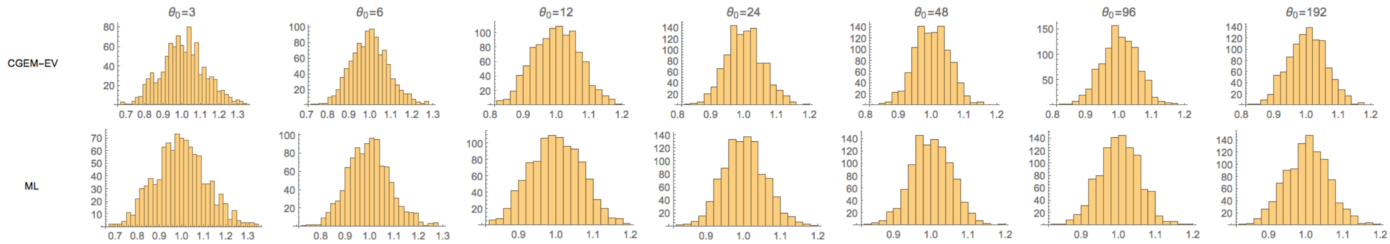


ML (truth as initial values)

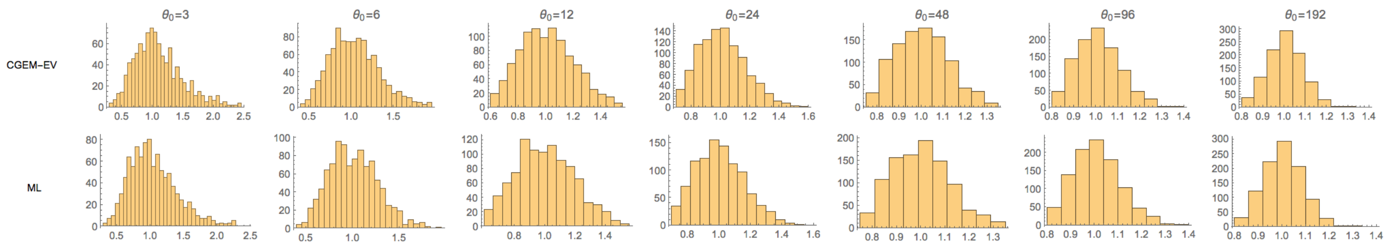


histograms of $\hat{c}/c_0$ (1000 time-series replicates) where $c_0 = b_0\,\theta_0^{2\nu}$ and $\hat{c} = \hat{b}\,\hat{\theta}^{2\nu}$, time-series length n=800

