NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

FITTING GAUSSIAN MARKOV RANDOM FIELDS TO GAUSSIAN FIELDS

by

Håvard Rue and Håkon Tjelmeland

PREPRINT
STATISTICS NO. 16/1999

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL
http://www.math.ntnu.no/preprint/statistics/1999/S16-1999.ps
Håvard Rue has homepage: http://www.math.ntnu.no/~hrue
E-mail: hrue@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Fitting Gaussian Markov random fields to Gaussian fields

HÅVARD RUE AND HÅKON TJELMELAND
DEPARTMENT OF MATHEMATICAL SCIENCES
NTNU, NORWAY

MAY 19, 1999

SUMMARY

This paper discusses the following task often encountered building Bayesian spatial models: construct a homogeneous Gaussian Markov random field (GMRF) on a lattice with correlation properties either as present in observed data or consistent with prior knowledge. The Markov property is essential in design of computational efficient Markov chain Monte Carlo algorithms used to analyse such models. We argue that we can restate both tasks into the problem of fitting a GMRF to a prescribed stationary Gaussian field on a lattice when both local and global properties are important. We demonstrate that using the Kullback-Leibler discrepancy often fails in this task, giving a severe undesirable behaviour of the correlation function for lags outside the neighbourhood. We propose a new criterion that resolves this difficulty, and demonstrate that GMRFs with small neighbourhoods can approximate Gaussian fields surprisingly well even with long correlation lengths. Finally, we discuss implications of our findings for likelihood based inference for general Markov fields when also global properties are important.

KEYWORDS: Conditional Autoregressive models, Empirical Bayes, Gaussian Markov random field, Gaussian fields, Maximum likelihood, Kullback-Leibler discrepancy, Markov chain Monte Carlo.
ADDRESS FOR CORRESPONDENCE: H. Rue, Department of Mathematical Sciences, The Norwegian University of Science and Technology, N-7491 Trondheim.
E-MAIL: Havard.Rue@math.ntnu.no
ACKNOWLEDGEMENTS: We thank Arnoldo Frigessi and Gudmund Høst for stimulating discussions. HR thanks the EU-TMR project on Spatial Statistics (ERB-FMRX-CT960095) for support and inspiration.

1 INTRODUCTION

A task often encountered in Bayesian spatial, spatio-temporal and imaging models, is to represent a spatial variable $\boldsymbol{x}$ by a stationary Gaussian field for stochastic modelling purpose. We assume that $\boldsymbol{x}$ is defined on an $n_{\mathrm{r}} \times n_{\mathrm{c}}$ lattice Λ, where $n_{\mathrm{r}} \leq n_{\mathrm{c}}$. In some cases we have one or several realisations of the variable of interest, and we can then use an empirical Bayes approach estimating a Gaussian field from the realisations. The common procedure is to subtract a parametric form of the mean value, estimate the variance and then fit a parametric correlation function to the empirical found from the data. The estimated Gaussian field can then be fed back into the Bayesian model. In those cases where empirical Bayes is not appropriate simply because the spatial variable $\boldsymbol{x}$ of interest is not observed, say, it is common to specify a parametric correlation function, perhaps using a hierarchical model for the parameters in it.

Bayesian spatial models are often analysed using Markov chain Monte Carlo (MCMC) techniques (refer to Gilks, Richardson & Spiegelhalter (1996) for an introduction to MCMC). Such techniques generates (in the limit) dependent realisations from the posterior density by running a Markov chain with the posterior as the equilibrium density. The Markov chain is usually constructed so that only a small set of variables, often single sites, are updated at each iteration. This is done by sampling from the conditional density of the selected variable or groups of variables, or a density close to it, given the rest. For this to be computational tractable, we often need a Markov property in the posterior, meaning that the conditional density for one variable conditioned on the rest, do not depend on all other variables but only those in the neighbourhood. Typically, neighbourhood variables are close either in space and/or time. It is common to ensure a Markov property in the posterior by using prior fields with Markov property. The observations in the posterior enter the Bayesian model through the likelihood and does not usually destroy the Markov properties present in the prior.

The above scenario requires that we provide a Gaussian field on Λ based on the observed data, but the Gaussian field should have a Markov property. This subclass of Gaussian fields (on Λ) is called Gaussian Markov random fields (GMRF) or conditional Autoregressive models (Cressie, 1993; Besag & Kooperberg, 1995). In the case where an empirical Bayes approach is not appropriate, we might have to specify the GMRF by a neighbourhood and a set of conditional densities uniquely identifying the joint density. Two important questions now arise.

1. How should we fit a GMRF to observed data, the estimated Gaussian field or to our prior knowledge about the spatial behaviour? Here, global as well as local properties of the GMRF are important.
2. Is a GMRF with a small neighbourhood capable to approximate the behaviour of a Gaussian field with long correlation length?

This paper is about these questions. We will first argue that it suffice to discuss the problem of how to fit a GMRF to a given Gaussian field on a lattice, no matter if there are observed data or not. This because the empirical correlations from observed data can be thought of as the correlation function of a Gaussian field. Further, for those cases we need to specify the field from our prior knowledge only, we find it easier to give the variance and a correlation function in a Gaussian field, which gives a good (and correct) intuition of the behaviour of the field as the marginal joint density of any two distinct sites are then known. A GMRF specified through conditional properties are much harder to interpret in terms of its correlation properties. This because the marginal joint density of two distinct sites, x_{ij}

and x_{kl}, does not only depend on the conditional density for x_{ij} and x_{kl} conditioned on the rest, but also the marginal density for the rest. We present counter-intuitive full conditionals that support this claim.

