DEPT. OF MATH. UNIVERSITY OF OSLO
STATISTICAL RESEARCH REPORT No. 7
ISBN 82–553–1317–6 November 2001
ISSN 0806–3842

Stationary Space Time Gaussian Fields and their time autoregressive representation

Geir Storvik*, Arnoldo Frigessi and David Hirst

Email: geirs@math.uio.no

Abstract

We compare two different modelling strategies for continuous space discrete time data. The first strategy is in the spirit of Gaussian kriging. The model is a general stationary space-time Gaussian field where the key point is the choice of a parametric form for the covariance function. Mostly, covariance functions which are used are separable in space and time. Non-separable covariance functions are useful in many applications, but construction of such is not easy.

The second strategy is to model the time-evolution of the process more directly. We consider models of the autoregressive type where the process at time t is obtained by convolving the process at time $t-1$ and adding some noise which are spatially correlated.

Under specific conditions, the two strategies describe two different formulations of the same stochastic process. We show how the two representations look in different cases. Furthermore, by transforming time-dynamic convolution models to Gaussian fields we can obtain new covariance functions and by writing a Gaussian field as a time-dynamic convolution model, interesting properties are discovered.

The computational aspects of the two strategies are discussed through experiments on a dataset of daily UK temperatures. Although algorithms for performing estimation, simulation and so on are easy to do for the first strategy, more computer efficient algorithms based on the second strategy can be constructed.

***Keywords*— space-time covariance structure, non-separable; stationary Gaussian Fields, convolution models, spectral representation, kalman filter**

Revised version submitted to Stochastical modelling

*Geir Storvik is an Associate Professor at the University of Oslo and part time researcher at the Norwegian Computing Center, Arnoldo Frigessi is a Chief Research Scientist at the Norwegian Computing Center and a part time Professor at the University of Oslo, David Hirst is a Senior Research Scientist at the Norwegian Computing Center.

1 Introduction

Gaussian random fields are the most common spatial models. They are either used for modelling the observed data directly [Ripley, 1981, Cressie, 1991] or as building blocks in hierarchical models [Diggle et al., 1998, KnorrHeld and Besag, 1998]. Recently, there has been a lot of interest in extending these models to spatio-temporal processes, where one of the main difficulties lies in specifying the space-time covariance structure. In this paper, we compare two different approaches. The first is similar to strategies used for spatial processes, in that parametric forms for the covariance functions are directly defined [Carroll et al., 1997, Jones and Zhang, 1997, Cressie and Huang, 1999]. The second strategy is to introduce latent structures that generate space-time correlation [Haas, 1995, Høst et al., 1995, Sølna and Switzer, 1996, Wikle and Cressie, 1999], and to assume that the residual noise has a much simpler coloring than the original process.

To be more precise, assume

$$y(t, \mathbf{x}) = \mu(t, \mathbf{x}) + z(t, \mathbf{x})$$

where t is time and $\mathbf{x}$ is the space variable. We assume $\mathbf{x} \in \mathcal{R}^2$. Here $\mu(t, \mathbf{x})$ is some deterministic regression part which we equate to zero for simplicity. $z(t, \mathbf{x})$ is a zero-mean time and space stationary stochastic field with covariance function $c(s; \mathbf{u})$, where s and $\mathbf{u}$ are temporal and spatial lags, respectively.

In the first modelling strategy, a specific parametric form for $c(\cdot; \cdot)$ is assumed. Care has to be taken in order to derive a non-negative definite function. Different classes of such functions have been presented in the literature. These are mostly either isotropic in all dimensions, making no distinction between time and space, or separable (i.e. $c(s; \mathbf{u}) = c_1(s)c_2(\mathbf{u})$). Recently Cressie and Huang [1999] have presented some non-separable space-time covariance functions. In practice it is always difficult to justify any particular form, especially with non-equally spaced data.

In the second modelling strategy the stochastic dependence is defined through latent variables which have a simpler dependence structure than $z(t, \mathbf{x})$ itself. One possibility is to make an explicit assumption about the time-evolution:

$$z(t, \mathbf{x}) = g(\mathbf{z}(t-1, \cdot), \varepsilon(t, \mathbf{x})) \tag{1}$$

where $\mathbf{z}(t-1, \cdot) = \{z(t-1, \mathbf{x}); \mathbf{x} \in \mathcal{R}^2\}$ is the spatial process at time $t-1$ and $\{\varepsilon(t, \mathbf{x})\}$ is a further stationary process. This approach can be natural if data are available at discrete equidistant time points. If we further assume g to be linear and $\varepsilon(t, \mathbf{x})$ to be Gaussian, then $\{z(t, \mathbf{x})\}$ will be a Gaussian field, with some covariance function. This will be the case of our interest. Note however that this second type of model can be extended to non-Gaussian noise, higher order temporal dependence, etc.

In this paper we compare these two strategies. In particular we consider interpretability, model checking, inference and computational difficulties.

In what follows we compute the covariance function of (1) explicitly for the linear-Gaussian case. We are then able to say what type of (non-separable) covariance we are implicitly assuming while using model (1). This provides us with a way to build non-separable covariance functions that are interpretable in term of the time-dynamic representation. We also proceed in the opposite direction. We take a Gaussian field with a non-separable covariance function (as proposed by Cressie and Huang [1999] for example) and see first if it can be represented as a linear-Gaussian version of the time-dynamic model (1) and second, if it can, we compute the covariance function of the noise ε and the linear transform from time $t-1$ to t. Such a representation can be used for interpretation of the Gaussian field.

Whenever there are two different modelling strategies that lead to the same model for the data, the following question arises: In practice can we consider only one of the two approaches (i.e. covariance or latent variables modelling)? Our findings clearly lead to a suggestive answer since each of the two modelling strategies can lead to models that would not likely be proposed when following the other approach.

We also note that there are Gaussian models that cannot be represented in the form (1) and in this case we explain why.

The outline of the paper is as follows: In section 2 we define our fields and fix notation. Section 3 consider the linear-Gaussian version of the time-dynamic model (1). In section 4 conditions for equality between the two representations are derived, while section 5 gives illustrative examples. A discussion on the two representations related to numerical computation is given in section 6.

2 Gaussian fields

A stationary, zero mean space time Gaussian field $z(t, \mathbf{x})$ is specified by the stationary covariance function

$$c(s; \mathbf{u}) = \mathsf{cov}[z(t, \mathbf{x}), z(t+s, \mathbf{x}+\mathbf{u})]$$

where s and $\mathbf{u}$ are, respectively, the time and spatial lags. Here we assume that t is continuous and real and $\mathbf{x}$ is, say, in $\mathcal{R}^2$. Covariance functions need to be positive definite. For this reason valid space time covariance functions are not easy to construct. Assuming the covariance function is continuous, positive definiteness is equivalent to the process having a spectral distribution function [Matérn, 1986][p. 12]. Assuming further the covariance function to be isotropic in the full spatio-temporal domain, all valid functions can

be represented as (possible infinite) mixtures of Bessel functions [Yaglom, 1987][p. 106]. Such representations have been used for defining "nonparametric" covariance functions (see Ecker and Gelfand [1997] and the references therein). Because the time-dimension has a different interpretation compared to the spatial ones, an isotropic assumption is often unrealistic. Other approaches are therefore needed for specifying covariance functions in space and time.

One simple way to build such a covariance function is to multiply a time stationary covariance function $c_1(s)$ with a space stationary covariance function $c_2(\mathbf{u})$

$$c(s;\mathbf{u}) = c_1(s)c_2(\mathbf{u}).$$

Such a space time covariance function is called *separable*. As discussed in Cressie and Huang [1999] the class of separable space time covariance functions is quite limited. In many applications we need more general types of correlations. Jones and Zhang [1997] and Cressie and Huang [1999] introduce new classes of covariance functions for stationary spatio-temporal processes, but the tool box is still limited.

The Cressie-Huang class will be of particular interest in the following. For every fixed $s \in \mathcal{R}$ let $\Gamma(s,\boldsymbol{\omega})$ be a function over the (spatial) frequency domain. If one can write

$$\Gamma(s,\boldsymbol{\omega}) = \rho(s,\boldsymbol{\omega})k(\boldsymbol{\omega})$$

and if the two following conditions are satisfied;

(CH1) For each $\boldsymbol{\omega} \in \mathcal{R}^2, \rho(\cdot;\boldsymbol{\omega})$ is a continuous autocorrelation function.

(CH2) $k(\boldsymbol{\omega}) > 0$ for all $\boldsymbol{\omega}$ and $\int k(\boldsymbol{\omega})d\boldsymbol{\omega} < \infty$,

then $\Gamma(s,\boldsymbol{\omega})$ can be Fourier transformed (over the spatial frequency $\boldsymbol{\omega}$) to produce a valid covariance function $c(s;\mathbf{u})$.