$$b_0 \left(= \frac{\mathrm{var}(Z)}{\mathrm{var}(\epsilon)} \right) = 100^2$$

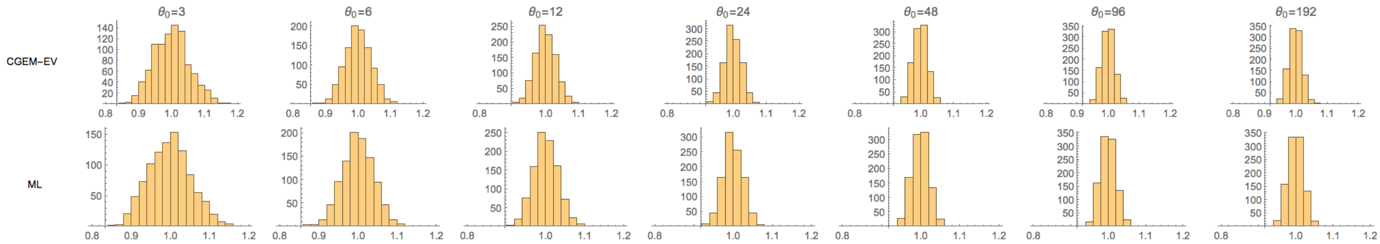


$$b_0 = 20^2$$

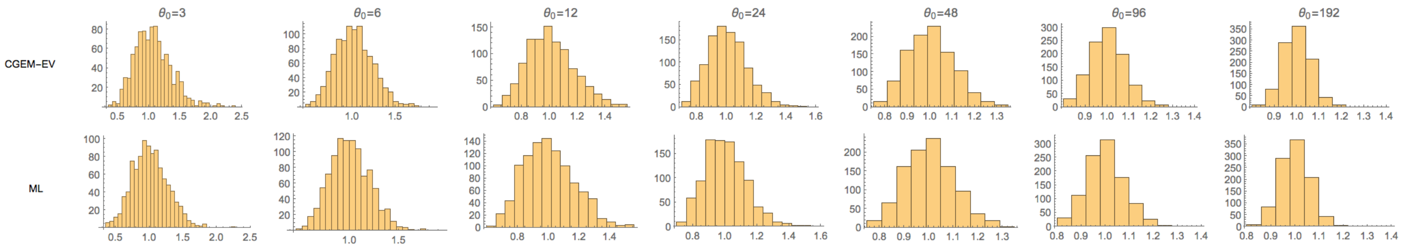


histograms of $\hat{c}/c_0$ (1000 replicates), time-series length n=5000

$b_0 = 1000^2$



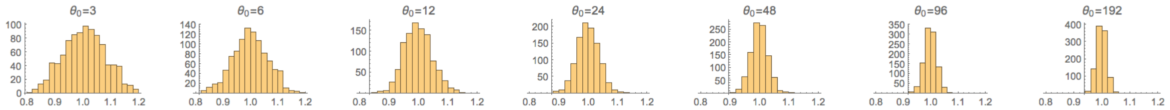
$b_0 = 20^2$



histograms of $\hat{c}/c_0$ (1000 replicates) only for CGEM-EV, time-series length $n = 20\,000$

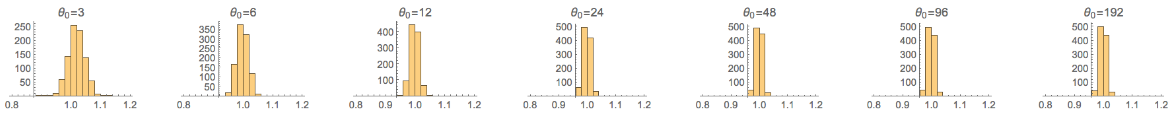
$b_0 = 1000^2$

CGEM−EV



$b_0 = 10\,000^2$

CGEM−EV



... CONCLUSIONS

□ for $\nu = 3/2$ (and $\nu = 1/2$, not presented here) the statistical efficiency is quite good, and
CGEM − EV is, in average, about 10 time faster than ML

□ for ν non halfinteger and n large,

- □ ML estimation requires $O(n^2)$ computation (see **ltsa** package),
- □ finding the root of CGEM − EV can be an $O(n \log(n))$ computation when iterative linear solvers exit (see [8] for $\nu = 1$)

References

[1] Rasmussen C.E., Williams C.K.I. (2006), *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press. www.gaussianprocess.org/gpml/chapters.

[2] Zhang H. (2012). "Asymptotics and Computation for Spatial Statistics," in *Advances and Challenges in Space-time Modelling of Natural Events (Lecture Notes in Statistics, Vol. 207)* (E. Porcu, J. M. Montero, and M. Schlather, eds.), New York: Springer, pp. 239—–252. doi:10.1007/978-3-642-17086-7_10.

[3] Kaufman C.G., Shaby, B.A. (2013). "The Role of the Range Parameter for Estimation and Prediction in Geostatistics," *Biometrika* 100 (2), pp. 473—–484. doi:10.1093/biomet/ass079.

[4] Girard D.A. (2012). "Asymptotic Near-Efficiency of the 'Gibbs-Energy and Empirical-Variance' Estimating Functions for Fitting Matérn Models to a Dense (Noisy) Series," 2012. arxiv.org/pdf/0909.1046v2.pdf.

[5] Petris Giovanni (2010). "An R Package for Dynamic Linear Models," *Journal of Statistical Software*, 36(12), 1-16. URL: http://www.jstatsoft.org/v36/i12/

[6] Girard D.A. (2014) "Estimating a Centered Ornstein-Uhlenbeck Process under Measurement Errors," *Wolfram Demonstrations Project*, Published: July 1, 2014. URL: demonstrations.wolfram.com/EstimatingACenteredOrnsteinUhlenbeckProcessUnderMeasurementE/

[7] Girard D.A. (2015) "Three Alternatives to the Likelihood Maximization for Estimating a Centered Matérn (3/2) Process, " *Wolfram Demonstrations Project*, URL: demonstrations.wolfram.com/ThreeAlternativesToTheLikelihoodMaximizationForEstimatingACe/

[8] Girard D.A. (2015) " Estimating a Centered Matérn (1) Process: Three Alternatives to the Likelihood Maximization via conjugate gradient linear solvers" *Wolfram Demonstrations Project*, URL: demonstrations.wolfram.com/EstimatingACenteredMatern1ProcessThreeAlternativesToMaximumL/

[9] Girard D.A. (2015) " On the convergence of a simple iterative algorithm for fitting Matérn isotropic models to large (noisy) datasets " To appear