Besag & Kooperberg (1995) performed a study where they fitted the coefficients in a non-homogeneous GMRF to account for the boundary effects in small lattices. The fit was done using the Kullback-Leibler discrepancy between a GMRF and a prescribed covariance matrix (i.e. Gaussian field), and solved numerically using an algorithm by Dempster (1972). The numerical algorithm for non-homogeneous GMRFs is only computational tractable for small lattices. We focus on homogeneous GMRFs on a torus in order to derive a computational feasible algorithm for large lattices, and demonstrate in Section 3 that the Kullback-Leibler discrepancy is not suitable for fitting a GMRF to a Gaussian field: correlations outside the neighbourhood can have severe undesirable behaviour that is not supported by the data. Hence, we do not obtain a GMRF that approximate well the correlation function of a Gaussian field using the Kullback-Leibler discrepancy. However, we propose an improved criterion which shows to give a good overall fit (in terms of correlations and variance) between the GMRF and the Gaussian field.

The plan of this paper is as follows. In Section 2 we present GMRFs and discuss the computational advantages of GMRFs over Gaussian fields. In Section 3 we discuss how to use the Kullback-Leibler discrepancy between the Gaussian field and the GMRF, as a criteria to fit a GMRF. The Kullback-Leibler discrepancy has close connections to both the principle of maximum-likelihood and maximum entropy, which is basic in statistical inference, and is thereby a natural choice. By using properties of covariance matrices for stationary fields on a torus, we are able to evaluate the discrepancy and obtain the best fit using numerical minimisation. The results obtained with the Kullback-Leibler discrepancy are not satisfactory, but still interesting. We propose in Section 4 an improved criterion for fitting GMRFs, which includes the Kullback-Leibler discrepancy as a special case. The results demonstrate that GMRFs are indeed capable to fit Gaussian fields even for long correlation lengths. In Section 5 we discuss the implications of our findings in Sections 3 and 4 to likelihood based inference for GMRFs and general Markov fields from real data.

2 PRELIMINARIES

A general Gaussian field $\boldsymbol{x}$ on an $n_{\mathrm{r}} \times n_{\mathrm{c}}$ lattice Λ, $n_{\mathrm{r}} \leq n_{\mathrm{c}}$, is a multivariate Gaussian variable with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. A Gaussian field on Λ is a Gaussian Markov random field (GMRF), if it satisfies a Markov property

$$\pi(x_{ij} \mid x_{kl}, (k,l) \in \Lambda \setminus \{(i,j)\}) = \pi(x_{ij} \mid x_{kl}, (k,l) \in \partial_{ij}),$$

where ∂_{ij} is the neighbourhood of (i,j). In most cases, ∂_{ij} is a set of sites close to (i,j) in some norm,

$$\partial_{ij} = \{(k,j) \in \Lambda \setminus \{(i,j)\} : \|(k,l) - (i,j)\| \leq c\}. \tag{1}$$

GMRFs have been used in many different areas, typically in situations where the Markov property is needed. Refer to Besag & Kooperberg (1995) for an overview.

The Markov property induced by the neighbourhood system $\{\partial_{ij}, (i,j) \in \Lambda\}$ imposes a zero-pattern structure in the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ as follows: $Q_{ij,kl}$ is zero unless (i,j) and (k,l) are neighbours. A more formal statement is that x_{ij} and x_{kl} are conditionally independent given the remaining variables *iff* $Q_{ij,kl} = 0$. The conditional mean and variance are expressed as follows

$$\mathrm{E}(x_{ij} \mid \boldsymbol{x}_{\Lambda\setminus\{(i,j)\}}) = -\sum_{(k,l)\in\partial_{ij}} \frac{Q_{ij,kl}}{Q_{ij,ij}} x_{kl} \qquad \text{and} \qquad \mathrm{Var}(x_{ij} \mid \boldsymbol{x}_{\Lambda\setminus\{(i,j)\}}) = 1/Q_{ij,ij} \tag{2}$$

which depend only on sites in the neighbourhood ∂_{ij}. The coefficients in $\boldsymbol{Q}$ have nice conditional interpretations: $-Q_{ij,kl}/Q_{ij,ij}$ is the conditional correlation coefficient between x_{ij} and x_{kl} conditioned on the rest, and $1/Q_{ij,ij}$ is the conditional variance of x_{ij} conditioned on the rest.

A GMRF have several benefits compared to a general Gaussian field, among these are the following.

MARKOV PROPERTY The Markov property of a GMRF is a valuable property, which is often necessary for model analysis using Markov chain Monte Carlo techniques as discussed in Section 1.

FAST AND EXACT SIMULATION A GMRF can be fast and exactly simulated using $\mathcal{O}(n_\mathrm{r}^3 n_\mathrm{c})$ flops using the algorithm of Rue (1999), which is easily extendable for also non-rectangular "lattices". Repeated iid samples are available in only $\mathcal{O}(n_\mathrm{r}^2 n_\mathrm{c})$ flops. The constant of proportionality in the flops-count depends on the size of the neighbourhood. For large lattices, there is also an exact simulation algorithm by Rue, Marthinsen & Husby (1999) which is designed specially for parallel computers. The algorithm runs with a speed-up close to the theoretical limit and has low memory requirement. Neither of the algorithms require homogeneous coefficients in the precision matrix. Thus we have full generality which is important in applications. In comparison, simulation algorithms for general Gaussian fields require $\mathcal{O}(n_\mathrm{r}^3 n_\mathrm{c}^3)$ flops. Further, the memory requirement is often lower for GMRFs and this is important for large fields. The fast simulation algorithms for GMRFs can also be used to perform large block-updates in a MCMC algorithm, see the previous point.

LIKELIHOOD The likelihood of a GMRF can be computed in $\mathcal{O}(n_\mathrm{r}^3 n_\mathrm{c})$ flops (Rue, 1999), which is important if the normalisation constant is needed in estimation or in a MCMC algorithm. Examples of the latter are cases where the coefficients in $\boldsymbol{Q}$ depends on hyperparameters or the region of interest varies. For a general Gaussian field, we need $\mathcal{O}(n_\mathrm{r}^3 n_\mathrm{c}^3)$ flops to compute the normalisation constant.