3 Time autoregressive spatial models

A different way to introduce space-time correlation is by means of time-dynamical hierarchical models as defined by (1). We will consider the special case

$$z(t,\mathbf{x}) = \int_{\mathcal{R}^2} h(\mathbf{v})z(t-1,\mathbf{x}+\mathbf{v})d\mathbf{v} + \varepsilon(t,\mathbf{x}), \quad \mathbf{x} \in \mathcal{R}^2, t = ...,-1,0,1,... \tag{2}$$

We name this a *time autoregressive spatial model* (of order 1). The spatial *convolution kernel* $h(\mathbf{v})$ has to be chosen to guarantee stationarity of $z(t,\mathbf{x})$. It decays rapidly from its

mode located in the origin. Often it is assumed that $h(\mathbf{v}) = h(||\mathbf{v}||)$. For instance it could be a Gaussian kernel. The noise process $\{\varepsilon(t, \mathbf{x})\}$ is assumed to be a stationary Gaussian field, uncorrelated in time but colored in space, with spatial covariance function $r(\mathbf{u})$, so that

$$\mathsf{cov}[\varepsilon(t, \mathbf{x}), \varepsilon(t', \mathbf{x} + \mathbf{u})] = \begin{cases} 0, & \text{if } t \neq t'; \\ r(\mathbf{u}), & \text{otherwise} \end{cases}$$

Define $H(\boldsymbol{\omega})$ to be the inverse Fourier transform of $h(\mathbf{v})$. The following theorem states conditions for when the process (2) is stationary in time.

Theorem 1
Assume that $|H(\boldsymbol{\omega})| < 1$ *for all* $\boldsymbol{\omega}$. *Assume further that* $\mathsf{var}[z(0, \mathbf{x})] < K$ *for all* $\mathbf{x}$. *Then, considering* $\{z(t, \mathbf{x})\}$ *as a process evolving in time, a unique limit distribution exists. Further, this distribution is stationary in space.*

Proof: See appendix A.

A corollary, which will be important for performing numerical computation is

Corollary 1
Given that the process $\{z(t, \mathbf{x})\}$ *is in its limit distribution, an alternative representation of the process is*

$$z(t, \mathbf{x}) = \int_{\omega} e^{i\boldsymbol{\omega}'\mathbf{x}} dZ(t, \boldsymbol{\omega}) \tag{3}$$

where

$$dZ(t, \boldsymbol{\omega}) = H(\boldsymbol{\omega}) dZ(t-1, \boldsymbol{\omega}) + dE(t, \boldsymbol{\omega}). \tag{4}$$

Proof: Follows directly from the proof of Theorem 1

Assuming the process defined by (2) is stationary with finite first and second order moments, its covariance function $c(s; \mathbf{u})$ is defined for all integer s, and all $\mathbf{u} \in \mathcal{R}^2$. It must satisfy (see appendix B)

$$c(0; \mathbf{u}) = \int_{\mathcal{R}^2} \int_{\mathcal{R}^2} h(\mathbf{v}) h(\mathbf{w}) c(0; \mathbf{u} + \mathbf{w} - \mathbf{v}) d\mathbf{v} d\mathbf{w} + r(\mathbf{u}), \qquad \forall \mathbf{u} \tag{5}$$

$$c(s; \mathbf{u}) = \int_{\mathcal{R}^2} h(\mathbf{v}) c(s-1; \mathbf{u} - \mathbf{v}) d\mathbf{v}, \qquad \forall \mathbf{u}, s = 1, 2, ... \tag{6}$$

Assume that the following functions are all well defined: the inverse Fourier transform $H(\boldsymbol{\omega})$ of the kernel function $h(\cdot)$; the inverse Fourier transform $R(\boldsymbol{\omega})$ of $r(\cdot)$; and the inverse Fourier transform $\Gamma(s; \boldsymbol{\omega})$ of $c(s; \cdot)$. Note that all (inverse) Fourier transforms are

taken in the spatial domain. Then, by taking inverse Fourier transforms on both sides of (5), we obtain

$$\Gamma(0;\boldsymbol{\omega}) = H(\boldsymbol{\omega})H(-\boldsymbol{\omega})\Gamma(0;\boldsymbol{\omega}) + R(\boldsymbol{\omega}).$$

This gives, for $H(\boldsymbol{\omega})H(-\boldsymbol{\omega}) \neq 1$,

$$\Gamma(0;\boldsymbol{\omega}) = \frac{R(\boldsymbol{\omega})}{1 - H(\boldsymbol{\omega})H(-\boldsymbol{\omega})}. \tag{7}$$

Similarly, by taking inverse Fourier transforms on both sides of (6), we obtain

$$\Gamma(s;\boldsymbol{\omega}) = \Gamma(0;\boldsymbol{\omega})H(-\boldsymbol{\omega})^{|s|}. \tag{8}$$

(Note that this equation is also valid for negative s since $c(-s,\mathbf{u}) = c(s,\mathbf{u})$.) The covariance function can then be obtained as the Fourier transform of $\Gamma(s;\boldsymbol{\omega})$ given by (7) and (8).

Under mild assumptions on $h(\cdot)$ and $r(\cdot)$, discussed in Theorem 2, the inverse Fourier transform $\Gamma(s;\boldsymbol{\omega})$ exists, and the covariance function of (2) can be computed at least numerically. In some cases it is possible to compute $c(s;\mathbf{u})$ analytically, as in the following example.

Example 1
Assume in (2) an isotropic Matern function for $h(\cdot)$

$$h(\mathbf{u}) = \pi^{-1}\theta a(\nu_1)\|\mathbf{u}\|^{\nu_1}K_{\nu_1}(\|\mathbf{u}\|),\ \nu_1 > 0, 0 < \theta < 1,$$

where $K_\nu(\cdot)$ is the modified Bessel function of the third kind of order ν while

$$a(\nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)}.$$

Assume further a difference between two Matern functions for $r(\cdot)$:

$$r(\mathbf{u}) = a(\nu_2)\|\mathbf{u}\|^{\nu_2}K_{\nu_2}(\|\mathbf{u}\|) - \theta^2 a(\nu_2 + 2\nu_1 + 2)\|\mathbf{u}\|^{\nu_2+2\nu_1+2}K_{\nu_2+2\nu_1+2}(\|\mathbf{u}\|)$$

where $\nu_2 > 0$ Then we obtain

$$\begin{aligned} H(\boldsymbol{\omega}) =& \theta(1 + \|\boldsymbol{\omega}\|^2)^{-\nu_1-1} \\ R(\boldsymbol{\omega}) =& \pi(1 + \|\boldsymbol{\omega}\|^2)^{-\nu_2-1} - \pi\theta^2(1 + \|\boldsymbol{\omega}\|^2)^{-(\nu_2+2\nu_1+2)-1}; \\ =& \pi(1 + \|\boldsymbol{\omega}\|^2)^{-\nu_2-1}[1 - \theta^2(1 + \|\boldsymbol{\omega}\|^2)^{-2(\nu_1+1)}], \end{aligned}$$

which gives

$$\Gamma(s;\boldsymbol{\omega}) = \pi\theta^{|s|}(1 + \|\boldsymbol{\omega}\|^2)^{-(\nu_2+(\nu_1+1)|s|)-1} \tag{9}$$

for $s = 0, \pm1, \pm2, ...$ This can be Fourier transformed so that we obtain

$$c(s;\mathbf{u}) = \theta^{|s|}a(\nu_2 + (\nu_1 + 1)|s|)\|\mathbf{u}\|^{\nu_2+(\nu_1+1)|s|}K_{\nu_2+(\nu_1+1)|s|}(\|\mathbf{u}\|) \tag{10}$$

For every fixed s this is a Matern covariance function. □

The example above illustrates an important issue. The time lag 1 autoregressive model (2) is defined only for $s \in \mathcal{Z}$. However, the covariance function $c(s; \mathbf{u})$ in (10) can be extended to continuous values $s \in \mathcal{R}$. It is a valid covariance function for a Gaussian space-time field defined in *continuous* time. This can be seen by noting that the covariance function is a member of the Cressie and Huang [1999] class (i.e. $\Gamma(s; \boldsymbol{\omega})$ given by (9) can be seen to satisfy the conditions (CH1) and (CH2)).

4 A double representation

We assume that the data $z(t, \mathbf{x})$ are available only at discrete, say integer, time points. One possible model is a time continuous spatial Gaussian field with a specific covariance function $c(s, \mathbf{u})$. Due to the discrete nature of the data, in the likelihood function the covariance function will appear only at integer time lags.

A second possible model is (2), for some kernel function $h(\cdot)$ and spatial covariance function $r(\cdot)$. This model is defined only in discrete time. However, as we saw in the example of the previous section, its space time covariance function can sometimes be extended to continuous time. In these cases there exists a continuous time spatial Gaussian field. The likelihood function of the data assuming model (2) or assuming such a corresponding continuous time spatial Gaussian field would be identical.

Hence there are cases in which the same statistical model can be described using two rather different forms: by means of a space-time covariance function $c(s, \mathbf{u})$ or through a spatial kernel $h(\mathbf{x})$ and a spatial covariance function $r(\mathbf{u})$. In this section we characterize precisely when such two representations are available by answering the two following questions:

1. When can the discrete time spatial covariance function of model (2) be extended to continuous time?
2. When can the likelihood of discrete time data, modeled as a continuous time spatial Gaussian field, be written using the time lag 1 autoregressive representation (2)?

The following theorem answers the first question.

