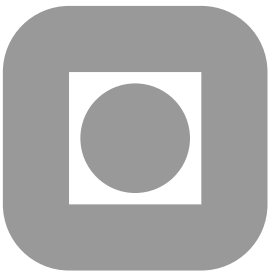
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

A note on the second order random walk model for irregular locations

by

Finn Lindgren & Håvard Rue

PREPRINT
STATISTICS NO. 6/2005



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL http://www.math.ntnu.no/preprint/statistics/2005/S6-2005.pdf
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.

A note on the second order random walk model for irregular locations

Finn Lindgren
Mathematical Statistics
Centre for Mathematical Sciences
Lund University, Sweden

Håvard Rue
Department of Mathematical Sciences
Norwegian University of Science and Technology, Norway

7th September 2005

Abstract

The second order random walk (RW2) model is commonly used for smoothing data and for modelling response functions. It is computationally efficient due to the Markov properties of the joint (intrinsic) Gaussian density. For evenly spaced locations the RW2 model is well established, whereas for irregularly spaced locations there is no well established construction in the literature. By considering the RW2 model as the solution of a stochastic differential equation (SDE), a discretely observed integrated Wiener process, it is possible to derive the density preserving the Markov properties by augmenting the state-space with the velocities. In this note, we derive a computationally more efficient RW2 model for irregular locations using a Galerkin approximation to the solution of the SDE without the need of augmenting the state-space. Numerical comparison with the exact solution demonstrates that the error in the Galerkin approximation is small and negligible in applications.

KEYWORDS: Galerkin approximation, Integrated Wiener process, Intrinsic Gaussian Markov random fields, Numerical methods for sparse matrices, Second order random walk.

ADDRESS FOR CORRESPONDENCE:
Finn Lindgren, Mathematical Statistics, Centre for Mathematical Sciences, Sölvegatan 18, Box 118, SE-221 00 Lund, Sweden

E-MAIL: finn@maths.lth.se

WWW-ADDRESS: http://www.maths.lth.se/matstat/staff/finn/

ACKNOWLEDGEMENTS: Finn Lindgren is supported by the Swedish Research Council (No. 621-2001-2819: Computational and algorithmic aspects of stochastic systems in engineering) and the Swedish Foundation for Strategic Research (Spatial statistics and image analysis for environment and medicine), and also wishes to thank the Department of Mathematical Sciences at NTNU for its hospitality during August 2005.

1 Introduction

The second order random walk (RW2) model for regular locations has the density

$$\pi(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2}\sum_{i=3}^{n}(x_i - 2x_{i-1} + x_{i-2})^2\right) \tag{1}$$

where $\boldsymbol{x} = \left[x_1, \ldots, x_n\right]^T$. The density is invariant under addition of $a+bi$ to x_i for any constants a and b, and is therefore improper with rank deficiency two. The term $x_{i+1} - 2x_i + x_{i-1}$ can be interpreted as an estimate of the second order derivative of a continuous time function $x(t)$ at $t = i$ using values of $x(t)$ at $t = i-1$, i, and $i+1$. Hence, the RW2 model is appropriate for representing "smooth curves" with small squared second derivative. We write "curves" to indicate that often in practice, a straight line is drawn (and is often implicit) in-between x_i and x_{i+1} as an interpolant.

The RW2 model (1) is much used in statistics, in basic tasks such as smoothing data and modelling response functions, where semi-parametric regression, smoothing and penalised likelihood are methods used (Fahrmeir and Tutz, 2001, Fahrmeir and Lang, 2001, Green and Silverman, 1994, Fahrmeir and Knorr-Held, 2000). All these tasks are important ingredients in many statistical models and are used in many areas in statistics. Rue and Held (2005, Ch. 1) gives numerous examples and references.

The popularity of (1) can be explained by two reasons. The RW2 model is quite flexible due to its invariance to addition of a linear trend, and also computationally convenient due to its Markov properties

$$\pi(x_i \mid \boldsymbol{x}_{-i}) = \pi(x_i \mid x_{i-2}, x_{i-1}, x_{i+1}, x_{i+2}) \tag{2}$$

for $2 < i < n-2$, and with trivial changes near the boundary. Here, $\boldsymbol{x}_{-i}$ denote all elements in $\boldsymbol{x}$ except for x_i. The Markov property allows both for fast calculations of the related full conditionals in Markov chain Monte Carlo algorithms, but also, more efficiently, for using direct simulation algorithms based on the Kalman filter, see for example (Kitagawa and Gersch, 1996). The RW2 model is also a Gaussian Markov random field (GMRF) for which more general and very efficient simulation algorithms exist based on numerical methods for sparse matrices, see Rue and Held (2005).

Although (1) is appropriate for regular locations, we often encounter situations where $\{x_i\}$ should represent a smooth "curve" at locations $\{s_i\}$ where the distance between the s_i's are not constant. One such example, is when (1) represents the effect of a covariate in a generalised linear model and the different values of the covariate is not regularly spaced, see Rue and Held (2005, Ch. 4) for many such examples. An alternative approach in such cases, is to use regular locations and to use the value of the interpolant at s_i. This however, often leads to increased dimension of the RW2 model, which implies increased computational effort. A better approach, is to extend the RW2 model (1) to deal with irregular locations.