For the neighbourhood of the GMRF, we assume a maximum norm is used in (1), so ∂_{ij} is a window of size $(2m+1)\times(2m+1)$, $m = 1, 2, \ldots$ centred at (i,j). The motivation for using a square and not for example a discrete circle, is that the computational cost of the sampling algorithm of Rue (1999), including the computation of the likelihood, is (almost) identical for a square of size $(2m+1)\times(2m+1)$ and a discrete circle with diameter $(2m+1)$. Hence, our choice gives us some extra parameters for free.

The Gaussian fields we use for simulation experiments in later sections, are all stationary and isotropic. We use the three commonly used Gaussian, exponential and spherical correlation functions (Cressie, 1993)

$$\begin{aligned}
\text{Exponential:}\quad & \rho(k,l) = \exp(-3d(k,l)/r)\\
\text{Gaussian:}\quad & \rho(k,l) = \exp(-3d(k,l)^2/r^2)\\
\text{Spherical:}\quad & \rho(k,l) = \begin{cases} 1 - \frac{2}{\pi}\left(\frac{d(k,l)}{r}\sqrt{1-(\frac{d(k,l)}{r})^2} + \arcsin\frac{d(k,l)}{r}\right), & d(k,l)\le r\\ 0, & d(k,l) > r \end{cases}
\end{aligned} \tag{3}$$

where $d(k,l)$ denotes the Euclidean distance from $(0,0)$ to (k,l). For the spherical correlation function r is the correlation length, whereas in the exponential and Gaussian cases r defines an effective range of correlation, i.e. the distance where the correlation drops below about 0.05. Slightly imprecise, in the remainder of this paper we use the terms correlation length and range for r in all three cases.

We denote by $\gamma(k,l)$ the covariance function for the Gaussian field given as the covariance between x_{00} and x_{kl}. Thus, $\rho(k,l) = \gamma(k,l)/\gamma(0,0)$ where $\gamma(0,0)$ denotes the variance. The corresponding quantities for the GMRF are $\widetilde{\rho}$ and $\widetilde{\gamma}$. The number of parameters in a $(2m+1) \times (2m+1)$ neighbourhood is $m(m+1)/2 + m + 1$ when we impose rotation and reflection invariance. We denote by θ_{kl} the element $Q_{00,kl}$. Figure 1 displays the parameters (and those who are equal due to rotation and reflection invariance) for $m = 2$.

3 FITTING A GMRF USING KULLBACK-LEIBLER DISCREPANCY

In this section we discuss the use of Kullback-Leibler discrepancy between two densities as a criterion for fitting a GMRF to a Gaussian field. We then show how to evaluate the discrepancy using only $\mathcal{O}(n_{\mathrm{r}} n_{\mathrm{c}} \log(n_{\mathrm{r}} n_{\mathrm{c}}))$ flops. This makes numerical minimisation of the Kullback-Leibler discrepancy feasible in practice for large fields.

The Kullback-Leibler discrepancy measure the discrepancy by approximating the density $f(\boldsymbol{x})$ by $\widetilde{f}(\boldsymbol{x})$, and is defined as

$$\mathrm{KL}(f, \widetilde{f}) = \int_{\boldsymbol{x}} f(\boldsymbol{x}) \log\left(\frac{f(\boldsymbol{x})}{\widetilde{f}(\boldsymbol{x})}\right) \, d\boldsymbol{x}. \tag{4}$$

$\mathrm{KL}(f, \widetilde{f})$ is not symmetric in the arguments and defines a pseudo-distance as the triangle inequality does not hold true. $\mathrm{KL}(f, \widetilde{f})$ has close connections to both the principle of maximum-likelihood and maximum entropy, which is basic in statistical inference. $\mathrm{KL}(f, \widetilde{f})$ is a natural candidate for a pseudo-metric between the Gaussian field and the approximating GMRF, and was also used by Besag & Kooperberg (1995) as discussed in Section 1.

For a stationary Gaussian field f and a GMRF $\widetilde{f}$ on a torus, the computation of $\mathrm{KL}(f, \widetilde{f})$ is especially efficient.

THEOREM 1 *Let f be a zero mean stationary Gaussian field on an $n_r \times n_c$ torus with covariance function $\gamma(k,l)$ and covariance matrix $\boldsymbol{\Sigma}$, and $\widetilde{f}$ a zero mean stationary GMRF on the same torus with precision matrix $\boldsymbol{Q}$ parametrised with $\boldsymbol{\theta}$, then*

$$KL(f, \widetilde{f}) = -\frac{1}{2} \sum_{i=0}^{n_r-1} \sum_{j=0}^{n_c-1} \left(\log(\lambda_{ij} q_{ij}(\boldsymbol{\theta})) - \lambda_{ij} q_{ij}(\boldsymbol{\theta}) + 1\right) \tag{5}$$

where

$$\lambda_{ij} = \sum_{k=0}^{n_r-1} \sum_{l=0}^{n_c-1} \gamma(k,l) \exp\left(-2\pi\iota(ki/n_r + lj/n_c)\right), \tag{6}$$

$$q_{ij}(\boldsymbol{\theta}) = \sum\sum_{(k,l)\in\partial_{00}\cup\{00\}} Q_{00,kl}(\boldsymbol{\theta}) \exp\left(-2\pi\iota(ki/n_r + lj/n_c)\right) \tag{7}$$

and $\iota = \sqrt{-1}$.