Theorem 2
Assume the process defined by (2) *is stationary in time. Let the inverse Fourier transforms* $H(\boldsymbol{\omega})$ *of* $h(\cdot)$ *and* $R(\boldsymbol{\omega})$ *of* $r(\cdot)$ *be well defined. For* $s \in \mathcal{Z}$, *let*

$$\Gamma(s; \boldsymbol{\omega}) = \frac{R(\boldsymbol{\omega})}{1 - H(\boldsymbol{\omega})H(-\boldsymbol{\omega})} H(-\boldsymbol{\omega})^{|s|}. \tag{11}$$

If $|H(\boldsymbol{\omega})| < 1$ for all $\boldsymbol{\omega} \in \mathcal{R}$ and $\int \Gamma(0;\boldsymbol{\omega})d\boldsymbol{\omega} < \infty$, then the Fourier transform $c(s;\mathbf{u})$ of $\Gamma(s;\boldsymbol{\omega})$ exists for all $s \in \mathcal{Z}$ and can be extended to $s \in \mathcal{R}$. $c(s;\mathbf{u})$ is the covariance function of the stationary solution of (2). Such a space-(continuous) time covariance function defines a space-(continuous)time Gaussian field.

Proof: Note that since $R(\boldsymbol{\omega})$ is the spectral density of the stationary process $\{\varepsilon(t,\cdot)\}$, $R(\boldsymbol{\omega}) > 0$ for all $\boldsymbol{\omega}$. Combined with the condition that $|H(\boldsymbol{\omega})| < 1$, this implies that $\Gamma(0;\boldsymbol{\omega}) > 0$ for all $\boldsymbol{\omega}$. Further, since $|H(\boldsymbol{\omega})| < 1$,

$$\int |\Gamma(s;\boldsymbol{\omega})|d\boldsymbol{\omega} \leq \int |\Gamma(0;\boldsymbol{\omega})|d\boldsymbol{\omega} = \int \Gamma(0;\boldsymbol{\omega})d\boldsymbol{\omega} < \infty$$

showing that the Fourier transform of $\Gamma(s;\boldsymbol{\omega})$ exists for all integer s.

To show that this extends to $s \in \mathcal{R}$, we use the main result in Cressie and Huang [1999], which specifies a sufficient condition for a function $c(s,\mathbf{u})$ with $s \in \mathcal{R}$ to be positive definite. In our case we now take $k(\boldsymbol{\omega}) = \Gamma(0;\boldsymbol{\omega})$ and $\rho(s;\boldsymbol{\omega}) = H(-\boldsymbol{\omega})^{|s|}$. Since $|H(-\boldsymbol{\omega})| < 1$, $\rho(s;\boldsymbol{\omega})$ is integrable, implying that it is an autocorrelation function, making the (CH1) criterion satisfied. In fact it is the autocorrelation function for the time continuous autoregressive model of order one. Also $k(\boldsymbol{\omega}) = \Gamma(0;\boldsymbol{\omega}) > 0$. Finally $\int k(\boldsymbol{\omega})d\boldsymbol{\omega} < \infty$, is directly assumed. This makes also (CH2) satisfied implying that c is a valid covariance function. □

Remarks:

1. Concerning the integrability of $\Gamma(0;\boldsymbol{\omega})$, let $h(\mathbf{x})$ be symmetric and unimodal with mode at $\mathbf{x} = \mathbf{0}$. This is a common situation. Then $H(\boldsymbol{\omega}) \leq H(\mathbf{0})$ for all $\boldsymbol{\omega}$, so that

   $$\int \Gamma(0;\boldsymbol{\omega})d\boldsymbol{\omega} \leq \frac{1}{1 - H(\mathbf{0})^2} \int R(\boldsymbol{\omega})d\boldsymbol{\omega} = r(\mathbf{0}) < \infty.$$

   This shows that this condition is also often fulfilled.

2. Note that none of the calculations actually involves the assumption of a *Gaussian* process. Spatial and spatio-temporal covariance functions are however mostly used in connection with Gaussian processes.

The next theorem answers the second question.

Theorem 3
Assume $c(s,\mathbf{u})$ is a valid covariance function for $s \in \mathcal{R}$ and $u \in \mathcal{R}^2$. Assume that for fixed s the (spatial) Fourier transform $\Gamma(s;\boldsymbol{\omega})$ of $c(s,\cdot)$ exists. Assume one can write

$$\Gamma(s;\boldsymbol{\omega}) = \Gamma(0;\boldsymbol{\omega})H(-\boldsymbol{\omega})^{|s|} \tag{12}$$

for some function $H(\boldsymbol{\omega})$ with the following properties:

(i) The Fourier transform $h(\cdot)$ of $H(\cdot)$ exists and $|H(\boldsymbol{\omega})| < 1$ for all $\boldsymbol{\omega}$.

(ii) $\Gamma(0;\boldsymbol{\omega}) > 0$ for all $\boldsymbol{\omega}$, and $\int \Gamma(0;\boldsymbol{\omega})d\boldsymbol{\omega} < \infty$.

Then there exists a stochastic process $\{z(t;\mathbf{x})\}$ defined for $s \in \mathcal{Z}$ and for all $\mathbf{x} \in \mathcal{R}^2$ which follows model (2) with covariance structure equal to $c(s;\mathbf{u})$. For this process $h(\cdot)$ and $r(\cdot)$ are uniquely defined as the the Fourier transforms of $H(\cdot)$ and

$$R(\boldsymbol{\omega}) = \Gamma(0;\boldsymbol{\omega})(1 - H(\boldsymbol{\omega})H(-\boldsymbol{\omega})). \tag{13}$$

Proof: Since $|H(\boldsymbol{\omega})| < 1$, by (13) $|R(\boldsymbol{\omega})| < \Gamma(0;\boldsymbol{\omega})$. Furthermore, since $\int \Gamma(0;\boldsymbol{\omega})d\omega < \infty$, a similar property is valid for $R(\boldsymbol{\omega})$, which implies the existence of the Fourier transform $r(\mathbf{x})$ of $R(\boldsymbol{\omega})$.

Now define the process $\{z(t;\mathbf{x})\}$ for integer time points s through (2) using the given choices of $h(\cdot)$ and $r(\cdot)$. By (7) and (8), we easily see that the defined process has covariance function $c(s;\mathbf{u})$.

In order to prove uniqueness, assume there exist some other functions $\tilde{h}(\mathbf{x})$ and $\tilde{r}(\mathbf{x})$ which through (2) define a new process $\tilde{z}$ with the same covariance function c. Then also $\Gamma(s;\boldsymbol{\omega})$ must be common for the two processes. Define $\widetilde{H}(\boldsymbol{\omega})$ and $\widetilde{R}(\boldsymbol{\omega})$ to be the inverse Fourier transforms of $\tilde{h}(\mathbf{x})$ and $\tilde{r}(\mathbf{x})$, respectively. By (8)

$$\Gamma(0;\boldsymbol{\omega})H(-\boldsymbol{\omega})^{|s|} = \Gamma(0;\boldsymbol{\omega})\widetilde{H}(-\boldsymbol{\omega})^{|s|},$$

showing that $\widetilde{H}(\boldsymbol{\omega}) = H(\boldsymbol{\omega})$. Further, by (7)

$$\frac{\widetilde{R}(\boldsymbol{\omega})}{1 - \widetilde{H}(\boldsymbol{\omega})\widetilde{H}(-\boldsymbol{\omega})} = \frac{R(\boldsymbol{\omega})}{1 - H(\boldsymbol{\omega})H(-\boldsymbol{\omega})}.$$

Hence

$$\widetilde{R}(\boldsymbol{\omega}) = \frac{1 - \widetilde{H}(\boldsymbol{\omega})\widetilde{H}(-\boldsymbol{\omega})}{1 - H(\boldsymbol{\omega})H(-\boldsymbol{\omega})}R(\boldsymbol{\omega}) = R(\boldsymbol{\omega}),$$

completing the proof. □

Note that the uniqueness of $h(\cdot)$ and $r(\cdot)$ is established for the lag 1 autoregression model (2) for the *given time-discretization*. If the continuous time scale were to be discretized with different time steps, say $\Delta \neq 1$, then different $h(\cdot)$ and $r(\cdot)$ functions would be needed. More precisely, from (12) we may write for an arbitrary time scale Δ and integer s

$$\Gamma(s\Delta;\boldsymbol{\omega}) = \Gamma(0;\boldsymbol{\omega})H_\Delta(-\boldsymbol{\omega})^{|s|}$$

where $H_\Delta(-\boldsymbol{\omega}) = H(-\boldsymbol{\omega})^\Delta$. Furthermore, we can now write

$$z(t,\mathbf{x}) = \int_{\mathcal{S}} h_\Delta(\mathbf{u}) z(t-\Delta, \mathbf{x}+\mathbf{u}) d\mathbf{u} + \varepsilon_\Delta(t,\mathbf{x}) \tag{14}$$

where h_Δ is the inverse Fourier transform of H_Δ and the spatial covariance function $r_\Delta(\mathbf{u})$ of $\{\varepsilon_\Delta(t,\mathbf{x})\}$ is defined through its spectral density

$$R_\Delta(\boldsymbol{\omega}) = \Gamma(0;\boldsymbol{\omega})[1 - H_\Delta(\boldsymbol{\omega})H_\Delta(-\boldsymbol{\omega})].$$

Note however that h_Δ may not exist for small Δ even though it does for $\Delta = 1$. A very important result is given by Brown et al. [2000]. They show that if $h_\Delta(\mathbf{x})$ exists for all Δ, if $h_\Delta(\mathbf{x}) > 0 \forall \mathbf{x}, \Delta$, and if $h_\Delta(\cdot)$ all have finite second moments, then the kernel function $h_\Delta(\mathbf{x})$ has to be Gaussian. This means that if other convolution functions are to be used, all the conditions above can not be satisfied. As we will see in the examples of the following section, typically h_Δ will be negative in some areas for small values of Δ or it does not exist at all.

5 Examples

We have seen that in certain situations there are two different modelling strategies that lead to the same likelihood function. In this section we will see some examples. They show that both modelling approaches are important: sometimes models seem natural if expressed with one type of parameterization, and less so once written in the other way.

We begin with a model that has a simple separable space time covariance function $c(s;\mathbf{u})$ and look to its representation (2).

Example 2 (Separable covariance functions)
Assume

$$c(s;\mathbf{u}) = c_1(s)c_2(\mathbf{u}).$$

Then,

$$\Gamma(s;\boldsymbol{\omega}) = c_1(s)C_2(\boldsymbol{\omega}),$$

where $C_2(\boldsymbol{\omega})$ is the spectral density of $c_2(\mathbf{u})$. In order to represent this process as (2), by Theorem 3 since $c_1(s)$ does not depend on $\boldsymbol{\omega}$, it must be of the form $\theta^{|s|}$ with $|\theta| < 1$. Because $H(\boldsymbol{\omega}) = \theta$, it follows that $h(x) = \theta\delta(x)$ where $\delta(x)$ is Dirac's delta. Hence separable covariance functions correspond to the simple time autoregressive model

$$z(t,\mathbf{x}) = \theta z(t-1,\mathbf{x}) + \varepsilon(t,\mathbf{x}).$$

where $\{\varepsilon(t,\mathbf{x}\}$ is spatially correlated with covariance function given by $r(\mathbf{u}) = c_2(0)(1-\theta^2)c_2(\mathbf{u})$. □

Example 3
We shall use Theorem 3 to obtain the time autoregressive representation of a slightly modified version of example 2 of Cressie and Huang [1999]. Let

$$c(s;\mathbf{u}) = \theta^{|s|}\frac{c_0}{|s|+c_0}e^{-||\mathbf{u}||^2 c_0/(|s|+c_0)}. \tag{15}$$

Cressie and Huang [1999] considered the case $\theta = 1$ (with a slightly different parametrization). Its inverse (spatial) Fourier transform is given by

$$\Gamma(s;\boldsymbol{\omega}) = (4\pi)^{-1}e^{-||\boldsymbol{\omega}||^2/4}[\theta e^{-||\boldsymbol{\omega}||^2/4c_0}]^{|s|}$$

The space-time covariance function (15) is non-separable. We take

$$H(\boldsymbol{\omega}) = \theta e^{-||\boldsymbol{\omega}||^2/4c_0}, \tag{16}$$

so that the conditions in Theorem 3 are fulfilled for $\theta < 1$. From equation (13) we have that

$$R(\boldsymbol{\omega}) = (4\pi)^{-1}e^{-||\boldsymbol{\omega}||^2}[1-\theta^2 e^{-||\boldsymbol{\omega}||^2/2c_0}], \tag{17}$$

giving

$$h(\mathbf{u}) = 4\theta e^{-c_0||\mathbf{u}||^2}$$

and

$$r(\mathbf{u}) = c_0 e^{-||\mathbf{u}||^2/c_0}[1-\theta^2\frac{c_0}{[2+c_0]\pi}e^{-||\mathbf{u}||^2 c_0/[2+c_0]}.$$

The kernel function $h(\cdot)$ turns out to be Gaussian, which is often the natural choice. However the corresponding spatial covariance function $r(\cdot)$ is rather unusual. In Figure 1 $r(\mathbf{u})$ is plotted for $c_0 = 1$ and different values of θ. The covariance function $r(\cdot)$ becomes negative for larger spatial lags. This is the case for any choice of c_0 and θ. $c_0 = \infty$ corresponds to a separable model.

We interpret the findings of this example in the following way. While it can be appropriate and natural to consider a likelihood built on a Gaussian field model with covariance function (15), it would be rather unlikely that a modeler would end up with the same likelihood if he started from a parameterization of type (2), since this would require considering a unnatural function $r(\cdot)$.

The spatial covariance function $r(\cdot)$ is forced to become negative probably because the Gaussian field model with covariance function given by (15) is not a natural candidate for the parameterization (2). The form of the kernel $h(\cdot)$ induces large positive correlations at large distances, which need to be compensated for by the error term $\varepsilon(t,\mathbf{x})$.

We make a further remark on this example: Limits of covariance functions are also valid covariance functions [Matérn, 1986]. Letting $\theta \to 1$, we obtain the covariance function

$$c(s;\mathbf{u}) = \frac{1}{|s|+c_0} e^{-||\mathbf{u}||^2/(|s|+c_0)}.$$

which is the one considered by Cressie and Huang [1999]. Considering the representation (2), we see from (16) and (17) that $H(\mathbf{0}) \to 1$ while $R(\mathbf{0}) \to 0$. Note that $R(\mathbf{0})$ can be interpreted as the variance of the "spatial average" of $\varepsilon(t,\mathbf{x})$ (properly scaled) meaning that at each time point a noise process with *average value equal to zero* is added. In most cases, such a process will not be very plausible. However, there might be situations where the average value *is* constant (energy equilibrium or mass balance for example), but where the values at specific sites change randomly. Note that in such cases, negative correlations at some sites are necessary.

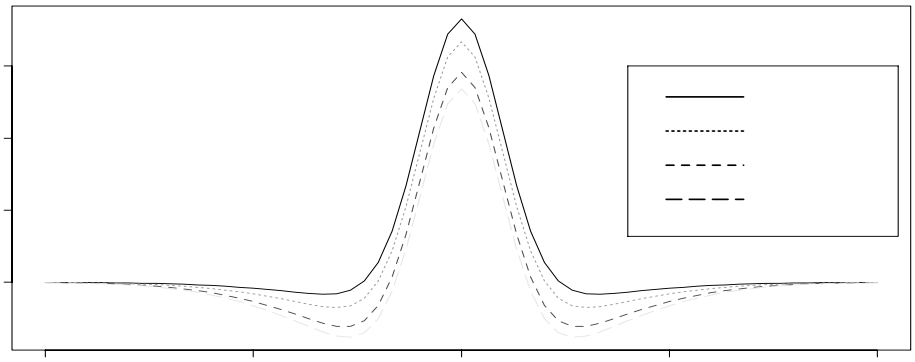

Figure 1: Plot of $r(\mathbf{u})$ for $\theta = 0.5, 0.7, 0.9$ and 0.99. The upper curve is for the lowest value of θ and so on. In all cases $c_0 = 1$.

□

Example 4 (Matern kernel and Matern spatial covariance function)
We now start with a plausible model (2) given by a Matern kernel and spatial covariance functions

$$h(\mathbf{u}) = \frac{\theta}{2^{\nu_1-1}\Gamma(\nu_1)} \left(2\alpha_1^{-1}\sqrt{\nu_1}||\mathbf{u}||\right)^{\nu_1} K_{\nu_1}\left(2\alpha_1^{-1}\sqrt{\nu_1}||\mathbf{u}||\right);$$
$$r(\mathbf{u}) = \frac{1}{2^{\nu_2-1}\Gamma(\nu_2)} \left(2\alpha_2^{-1}\sqrt{\nu_2}||\mathbf{u}||\right)^{\nu_2} K_{\nu_2}\left(2\alpha_2^{-1}\sqrt{\nu_2}||\mathbf{u}||\right).$$

The α_i's are (positive) spatial scaling parameters while the ν_i's are shape parameters, both greater than 0.5. These Matern functions are very flexible and include as special cases the exponential ($\alpha_i = 0.5$) and the Gaussian function ($\nu_i = \infty$). We obtain

$$H(\boldsymbol{\omega}) = \theta\pi\alpha_1^2(1 + \frac{1}{4}\alpha_1^2\nu_1^{-1}\|\boldsymbol{\omega}\|^2)^{-\nu_1-1}$$
$$R(\boldsymbol{\omega}) = \pi\alpha_2^2(1 + \frac{1}{4}\alpha_2^2\nu_2^{-1}\|\boldsymbol{\omega}\|^2)^{-\nu_2-1}$$

showing that $|H(\boldsymbol{\omega})| < 1$ when $\theta < \pi^{-1}\alpha_1^{-2}$. Since $h(\mathbf{x})$ in this case is real, symmetric and unimodal, the integrability of $\Gamma(0;\boldsymbol{\omega})$ is directly obtained (see remark 2 after Theorem 2).