PROOF: The proof uses standard results about Toeplitz circulant matrices, see Brockwell & Davis (1987). An $n \times n$ matrix $\boldsymbol{C}$ is Toeplitz circulant iff $C_{ij} = c_{i-j \bmod n}$ for some vector $\boldsymbol{c}$. The eigenvectors $\boldsymbol{V}$ and eigenvalues $\boldsymbol{D}$ for Toeplitz circulant matrices, $\boldsymbol{C} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^H$ (H denotes conjugate transpose), are explicitly known: $\boldsymbol{D} = \mathrm{diag}(d_0, \ldots, d_{n-1})$, where $d_i = \sum_{j=0}^{n-1} c_j \exp(-2\pi\iota ij/n)$, and $\boldsymbol{V}$ is the unitary matrix with entries $V_{ij} = n^{-1/2}\exp\left(-2\pi\iota ij/n\right)$. Hence, all Toeplitz circulant matrices have the same set of eigenvectors, and the eigenvalues are found by discrete Fourier transform of the first row of $\boldsymbol{C}$. Since $\boldsymbol{C}^{-1} = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{V}^H$, the inverse of a Toeplitz circulant matrix is also Toeplitz circulant. In our two-dimensional case, the covariance matrix of the Gaussian field $\boldsymbol{\Sigma}$ is a Toeplitz block circulant matrix and so is the precision matrix of the GMRF $\boldsymbol{Q}$. The result about Toeplitz circulant matrices generalises to Toeplitz circulant block matrices: the eigenvalues are the results of a two-dimensional discrete Fourier transform of the first line interpreted as an $n_\mathrm{r} \times n_\mathrm{c}$ matrix and all Toeplitz circulant block matrices have the same set of eigenvectors. Denote by λ_{ij} the eigenvalues of $\boldsymbol{\Sigma}$ and by $q_{ij}(\boldsymbol{\theta})$ the eigenvalues of $\boldsymbol{Q}$ which depend on the parametrisation $\boldsymbol{\theta}$. The expressions for λ_{ij} and $q_{ij}(\boldsymbol{\theta})$ are given in (6) and (7). Hence we have that

$$\begin{aligned}
\mathrm{KL}(f,\widetilde{f}) &= \int_{\boldsymbol{x}} f(\boldsymbol{x}) \log\left(\frac{f(\boldsymbol{x})}{\widetilde{f}(\boldsymbol{x})}\right)\, d\boldsymbol{x} \\
&= -\log(|\boldsymbol{\Sigma}|^{1/2}|\boldsymbol{Q}|^{1/2}) \;-\frac{1}{2}\int_{\boldsymbol{x}} f(\boldsymbol{x})\boldsymbol{x}^T(\boldsymbol{\Sigma}^{-1} - \boldsymbol{Q})\boldsymbol{x}\, d\boldsymbol{x} \\
&= -\frac{1}{2}\sum_{i=0}^{n_\mathrm{r}-1}\sum_{j=0}^{n_\mathrm{c}-1} \log(\lambda_{ij} q_{ij}(\boldsymbol{\theta})) - \frac{1}{2}\sum_{i=0}^{n_\mathrm{r}-1}\sum_{j=0}^{n_\mathrm{c}-1}\left(\frac{1}{\lambda_{ij}} - q_{ij}(\boldsymbol{\theta})\right)\lambda_{ij}
\end{aligned} \tag{8}$$

and the result follows. ■

The result shows that we can express the Kullback-Leibler discrepancy between a Gaussian field and a GMRF in terms of the eigenvalues of the respective covariance and precision matrices. The eigenvalues are explicitly known as the two-dimensional discrete Fourier transform of the first row of $\boldsymbol{\Sigma}$ and $\boldsymbol{Q}(\boldsymbol{\theta})$, respectively, interpreted as an $n_\mathrm{r} \times n_\mathrm{c}$ array. All eigenvalues λ_{ij} and $q_{ij}(\boldsymbol{\theta})$ are real since $\boldsymbol{\Sigma}$ and $\boldsymbol{Q}(\boldsymbol{\theta})$ are symmetric. We can use a fast two-dimensional discrete Fourier transform algorithm and evaluate $\mathrm{KL}(f,\widetilde{f})$ in $\mathcal{O}(n_\mathrm{r}n_\mathrm{c}\log(n_\mathrm{r}n_\mathrm{c}))$ flops, and obvious even faster for successive evaluations with a bit more coding. We can then solve numerically the minimisation problem to obtain the best fit, $\boldsymbol{\theta}_{\mathrm{KL}}$,

$$\boldsymbol{\theta}_{\mathrm{KL}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=0}^{n_\mathrm{r}-1}\sum_{j=0}^{n_\mathrm{c}-1}\left[\lambda_{ij}q_{ij}(\boldsymbol{\theta}) - \log(q_{ij}(\boldsymbol{\theta}))\right], \qquad q_{ij}(\boldsymbol{\theta}) > 0,\ \forall i,j. \tag{9}$$

The constraint that all eigenvalues are positive is equivalent to positive definiteness. This constraint is linear in $\boldsymbol{\theta}$, however, the dimensions involved are usually so large that we can not treat it as a linear inequality constraint. We approach the constrained minimisation problem as an unconstrained one, using a simplex-based iterative algorithm adding a big penalty if some of the eigenvalues are negative. This approach works satisfactory in practice.

Figure 2 shows the fitted correlation functions $\widetilde{\rho}_{\mathrm{KL}}(k,l)$, along the main axis, for a GMRF with a 5×5 neighbourhood for the three Gaussian fields in Section 2 (exponential, Gaussian, and spherical) along the main axis with correlation length $r = 20$. (The fit along the diagonal axis is better due to the square neighbourhood.) Table 1 shows the upper right part (compare with Figure 1) of the coefficients in the fitted GMRF for the three correlations functions converted to the interpretable quantities $1/\sqrt{\theta_{00}}$ and $-\theta_{kl}/\theta_{00}$, $(k,l) \in \partial_{00}$. The match for the exponential correlation function is reasonably good, which is natural as the exponential correlation function has Markov property on the line. The Gaussian and spherical correlation functions are more difficult to fit, and the match is not satisfactory. The fitted GMRFs have not similar global behaviour as the corresponding Gaussian fields. Even though larger neighbourhoods improve the fit somewhat, the fundamental behaviour of $\widetilde{\rho}_{\mathrm{KL}}$ outside ∂_{00} is still present for all neighbourhoods of reasonable size. The question is whether the inaccurate match is due to the fitting procedure or a feature of the GMRF itself.

To gain more insight into the fitting procedure based on KL discrepancy, assume the GMRF is parametrised with $(2m+1)^2$ free parameters indexed as θ_{kl}, $(k,l) \in \partial_{00} \cup \{00\}$. The (k',l') component of the gradient of $\mathrm{KL}(f, \widetilde{f})$, evaluated at the optimal fit, becomes

$$\sum_{i=0}^{n_{\mathrm{r}}-1} \sum_{j=0}^{n_{\mathrm{c}}-1} \left(\lambda_{ij} - \frac{1}{q_{ij}(\boldsymbol{\theta}_{\mathrm{KL}})} \right) \frac{\partial}{\partial \theta_{k'l'}} q_{ij}(\boldsymbol{\theta}_{\mathrm{KL}}) = 0, \tag{10}$$

where

$$\frac{\partial}{\partial \theta_{k'l'}} q_{ij}(\boldsymbol{\theta}_{\mathrm{KL}}) = \exp\left(-2\pi\iota(k'i/n_{\mathrm{r}} + l'j/n_{\mathrm{c}})\right). \tag{11}$$

Using that all eigenvalues are real, (10) defines the inverse discrete Fourier transform of $\{\lambda_{ij}\}$ and $\{1/q_{ij}(\boldsymbol{\theta}_{\mathrm{KL}})\}$ at (k',l'). Since $\{1/q_{ij}(\boldsymbol{\theta}_{\mathrm{KL}})\}$ are the eigenvalues of the inverse precision matrix $\boldsymbol{Q}(\boldsymbol{\theta}_{\mathrm{KL}})$, (10) reads

$$\gamma(k',l') = \widetilde{\gamma}_{\mathrm{KL}}(k',l'), \qquad (k',l') \in \partial_{00} \cup \{00\}. \tag{12}$$

This result gives an alternative interpretation of the optimal fit using the KL discrepancy. All covariances within the neighbourhood $\partial_{00} \cup \{00\}$ are fitted exactly and all the remaining ones are determined by invertion of the fitted precision matrix. (Dempster (1972) shows this result for maximum-likelihood estimators of the precision matrix in general, see also Whittaker (1990) for a recent treatment.) Using (12) we can now better understand the results in Figure 2. Only correlations within the 5×5 neighbourhood are fitted and these determine the behaviour of the fitted correlation function for larger lags. All Gaussian fields with similar correlations within the neighbourhood will be fitted with the same GMRF. Hence, we can not expect to fit the correlation function for a Gaussian field with a GMRF for lags not in $\partial_{00} \cup \{00\}$ using the KL discrepancy. This fact has implications for likelihood based inference for GMRF from real data which we discuss more in Section 5.

4 Fitting a GMRF using matched correlation function

In this section we introduce our criterion for fitting a GMRF to a stationary Gaussian field on a torus. The criterion takes all lags of the correlation function into account and is therefore able to fit both local and global properties.

The KL discrepancy fit the variance and the correlation function $\widetilde{\rho}(k,l)$ for the GMRF to $\rho(k,l)$ using only lags in ∂_{00}. We extend this idea to take also lags outside ∂_{00} into account. We define

$$\mathcal{D}(f,\widetilde{f}) = \sum_{(k,l)\in\Lambda\setminus\{00\}}\sum (\rho(k,l) - \widetilde{\rho}(k,l))^2\, W(k,l,\rho(k,l)), \tag{13}$$

where $W(k,l,\rho(k,l)) \geq 0$ is a user-defined weight function. Our criterion does not depend on the variances of the two fields, but the variance is trivial to fit exactly via a scaling.

$\mathcal{D}(f,\widetilde{f})$ defines a proper metric between the correlation functions ρ and $\widetilde{\rho}$ if we restrict the weight function $W(k,l,\rho(k,l))$ not to depend on $\rho(k,l)$ and to be strictly positive for all (k,l). Since the correlation function uniquely identifies the Gaussian density scaled to have unit variance, then $\mathcal{D}(f,\widetilde{f})$ is also a proper metric between the densities scaled to have unit variance in this case. In other cases, $\mathcal{D}(f,\widetilde{f})$ has the interpretation of a discrepancy between f and $\widetilde{f}$.

We can compute the correlation function $\widetilde{\rho}$ of our GMRF via the covariance function $\widetilde{\gamma}$ which is given as the inverse discrete two-dimensional Fourier transform of $\{q_{ij}(\boldsymbol{\theta})\}$,

$$\widetilde{\gamma}(k,l) = \frac{1}{n_{\mathrm{r}}n_{\mathrm{c}}}\sum_{i=0}^{n_{\mathrm{r}}-1}\sum_{j=0}^{n_{\mathrm{c}}-1}\frac{1}{q_{ij}(\boldsymbol{\theta})}\exp\left(2\pi\iota(ki/n_{\mathrm{r}} + lj/n_{\mathrm{c}})\right). \tag{14}$$

Hence, for reasonable choices of n_{r} and n_{c} we can compute (13) using $\mathcal{O}(n_{\mathrm{r}}n_{\mathrm{c}}\log(n_{\mathrm{r}}n_{\mathrm{c}}))$ flops.

The variance of a GMRF can always be fitted exactly by a scaling of the precision matrix. Therefore we minimise $\mathcal{D}(f,\widetilde{f})$ with respect to $\boldsymbol{\theta}$ for fixed $\theta_{00} = 1$, and then scale the solution to fit the variance. We approach the constrained minimisation problem in the same manner as in Section 3 adding a large penalty if $\boldsymbol{\theta}$ does not define a positive definite precision matrix, i.e. gives at least one non-positive eigenvalue.