From equations (7) and (8) we have that

$$\Gamma(s;\boldsymbol{\omega}) = \frac{\pi\alpha_2^2(1 + \frac{1}{4}\alpha_2^2\nu_2^{-1}\|\boldsymbol{\omega}\|^2)^{-\nu_2-1}}{1 - (\theta\pi\alpha_1^2)^2(1 + \frac{1}{4}\alpha_1^2\nu_1^{-1}\|\boldsymbol{\omega}\|^2)^{-2(\nu_1+1)}}(\theta\pi\alpha_1^2)^{|s|}(1 + \frac{1}{4}\alpha_1^2\nu_1^{-1}\|\boldsymbol{\omega}\|^2)^{-(\nu_1+1)|s|}. \tag{18}$$

The Fourier transform $c(s;\mathbf{x})$ of $\Gamma(s;\boldsymbol{\omega})$ is a legal covariance function. However, no analytical expression is available, except when ν_1 and ν_2 are integers [Vecchia, 1985]. Even in the integer cases, complex expressions are involved. This shows that in this example it is extremely unlikely that a modeler designing parametric space-time covariance functions $c(s;\mathbf{x})$ for Gaussian fields would consider such a model and the corresponding likelihood. This shows that modelling within framework (2) can be useful and sometimes unavoidable.

It is possible to compute $c(s;\mathbf{x})$ numerically, given values of all the parameters. This allows us to investigate the main features of $c(s;\mathbf{x})$.

Figure 2 is a plot of $c(s;d)$, where $d = \text{sign}(\mathbf{u})\|\mathbf{u}\|$, for $\alpha_1 = \alpha_2 = 2$ and $\nu_1 = \nu_2 = 1$. We notice the sharp discontinuity of the derivative with respect to $\mathbf{u}$, and a somewhat slow decay.

In Figure 3 we plot some projections to enhance the properties of the covariance function $c(s;\mathbf{u}) = c(s;\|\mathbf{u}\|)$.

In Figure 3 (a) we see that $c(s = 0;\mathbf{u})$ (plotted with a solid line) has heavier tails in $\mathbf{u}$ compared to $r(\mathbf{u})$(dashed). This is due to the fact that the autoregressive term in (2) contributes positively to the covariance in space. The exponentially shaped temporal correlation function $c(s;\mathbf{u} = \mathbf{0})$, plotted in Figure 3 ($b$), represents the time autoregressive structure.

It is possible to approximate $c(s;\mathbf{u})$ given through (18) by means of a separable space-time covariance function. One approach is the following:

$$\tilde{c}(s;\mathbf{u}) = \frac{c(s;\mathbf{0}) \cdot c(0;\mathbf{u})}{c(0,\mathbf{0})}.$$

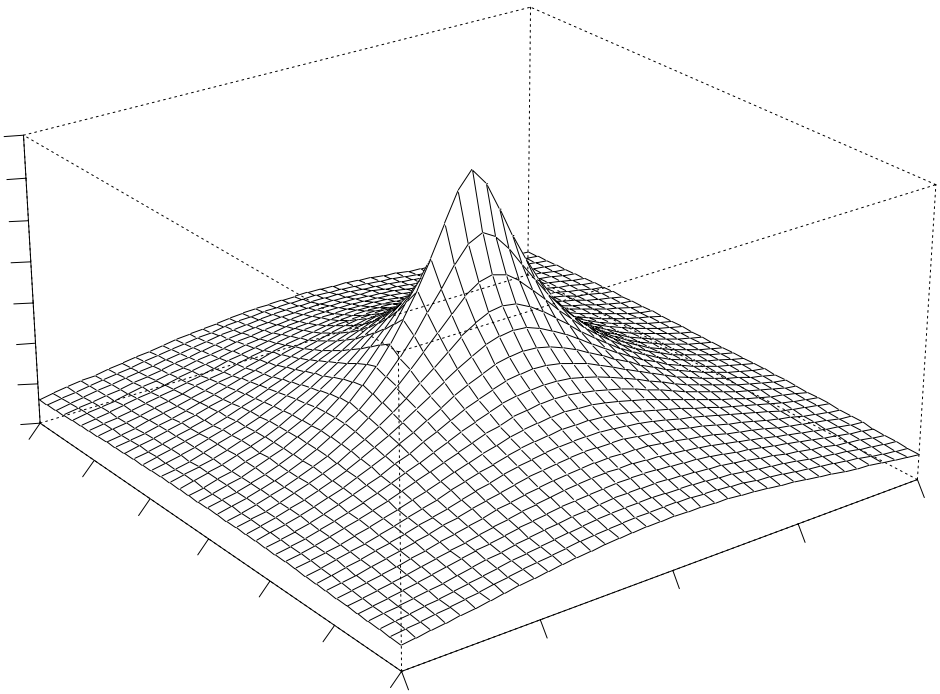

Figure 2: Plot of covariance function $c(s; ||\mathbf{u}||)$ for example 4 with $\nu_1 = \nu_2 = 1$, $\alpha_1 = \alpha_2 = 2$ and $\theta = 0.07$.

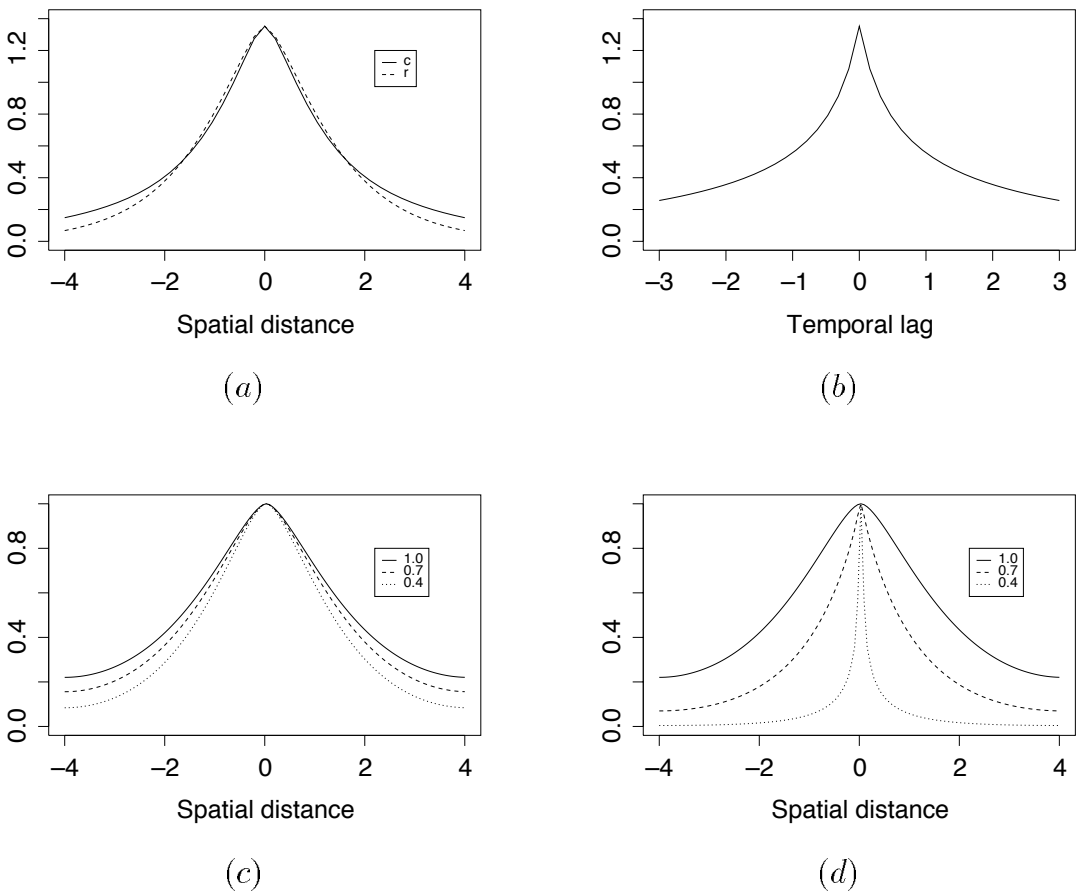

Figure 3: Example 4 with $\nu_1 = \nu_2 = 1$, $\alpha_1 = \alpha_2 = 2$ and $\theta = 0.9/(\pi\alpha_1^2)$. (a): Plot of $c(0; \|\mathbf{u}\|)$ (solid line) and $r(\|\mathbf{u}\|)$ (dashed line), (b): Plot of $c(s; \mathbf{0})$. (c): Plot of $r_\Delta(\|\mathbf{u}\|)$ for $\Delta = 1, 0.7, 0.4$ (scaled such that maximum value is equal to one). (d): Plot of $h_\Delta(\|\mathbf{u}\|)$ for $\Delta = 1, 0.7, 0.4$ (scaled such that maximum value is equal to one).

The covariance function $\check{c}(s;\mathbf{u})$ is now separable with $\check{c}(s;\mathbf{0}) = c(s;\mathbf{0})$ and $\check{c}(0;\mathbf{u}) = c(0;\mathbf{u})$. In Figure 4 the ratio $c(s;||\mathbf{u}||)/\check{c}(s;||\mathbf{u}||)$ is plotted. As expected, due to the structure of model(2), the correlations along the space-time "diagonals" are considerably larger than those we would have obtained from a separable covariance function. It remains to be seen whether the separable approximation would still work well enough in a practical setting despite its significantly different theoretical properties.