The KL discrepancy is a special case of (13) choosing $W(k,l,\rho(k,l))$ positive if $(k,l) \in \partial_{00}$ and zero otherwise. We choose

$$W(k,l,\rho(k,l)) = \frac{1}{2\pi d(k,l)}, \tag{15}$$

where $d(k,l)$ is the Euclidean distance on the torus between (k,l) and $(0,0)$. With this choice we obtain a fitted GMRF with good average properties, as the fitted correlation function should be close to $\rho(k,l)$ for all lags. Further, $\mathcal{D}(f,\widetilde{f})$ is a proper metric. With (15) the contribution along each radii has roughly the same weight. Other choices are indeed possible and the weight function should be chosen by considering the application at hand.

We have fitted the exponential, Gaussian and spherical correlation functions with ranges 10, 20, and 30, to a GMRF with neighbourhoods from 3×3 up to 9×9 using the $\mathcal{D}(f,\widetilde{f})$ criterion. The presented results are for a torus size of 256×256, but we also tried larger arrays and the results were almost identical. The results are displayed in Figures 3, 4 and 5, one figure for each of the three correlation functions. The figures show the differences between fitted and target correlation functions. Again the fit along the diagonal axis is better compared to the horizontal axis due to the square neighbourhood.

Similar to the GMRF on the line, the auto-regressive process, the correlation function of a GMRF on a torus consists of exponentials and exponential decaying sin's, see for example Besag & Kooperberg (1995). The exponential correlation function is therefore the easiest case, providing a near perfect match for a neighbourhood larger than 3×3. The Gaussian and spherical correlation functions represent difficult cases, as they consist of exponential decaying $\sin$'s of all orders. Despite this fact, the fit is surprisingly accurate even for a 5×5 neighbourhood. The accuracy of the fit increases as the neighbourhood increases, but so does the computational cost related to the GMRF. Hence, the optimal size of the neighbourhood depends on the application. If we compare with the (lack of) accuracy as is common when fitting correlation functions to real data, our fitted GMRF could equally well be used instead of the fitted Gaussian field.

Table 2 shows the upper right part (compare with Figure 1) of the coefficients in the fitted GMRF for the three correlations functions with range 20 and a 5×5 neighbourhood, converted to the interpretable quantities $1/\sqrt{\theta_{00}}$ and $-\theta_{kl}/\theta_{00}$, $(k, l) \in \partial_{00}$. We find the full conditionals counter intuitive, when we take into account that the correlation functions we are trying to match are all positive. The full conditionals contain both positive and negative conditional correlations and the different patters are not easy to interpret. The coefficients are not easier to interpret for larger neighbourhoods. Thus, our intuition of the full conditionals can not be trusted. Therefore, we suggest that modelling of a Gaussian spatial varying variable with Markov property (from prior knowledge only), should be done through a correlation function and then fitted to a GMRF with a prescribed neighbourhood using the $\mathcal{D}(f, \widetilde{f})$ criterion for numerical minimisation. Specifying a joint density implicitly via the full conditionals is common in modern statistical modelling, and sometimes there is no viable alternative. However, our examples show that we should be careful about trusting to much in our intuition about the global properties of conditional specified models. To obtain an accurate fit of the Gaussian field, we need more accurate coefficients than those displayed in Table 2.

Note that the $\boldsymbol{Q}$ matrix derived from the coefficients in Table 2 are far from being diagonal dominant, $Q_{ij,ij} > \sum_{kl \neq ij} |Q_{ij,kl}|$. Diagonal dominance is a local property and is a sufficient but not necessary criterion for a matrix to be positive definite (which is a global property). Hence, if we only consider diagonal dominant parametrisations we do not seem to get the interesting models: none of our fits are diagonal dominant.

5 A comment on likelihood based inference for GMRF and Markov fields

Combining our findings in Sections 3 and 4, we now discuss likelihood based inference from real data in GMRF and in more general Markov field models.

The first observation is that our fitted GMRF using the Kullback-Leibler discrepancy is the limiting maximum-likelihood estimator for the parameters in the GMRF when the number of replicates of the Gaussian field tends to infinity. Unless the observed data are realisations of a GMRF, we can not expect the estimated correlation function outside ∂_{00} to follow the empirical correlation function found from the data, see Figure 2.

The second observation is that a GMRF can fit a Gaussian field surprisingly well using the proposed criteria (13) rather than the Kullback-Leibler discrepancy. As the correlation function identifies a zero mean and standardised Gaussian density, we have two densities close in norm but the limiting maximum-likelihood estimators of the parameters in the GMRF with data from these two densities are quite different (compare Tables 1 and 2). Hence, the maximum-likelihood estimators for GMRF seems not to be robust for model error.

From our two observations we conclude that likelihood based inference for parameters in GMRFs are questionable for real data, if (also) the global properties of the estimated GMRFs are important. We have investigated empirically the use of (13) as a criterion for estimating the parameters in a GMRF. Although we loose some efficiency if the GMRF is the true model, we gain a lot if the data is a Gaussian field close in norm to the GMRF: The estimated correlation function using (13) have a reasonable behaviour compared to the empirical, but the parameters in the GMRF have a larger variance if the GMRF is the wrong model.

We believe that a similar behaviour is also present in many other Markov fields estimating parameters from real data using maximum-likelihood. (Although this is computational difficult, Tjelmeland & Besag (1998) shows that it is feasible.) Our argument goes as follows.