In Figure 3 (c) the correlation function $\rho_\Delta(\mathbf{u}) = r_\Delta(\mathbf{u})/r_\Delta(\mathbf{0})$ is plotted for the different time-discretizations $\Delta = 1.0, 0.7, 0.4$. Although slowly changing, the spatial correlations become smaller for decreasing Δ. Also the variance decreases (not shown).

In Figure 3 (d) the convolution function h_Δ is plotted (scaled such that the maximum value is equal to 1). This function is changing much faster and becomes closer to a Dirac's delta function for decreasing Δ. Both the variance and the spatial correlation of the noise term decrease when Δ gets smaller, although the changes are slower than those observed for the convolution function.

For Δ small, the representation (2) is no longer valid. This can be seen by noting that

$$\begin{aligned} H_\Delta(\boldsymbol{\omega}) =& (\theta\pi\alpha_1^2)^\Delta (1 + \frac{1}{4}\alpha_1^2\nu_1^{-1}||\boldsymbol{\omega}||^2)^{-\Delta(\nu_1+1)} \\ =& (\theta\pi\alpha_1^2)^\Delta (1 + \frac{1}{4}\alpha_1^2\nu_1^{-1}||\boldsymbol{\omega}||^2)^{-\nu_\Delta-1} \end{aligned}$$

where $\nu_\Delta = \Delta(\nu_1 + 1) - 1$. For $\nu_\Delta \leq -0.5$ (corresponding to $\Delta \leq 1/2(1+\nu_1)$), the function is not integrable, and so the Fourier transform does not exist. Note that for a Gaussian convolution function ($\nu_1 = \infty$), the representation (2) is valid for all Δ, which is in agreement with the results of Brown et al. [2000]. □

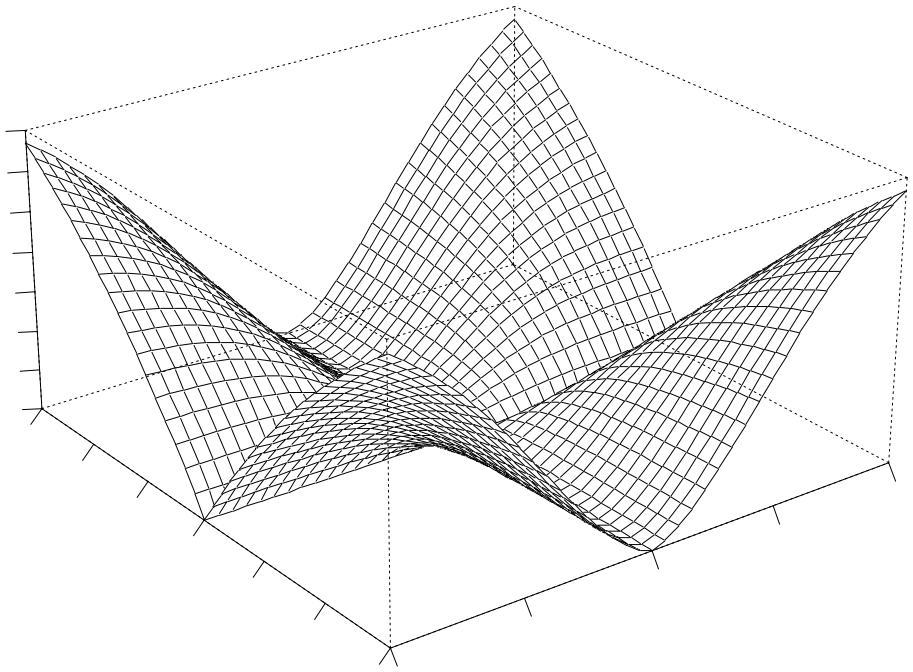

Figure 4: Plot of $c(s;\|\mathbf{u}\|)/\tilde{c}(s;\|\mathbf{u}\|)$ for example 4 with $\nu_1 = \nu_2 = 1$ and $\alpha_1 = \alpha_2 = 2$.

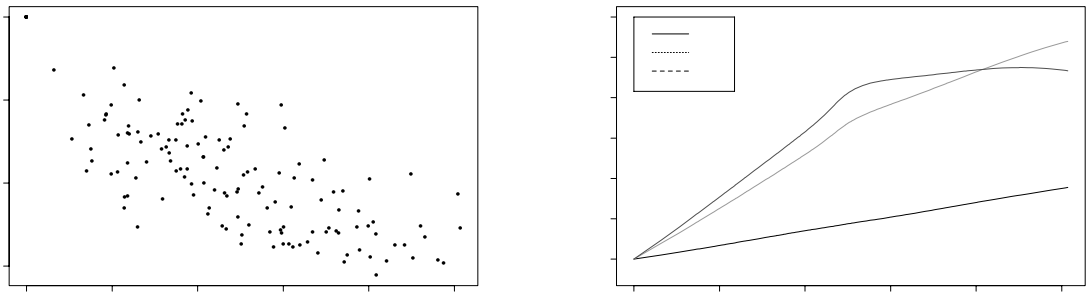

Figure 5: Left panel: Empirical estimates of correlation against distance. Right panel: Non-parametric estimates of $c(s; ||\mathbf{u}||)/c(0; ||\mathbf{u}||)$ for $s = 1, 2, 3$.

6 Experimental results and computational considerations

In the previous sections, we have seen that time autoregressive spatial models of type (2) are Gaussian processes with a covariance structure that can be found at least numerically, and that some models defined by their covariance structure correspond to time autoregressive spatial models. If we have a model that can be described in both ways, we have by re-parameterization two main options for performing numerical computation. In this section we discuss the advantages and disadvantages of working with the time autoregressive spatial structure or the covariance directly.

In order to illustrate the two modelling approaches, we will consider a dataset of daily UK temperatures from $n = 17$ measurement stations (chosen by ???) in the time-period 1959 to 1995. Although an important part of an analysis of such data is specification of a trend model, a non-parametric estimate of the trend was removed from the data in order to concentrate on the covariance structure. Figure 5 (left panel) shows empirical estimates of correlation against distance showing a clear spatial structure. To make a visual check for non-separability, non-parametric estimates (using the S-PLUS function lowess) of $c(s; ||\mathbf{u}||)/c(0; ||\mathbf{u}||)$ are plotted in the right panel of Figure 5 (normalized to have value equal to one for $||\mathbf{u}|| = 0$). With a separable covariance structure, these lines should be horizontal, while departure from constancy indicate separability. Increase in $c(s; ||\mathbf{u}||)/c(0; ||\mathbf{u}||)$ for small values of $||\mathbf{u}||$, which clearly is present for these data, indicate spatial convolution over time.

Consider calculation of likelihoods of such data. For simplicity we will assume the data (after removing the temporal trend) has expectation zero. One approach is to calculate the full covariance matrix for the given data and use the definition of multivariate normal

densities. Such an approach is easy to implement and will work well for small datasets. With large datasets, such an approach will be impractical (in general the computational complexity, mainly involving calculation of the cholesky decomposition of the covariance matrix, is of order N^3 where N is the number of observations). Approximations can be applied (and have been used in the literature) utilizing that for large distances in space or time correlations will be negligible. Such approaches will however be somewhat ad hoc and also require more complex implementation.

An alternative approach is based on the spectral representation of the time autoregressive spatial model given in equations (3) and (4)). Together, these two equations define a dynamic state space model [West and Harrison, 1997] with an infinite dimensional state vector. Discretizing $dZ(t, \boldsymbol{\omega})$ to a finite grid in the frequency space makes Kalman filtering techniques possible to apply (the full details of this computational procedure will be reported elsewhere).

The two computational approaches are compared on the given dataset. An extension of the covariance structure of example 3, incorporating a variance term and scalings both in time and space, i.e.

$$c_e(s; \mathbf{u}) = \begin{cases} \sigma_0^2 + \sigma^2 c(0; \mathbf{0}) & \text{if } s = \mathbf{u} = 0 \\ \sigma^2 c(s; a\mathbf{u}) & \text{otherwise} \end{cases}$$

with σ_0^2, σ^2, a all positive, was used. Likelihoods were calculated using parameters obtained by maximum likelihood based on data from the first 100 time points. The maximum likelihood values were $\theta = 0.99, a = 0.46, c_0 = 9.79, \sigma_0^2 = 99.48$ and $\sigma^2 = 432.54$.

In the left panel of figure 6, the likelihood is calculated based on data from the first t days for $t = 1, ..., 100$. The scale on the x-axis is the number of observations on which the likelihood is based (because of some missing data, this is a bit smaller than $t \times n$). The scale on the y-axis is log-likelihood divided by the number of observations. This division is made in order to get likelihoods on approximatively the same scale. The solid curve is log-likelihood based on direct calculation of the complete covariance matrix, while the dashed and dotted lines are based on the approximative state space model combined with Kalman filtering (using a 12×6 and a 8×4 grid, respectively) for approximating $dZ(t, \boldsymbol{\omega})$). We see that the approximations perform very well (it is actually not possible to see difference between the lines). We also see that errors do not accumulate in time, which can be shown theoretically. This is in contrast to an approach where the process $\{z(t, \mathbf{x})\}$ itself is discretized in space.

In the right hand panel, computer time used for each likelihood calculation is shown. The linear increase in Kalman filtering based likelihood calculations is clearly seen, while a more dramatic increase in computer time is seen for the covariance matrix based method.

Figure 7 gives profile likelihoods based on 100 time points with varying a while all other parameters are fixed equal to their maximum likelihood estimates. Solid lines shows the

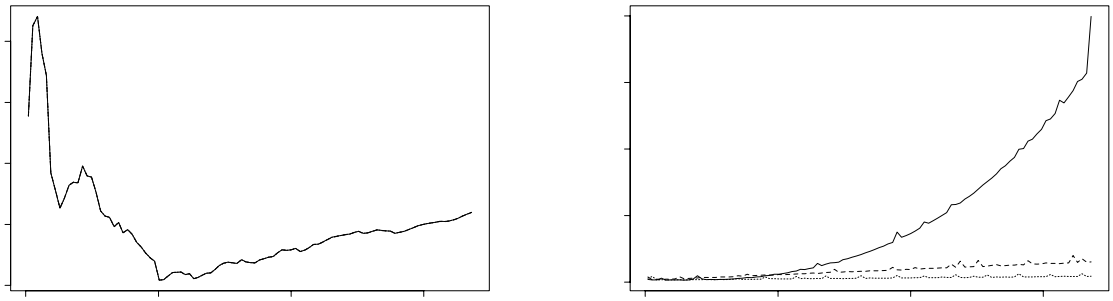

Figure 6: Left panel: Likelihoods as functions of number of observations. Right panel: Computer times as functions of number of observations. Solid line: Calculations based on full covariance matrix. Dashed line: Kalman filter with a 12×6 grid in the spectral domain. Dotted line: Kalman filter with a 8×4 grid.

likelihood function calculated from the full covariance function, while the other lines shows the results based on Kalman filtering using a 12×6 grid (dashed) and a 8×4 grid points (dotted). With 72 grid points, the two approaches are indistinguishable, but also for 32 grid points the approximation is very good.

In the case of no missing data, the $N \times N$ covariance matrix of the complete dataset will have a Toeplitz form. Such structures can be utilized to obtain construct algorithms requiring N^2 (or even $N(logN)^2$) operations [Press et al., 1992, page 95], compared to N^3 operations in the general case. Writing $N = Tn$ where T is the number of data points, while n is the number of spatial locations, the Kalman filter approach require an order of Tn^3 operations which still will be beneficial when T is large. Further, the Kalman filter approach can easily handle missing data, which not is the case for the approach based on using the full covariance matrix.

The properties of the two modelling strategies for performing numerical calculation of likelihoods can be transfered to prediction and conditional simulation because cholesky decomposition of the covariance matrix and running the Kalman filter will be the main computational tasks also in these cases. In particular, conditional simulation based on the spectral state space representation can be performed by the simulation smoother [De Jong and Shephard, 1995].

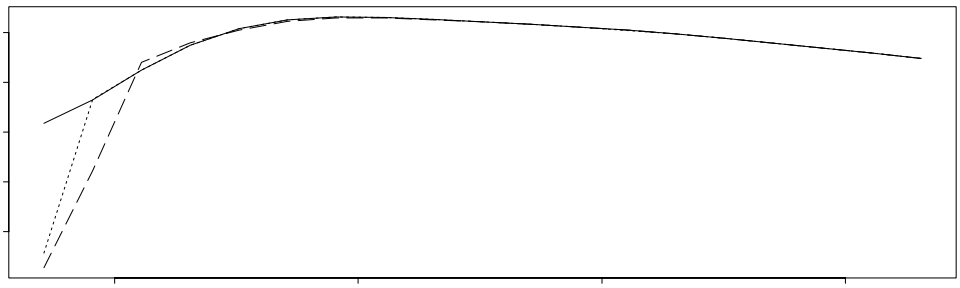

Figure 7: Profile likelihood as a function of a. All other parameters fixed at their maximum likelihood values. Solid line: Calculations based on full covariance matrix. Dashed line: Kalman filter with a 12×6 grid in the spectral domain. Dotted line: Kalman filter with a 8×4 grid.

7 Discussion

In this paper we have discussed two alternative formulations of a spatio-temporal model. We have shown how in some circumstances they can lead to the same process, and how in this case one form can be translated into the other. We have considered the computational and interpretational advantages and disadvantages of both representations.

One aim has been to make it easier to choose a model that could represent a plausible physical process. Often a parametric form for a covariance function is chosen without regard to any interpretation. In some circumstances informative empirical covariance functions can be calculated, and used to inform this choice. In most applications however, either space or time is sparsely sampled, making the empirical covariance function much less informative. In this situation it may be very difficult to justify a particular parametric form, but because of the physical interpretation of the time autoregressive spatial model, specification of the convolution function $h(\cdot)$ and the spatial covariance function $r(\cdot)$ may be easier. Equally the fact that some models, particularly but not exclusively separable ones, cannot be interpreted as time autoregressive spatial models may itself help modelling.

Our discussion has been concentrated on stationary processes. Recently there has been interest in the problem of spatial heterogeneity. Two main approaches have been applied. In Sampson [1992], heterogeneity was modelled through deformation of the spatial domain, using stationary covariance structures on the deformed space. Higdon et al. [1998]

model spatial processes through local convolutions of stationary processes, heterogeneity being incorporated through allowing the convolution kernels to vary smoothly over space (a parametric version of this approach was used in Hirst et al. [2001]). In each case, the spatial heterogeneity is modelled through transformation of stationary spatial processes. The stationary processes discussed in this paper can be used as building blocks for modelling spatial heterogeneity using either of the two representations, with temporal dependence being modelled through time-autoregressive spatial convolution of the underlying stationary processes.

Acknowledgements

We are grateful to the British Atmospheric Data Center which provided us with access to the Met Office Land Surface Observation Stations Data. The authors also wish to thank the anonymous referees for valuable comments on this work. The authors are supported by grant n. 121144/420 of the Norwegian research Council, by the ESF HSSS programme and by the EU-TMR programme ERB-FMRX-CT96-0095.

A Proof of Theorem 1

Proof: Using (2) recursively, we can write

$$z(t,\mathbf{x}) = b_t(\mathbf{x}) + \widetilde{z}(t,\mathbf{x})$$

where

$$b_t(\mathbf{x}) = \int_{\mathbf{u}_1,\ldots,\mathbf{u}_t} z(0,\mathbf{x}+\sum_{s=1}^{t}\mathbf{u}_s)\prod_{s=1}^{t}h(\mathbf{u}_s)d\mathbf{u}_s$$

while $\{\widetilde{z}(t,\mathbf{x})\}$ is a process defined by

$$\widetilde{z}(0,\mathbf{x}) = 0, \qquad \forall \mathbf{x} \tag{19}$$

$$\widetilde{z}(t,\mathbf{x}) = \int_{\mathbf{u})h(\mathbf{u})\widetilde{z}(t-1,\mathbf{x}+\mathbf{u})} d\mathbf{u} + \varepsilon(t,\mathbf{x}), \qquad t>0 \tag{20}$$

We will show that $b_t(\mathbf{x})$ converges towards zero in probability while $\{\widetilde{z}(t,\mathbf{x})\}$ converges towards an unique stationary distribution.

Consider first $b_t(\mathbf{x})$. Now

$$\begin{aligned}\mathsf{var}[b_1(\mathbf{x})] =& \mathsf{var}[\int_{\mathbf{u}} h(\mathbf{u}) z(0, \mathbf{x}+\mathbf{u}) d\mathbf{u}] \\ =& \int_{\mathbf{u}} \int_{\mathbf{v}} h(\mathbf{u}) h(\mathbf{v}) \mathsf{cov}[z(0, \mathbf{x}+\mathbf{u}), z(0, \mathbf{x}+\mathbf{v})] d\mathbf{u} d\mathbf{v} \\ \leq& \int_{\mathbf{u}} \int_{\mathbf{v}} h(\mathbf{u}) h(\mathbf{v}) K d\mathbf{u} d\mathbf{v} \\ =& K H(\mathbf{0})^2.\end{aligned}$$

Repeating this argument recursively, we get that

$$\mathsf{var}[b_t(\mathbf{x})] \leq K H(\mathbf{0})^{2t}.$$

Using now that $|H(\mathbf{0})| < 1$, we see that the variance converges towards zero, showing that the process itself converges towards zero in probability.