Consider a Markov field on an exponential family canonical form

$$\pi(\boldsymbol{x}) \propto \exp\left(\sum_{k=1}^{K} \beta_k \phi_k(\boldsymbol{x})\right), \tag{16}$$

where the complete sufficient statistics $\{\phi_k(\boldsymbol{x})\}$ are known functions and $\boldsymbol{\beta} = \{\beta_k\}$ are unknown parameters. The equations for maximum likelihood can then be formulated as

$$\mathrm{E}_{\boldsymbol{\beta}}\left(\phi_k(\boldsymbol{x})\right) = \frac{1}{M}\sum_{m=1}^{M} \phi_k(\boldsymbol{x}^{(m)}), \qquad k = 1, \dots, K \tag{17}$$

where $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(M)}$ denotes the observed scenes. As an example, Tjelmeland & Besag (1998) used $\phi_k(\boldsymbol{x})$ equal to number of clique configurations within specified sets to model binary Markov fields. For a zero mean, homogeneous GMRF, (17) becomes

$$\mathrm{E}_{\boldsymbol{\beta}}\left(\sum_{(i,j)\in\Lambda}\sum x_{ij}x_{i+k,j+l}\right) = \frac{1}{M}\sum_{m=1}^{M}\left[\sum_{(i,j)\in\Lambda}\sum x_{ij}^{(m)}x_{i+k,j+l}^{(m)}\right] \tag{18}$$

for all $(k,l) \in \partial_{00} \cup \{00\}$. Note that (18) corresponds to the previous found expression in (12) as $M \to \infty$, and again, the GMRF uses only second order moments within the neighbourhood to estimate the second order properties for all lags. Similarly, (17) shows that only the empirical averages of the local statistics $\phi_k(\boldsymbol{x})$ are used to estimate $\boldsymbol{\beta}$, and hence these statistics also implicitly define the global properties of the Markov field.

We do expect that local Markov fields are able to fit also global properties to some extent, but by using likelihood based inference one completely ignores the global properties in the estimation process. Hence, we believe that criteria similar to (13) will improve on likelihood based approaches to fit both local and global properties with local Markov models. For the estimation problem of Tjelmeland & Besag (1998), for example, one could in addition to terms corresponding to the sufficient statistics also include pairwise (or higher order) interaction terms for lags outside the neighbourhood. However, more research along these lines are needed.

REFERENCES

BESAG, J. & KOOPERBERG, C. (1995). On conditional and intrinsic autoregression, *Biometrika* 82(4): 733–746.

BROCKWELL, P. J. & DAVIS, R. A. (1987). *Time Series: Theory and Methods*, Springer, Berlin.

CRESSIE, N. A. C. (1993). *Statistics for spatial data*, 2 edn, John Wiley, New York.

DEMPSTER, A. P. (1972). Covariance selection, *Biometrics* 28(1): 157–175.

GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. (1996). *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.

RUE, H. (1999). A fast and exact simulation algorithm for general Gaussian Markov random fields, *Statistics no. 8*, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

RUE, H., MARTHINSEN, A. & HUSBY, E. G. (1999). A split and conquer algorithm for exact simulation of general Gaussian Markov random fields, *Statistics no. 11*, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

TJELMELAND, H. & BESAG, J. (1998). Markov random fields with higher order interactions, *Scandinavian Journal of Statistics* 25(3): 415–433.

WHITTAKER, J. (1990). *Graphical models in applied multivariate statistics*, Wiley series in probability and mathematical statistics. Probability and mathematical statistics, Chichester: Wiley.

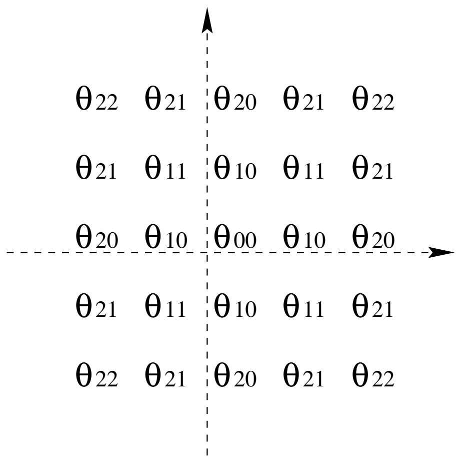
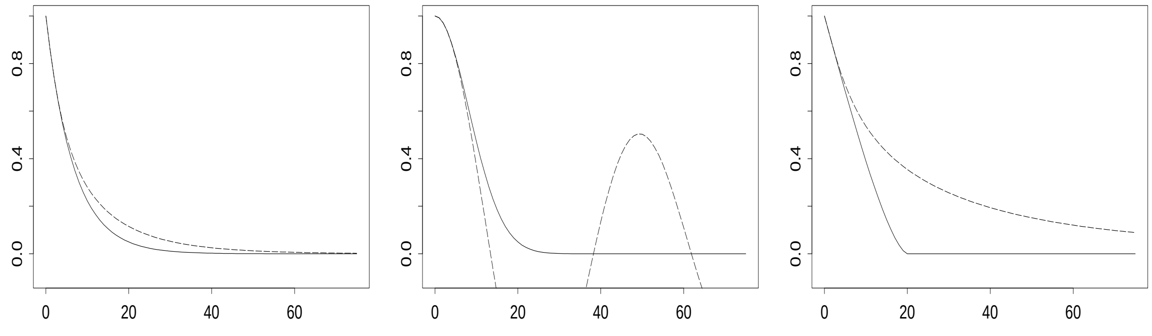
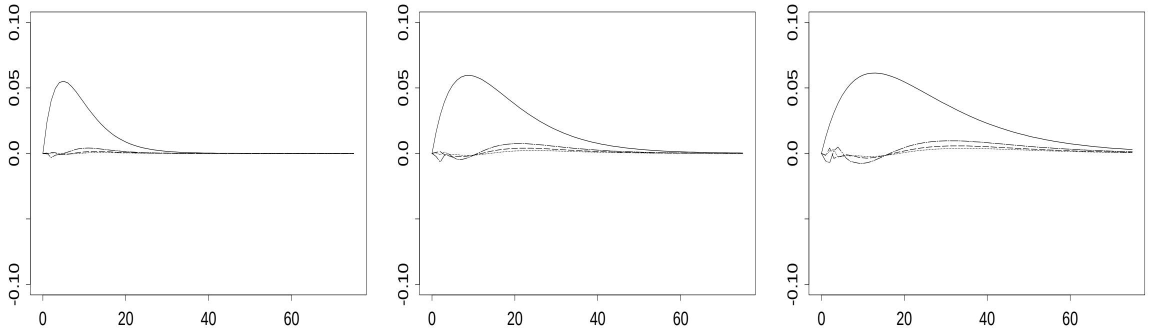


FIGURE 1: The 6 parameters for a GMRF with a 5×5 neighbourhood, labelled as $\theta_{00}, \theta_{10}, \theta_{20}, \theta_{11}, \theta_{21}$ and θ_{22}.