Consider next the process $\{\widetilde{z}(t, \mathbf{x})\}$. Because $\{\varepsilon(t, \mathbf{x})\}$ is stationary in space, by the spectral representation theorem (Bouleau and Lepinglé [1994, Theorem B.2.4]) there exist processes $\{E(t, \boldsymbol{\omega})\}$ such that

$$\varepsilon(t, \mathbf{x}) = \int_{\omega} e^{i\boldsymbol{\omega}'\mathbf{x}} dE(t, \boldsymbol{\omega})$$

where $\mathsf{E}[dE(t, \boldsymbol{\omega})] = 0$ and the processes are orthogonal, i.e.

$$\mathsf{cov}[dE(t, \boldsymbol{\omega}), dE(t, \boldsymbol{\omega}')] = 0.$$

for $\boldsymbol{\omega} \neq \boldsymbol{\omega}'$. Now $\widetilde{z}(1, \mathbf{x}) = \varepsilon(1, \mathbf{x})$, giving $d\widetilde{Z}(1, \boldsymbol{\omega}) = dE(1, \boldsymbol{\omega})$. Using (20), we obtain

$$\begin{aligned}\widetilde{z}(2, \mathbf{x}) =& \int_{\mathbf{v}} h(\mathbf{v}) \int_{\omega} e^{i\boldsymbol{\omega}(\mathbf{x}+\mathbf{v})} d\widetilde{Z}(1, \boldsymbol{\omega}) d\mathbf{v} + \int_{\omega} e^{i\boldsymbol{\omega}\mathbf{x}} dE(2, \boldsymbol{\omega}) \\ =& \int_{\omega} \int_{\mathbf{v}} h(\mathbf{v}) e^{i\boldsymbol{\omega}'\mathbf{v}} d\mathbf{v} e^{i\boldsymbol{\omega}'\mathbf{x}} d\widetilde{Z}(1, \boldsymbol{\omega}) + \int_{\omega} e^{i\boldsymbol{\omega}'\mathbf{x}} dE(2, \boldsymbol{\omega}) \\ =& \int_{\omega} H(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}'\mathbf{x}} d\widetilde{Z}(1, \boldsymbol{\omega}) + \int_{\omega} e^{i\boldsymbol{\omega}\mathbf{x}} dE(2, \boldsymbol{\omega}) \\ =& \int_{\omega} e^{i\boldsymbol{\omega}'\mathbf{x}} [H(\boldsymbol{\omega}) d\widetilde{Z}(1, \boldsymbol{\omega}) + dE(2, \boldsymbol{\omega})]\end{aligned}$$

showing that $\{\widetilde{z}(2, \mathbf{x})\}$ is a process with spectral representation

$$\widetilde{z}(2, \mathbf{x}) = \int_{\omega} e^{i\boldsymbol{\omega}'\mathbf{x}} d\widetilde{Z}(2, \boldsymbol{\omega})$$

where $d\widetilde{Z}(2, \boldsymbol{\omega}) = H(\boldsymbol{\omega}) d\widetilde{Z}(1, \boldsymbol{\omega}) + dE(2, \boldsymbol{\omega})$. The zero-mean and orthogonality properties of $\{d\widetilde{Z}(1, \boldsymbol{\omega})\}$ and $\{dE(2, \boldsymbol{\omega})\}$ implies that also $\{d\widetilde{Z}(2, \boldsymbol{\omega})\}$ has zero mean and is orthogonal,

resulting in that $\{z(2,\mathbf{x})\}$ is a process stationary in space (Bouleau and Lepinglé [1994, Proposition B.2.3]). By induction we have

$$\widetilde{z}(t,\mathbf{x}) = \int_{\omega} e^{i\boldsymbol{\omega}'\mathbf{x}} d\widetilde{Z}(t,\boldsymbol{\omega}) \tag{21}$$

where

$$d\widetilde{Z}(t,\boldsymbol{\omega}) = H(\boldsymbol{\omega}) d\widetilde{Z}(t-1,\boldsymbol{\omega}) + dE(t,\boldsymbol{\omega}). \tag{22}$$

with similar orthogonality properties. The process is therefore stationary in space for all $t \geq 0$.

Equations (21) and (22) is an alternative representation of the model (20). Now for each $\boldsymbol{\omega}$, (22) defines an ordinary AR(1) process which has an unique stationary distribution provided $|H(\boldsymbol{\omega})| < 0$. Because $\{\widetilde{z}(t,\mathbf{x})\}$ is completely defined by $\{d\widetilde{Z}(t,\boldsymbol{\omega})\}$, an unique stationary distribution for $\{z(t,\mathbf{x})\}$ must exist. Further, this unique stationary distribution must be stationary in space since $\{\widetilde{z}(t,\mathbf{x})\}$ is stationary in space for all t. □

B Proof of equations (5) and (6)

We have from equation (2)

$$\begin{aligned}
c(0;\mathbf{u}) =& \mathsf{cov}[z(t;\mathbf{x}), z(t;\mathbf{x}+\mathbf{u})] \\
=& \mathsf{cov}[\int_{\mathcal{R}^2} z(t-1;\mathbf{x}+\mathbf{v})d\mathbf{v} + \varepsilon(t;\mathbf{x}), \int_{\mathcal{R}^2} z(t-1;\mathbf{x}+\mathbf{u}+\mathbf{w})d\mathbf{w} + \varepsilon(t;\mathbf{x})] \\
=& \int_{\mathcal{R}^2}\int_{\mathcal{R}^2} h(\mathbf{v})h(\mathbf{w})\mathsf{cov}[z(t-1,\mathbf{x}+\mathbf{v}), z(t-1,\mathbf{x}+\mathbf{u}+\mathbf{w})]d\mathbf{v}d\mathbf{w} + \\
& \mathsf{cov}[\varepsilon(t,\mathbf{x}), \varepsilon(t,\mathbf{x}+\mathbf{u})] \\
=& \int_{\mathcal{R}^2}\int_{\mathcal{R}^2} h(\mathbf{v})h(\mathbf{w})c(0;\mathbf{u}+\mathbf{w}-\mathbf{v}d\mathbf{v}d\mathbf{w} + r(\mathbf{u}),
\end{aligned}$$

where we have used independence between $\{\varepsilon(t,\mathbf{x})\}$ and the past. This shows (5). Similarly,

$$\begin{aligned}
c(s;\mathbf{u}) =& \mathsf{cov}[z(t;\mathbf{x}), z(t-s;\mathbf{x}+\mathbf{u})] \\
=& \int_{\mathcal{R}^2} h(\mathbf{v})\mathsf{cov}[z(t-1,\mathbf{x}+\mathbf{v}), z(t-s,\mathbf{x}+\mathbf{u})]d\mathbf{v} + \mathsf{cov}[\varepsilon(t,\mathbf{x}), z(t-s;\mathbf{x}+\mathbf{u})] \\
=& \int_{\mathcal{R}^2}\int_{\mathcal{R}^2} h(\mathbf{v})c(s-1;\mathbf{u}-\mathbf{v})d\mathbf{v}d\mathbf{w} + 0,
\end{aligned}$$

showing (6).

References

N. Bouleau and D. Lepinglé. *Numerical methods for stochastic processes*. Wiley series in Probability and mathematical statistics. John Wiley & Sons, Inc, 1994.

P. E. Brown, K. F. Kåresen, G. O. Roberts, and S. Tonellato. Blur-generated non-separable space-time models. *Journal of Royal Statistical Society, Series B*, 62(4):847–860, 2000.

R. J. Carroll, R. Chen, E. I. George, T. H. Li, H. J. Newton, H. Schmiedel, and N. Wang. Ozone exposure and population density in Harris county, Texas. *Journal of the American Statistical Association*, 92(438):392–415, 1997. With discussion.

N. Cressie. *Statistics for Spatial Data*. New York: Wiley, 1991.

N. Cressie and H. Huang. Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1340, 1999.

P. De Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82(2):339–350, 1995.

P. J. Diggle, J. A. Tawn, and R. A. Moyeed. Model-based geostatistics. *Journal of Royal Statistical Society, Series C (Applied Statistics)*, 47(2):299–326, 1998.

M. D. Ecker and A. E. Gelfand. Bayesian variogram modeling for an isotropic spatial process. *J. Agricultural, Biological, and Environmental Statistics*, 2:4, 1997.

T. C. Haas. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association*, 90(432):1189–1199, 1995.

D. Higdon, J. Swall, and J. Kern. Non-stationary spatial modeling. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*, pages 761–768. Oxford University Press, Oxford, 1998.

D. Hirst, G. Storvik, and A. R. Syversveen. A hierarchical modelling approach to combining environmental data at different scales. Submitted for publication, 2001.

G. Høst, H. Omre, and P. Switzer. Spatial interpolation errors for monitoring data. *Journal of the American Statistical Association*, 90(431):853–861, 1995.

R. H. Jones and Y. Zhang. Models for continuous stationary space-time processes. In T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G Warren, and R. D. Wolfinger, editors, *Modelling longitudal and spatially correlated data*, volume 122 of *Lecture Notes in Statistics*, pages 289–298. Springer-Verlag, New-York, 1997.

L KnorrHeld and J. Besag. Modelling risk from a disease in time and space. *Statistics in medicine*, 17(18):2045–2060, 1998.

B. Matérn. *Spatial Variation*. Number 36 in Lecture Notes in Statistics. Springer Verlag, 1986.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannert. *Numerical Recipes in C*. Cambridge University Press, 1992.

B. D. Ripley. *Spatial Statistics*. Wiley series in probability and mathematical statistics. John Wiley & Sons, New York, 1981.

P. Sampson, P. D.and Guttorp. Nonparametric-estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119, 1992.

K. Sølna and P. Switzer. Time trend estimation for a geographic region. *Journal of the American Statistical Association*, 91(434):577–589, 1996.

A. V. Vecchia. A general class of models for stationary two-dimensional random processes. *Biometrika*, 72(2):281–291, 1985.

M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1997.

C. K. Wikle and N. Cressie. A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86(4):815–829, 1999.

A. Yaglom. *Correlation Theory of Stationary and Related Random Functions I*. Springer-Verlag, New York, 1987.