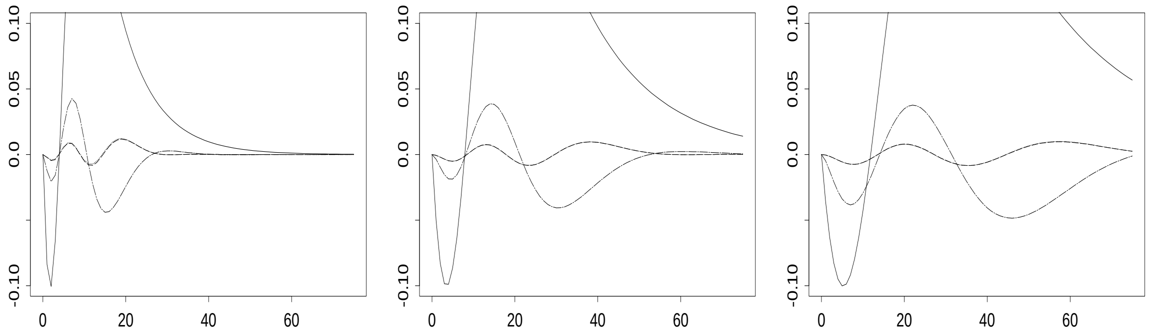
(a) (b) (c)

FIGURE 2: Target (solid) and fitted (dashed) correlation functions for 5×5 neighbourhood using $\mathrm{KL}(f, \widetilde{f})$ for the Gaussian field with (a) exponential, (b) Gaussian and (c) spherical correlation functions with correlation length $r = 20$.



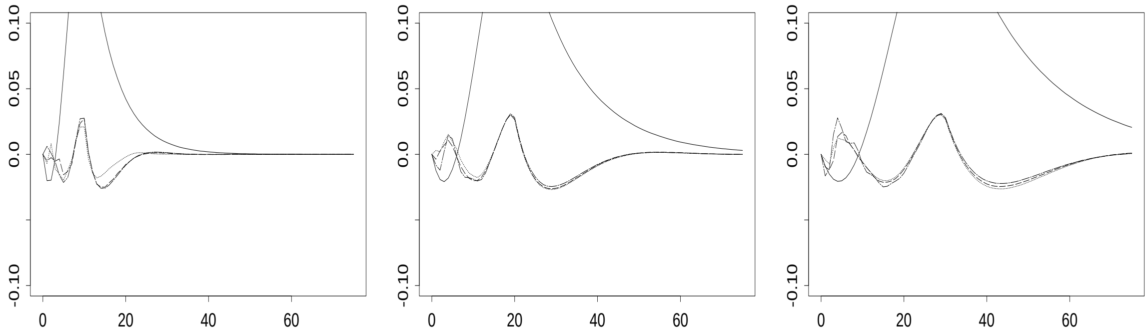
(a) (b) (c)

FIGURE 3: Difference between fitted and target correlation functions for a Gaussian field with exponential correlation function. Figure (a) is for correlation length $r = 10$ and with neighbourhood sizes 3×3 (solid), 5×5 (dash-dot), 7×7 (dashed) and 9×9 (dotted). Figures (b) and (c) shows similar results for range $r = 20$ and $r = 30$, respectively.



(a) (b) (c)

FIGURE 4: Difference between fitted and target correlation functions for a Gaussian field with Gaussian correlation function. Figure (a) is for correlation length $r = 10$ and with neighbourhood sizes 3×3 (solid), 5×5 (dash-dot), 7×7 (dashed) and 9×9 (dotted). Figures (b) and (c) shows similar results for range $r = 20$ and $r = 30$, respectively.



(a) (b) (c)

FIGURE 5: Difference between fitted and target correlation functions for a Gaussian field with spherical correlation function. Figure (a) is for correlation length $r = 10$ and with neighbourhood sizes 3×3 (solid), 5×5 (dash-dot), 7×7 (dashed) and 9×9 (dotted). Figures (b) and (c) shows similar results for range $r = 20$ and $r = 30$, respectively.

-0.027	-0.012	-0.011	-0.033	-0.006	0.013	-0.020	-0.007	-0.021
0.258	0.054	-0.012	0.475	-0.192	-0.006	0.267	0.037	-0.007
0.339	0.258	-0.027	0.007	0.475	-0.033	0.221	0.267	-0.020
Exponential			Gaussian			Spherical		

TABLE 1: Coefficients in the fitted GMRF using the $\mathrm{KL}(f, \widetilde{f})$ criterion and a 5×5 neighbourhood for exponential, Gaussian and spherical correlation function with range $r = 20$. The table shows the upper right part of Figure 1 and displays the conditional standard deviation and conditional correlations.

-0.150	0.080	-0.060	0.026	-0.051	0.037	-0.349	0.232	-0.117
0.285	0.015	0.080	0.471	-0.183	-0.051	0.495	-0.243	0.232
0.330	0.285	-0.150	0.014	0.471	0.026	0.138	0.495	-0.349
Exponential			Gaussian			Spherical		

TABLE 2: Coefficients in the fitted GMRF using the $\mathcal{D}(f, \widetilde{f})$ criterion and a 5×5 neighbourhood for exponential, Gaussian and spherical correlation function with range $r = 20$. The table shows the upper right part of Figure 1 and display the conditional standard deviation and conditional correlations.