Let $s_1 < s_2 < \ldots < s_n$ be the set of (fixed) locations and x_i be the corresponding response at s_i, for $i = 1, \ldots, n$. There are two approaches to construct a RW2 model for irregular locations. The first is to use a weighted version of (1) where the weights are selected by some ad-hoc argument, see Fahrmeir and Knorr-Held (2000). The second is to consider the RW2 model as a discretely observed continuous time process $x(t)$, were $x(t)$ is Gaussian and is the solution of

$$\Delta x(t) = \frac{\mathrm{d}W(t)}{\mathrm{d}t}, \tag{3}$$

where $\Delta = \frac{\mathrm{d}^2}{\mathrm{d}t^2}$ and $W(t)$ is the standard Wiener process. Such an approach can be motivated using the connection between smoothing splines and integrated Wiener processes (Wahba, 1978), and the construction of the first order random walk for irregular locations, see for example Rue and Held (2005, Ch. 3).

The solution of (3) does not have any Markov properties, meaning that $\pi(x_i \mid \boldsymbol{x}_{-i})$ does not simplify; the precision (the inverse covariance) matrix $\boldsymbol{Q}$ is dense. Refer to Rue and Held (2005, Ch. 2) for details on this issue. The solution of (3) does however have a Markov property on an augmented space

$$\widetilde{\boldsymbol{x}} = \left[x_1, x_1', \ldots, x_n, x_n'\right]^T \tag{4}$$

where also the derivatives (velocities) at $\{s_i\}$ are included. For this vector with $2n$ elements, Jones (1981) and Wecker and Ansley (1983) showed how to derive the joint density that does possess a Markov property, but the computations takes about $9/2$ the time as for the RW2 model for regular locations; see Rue and Held (2005, Ch. 3.5) for details. This fact, in addition to increased complexity, naturally requires a RW2 model for irregular locations such that (2) holds, but where the precision matrix is such that when n increases, we converge (in some sense) to the solution of (3). In this note, we give such a formulation using a Galerkin approach (see Thomée, 1984, in essence the common *finite element* method) to solve the SDE, and demonstrate that this approximation is more than appropriate to use in applications and that the error is quite small.

The plan for the rest of this note is as follows. First, the Galerkin method is used to construct a Markov random field model approximation to the SDE (3). Second, some theoretical and practical convergence properties are discussed. Third, and finally, a simple time-series representation of the model is derived.

2 Construction

We seek solutions to the differential (or diffusion) equation (SDE) $\Delta x(t) = \mathrm{d}W(t)/\mathrm{d}t$. Let $\langle f,\, g\rangle$ denote the inner-product $\int f(t)g(t)\,\mathrm{d}t$. A key observation is that by the definition of *weak solutions* to the SDE, the identity

$$\langle \phi,\, \Delta x\rangle = \langle \phi,\, \mathrm{d}W/\mathrm{d}t\rangle \tag{5}$$

must hold, for all appropriate *test functions* $\phi(t)$. Now, let Ω be the space of all possible solutions to the SDE, and let $\{\psi_i\}_{i=1,\ldots,n}$ be a set of basis functions for some subspace $\widetilde{\Omega} \subset \Omega$. A *Galerkin approximation* $\widetilde{x}$ to the SDE is constructed as a linear combination of the basis functions,

$$\widetilde{x}(t) = \sum_{i=1}^{n} \psi_i(t) y_i \tag{6}$$

such that the joint distribution of all scalar products $\langle \psi_i,\, \Delta\widetilde{x}\rangle$ equals the joint distribution of all $\langle \psi_i,\, \Delta x\rangle$. The problem is thus reduced to finding the distribution of the weights $\boldsymbol{y} = \big[y_1, \ldots, y_n\big]^T$.

Let $s_1 < s_2 < \cdots < s_n$ be a sequence of discretisation points, and let $d_i = s_{i+1} - s_i$ denote the distances between these points. We construct a Galerkin approximation for $\widetilde{\Omega}$ as the set of continuous, piecewise linear functions with derivative discontinuities at s_i. A set of basis functions ψ_i ,$i = 1, \ldots, n$, is given by

$$\psi_i(t) = \begin{cases} 0, & t < s_{i-1}, \text{ undefined for } i \leq 2, \\ \frac{t - s_{i-1}}{d_{i-1}}, & s_{i-1} \leq t < s_i, \text{ undefined for } i = 1, \\ 1 - \frac{t - s_i}{d_i}, & s_i \leq t < s_{i+1}, \text{ undefined for } i = n, \\ 0, & s_{i+1} \leq t, \text{ undefined for } i \geq n-1. \end{cases} \tag{7}$$

The second order derivatives of $\psi_3, \ldots, \psi_{n-2}$ can be expressed as

$$\Delta\psi_i(t) = \frac{1}{d_{i-1}}\delta_{s_{i-1}}(t) + \left(\frac{-1}{d_i} - \frac{1}{d_{i-1}}\right)\delta_{s_i}(t) + \frac{1}{d_i}\delta_{s_{i+1}}(t), \quad i = 3, \ldots, n-2, \tag{8}$$

where $\delta_s(t)$ is a Dirac's delta function at s. Similar expressions hold for ψ_1, ψ_2, ψ_{n-1}, and ψ_n, with the exception that there are no terms containing $\delta_{s_{-1}}$, δ_{s_1}, δ_{s_n}, or $\delta_{s_{n+1}}$.

Using the identity (5) for each basis function ψ_i, we obtain

$$\big[\langle \psi_i,\, \Delta x\rangle\big]_{i=1,\ldots,n} = \big[\langle \psi_i,\, \mathrm{d}W/\mathrm{d}t\rangle\big]_{i=1,\ldots,n}, \tag{9}$$

where the right-hand side has a Gaussian distribution, with expectation 0 and tridiagonal covariance matrix

$\boldsymbol{B} = \left[\langle \psi_i,\, \psi_j \rangle\right]_{i,j=1,\ldots,n}$ with interior elements

$$B_{i,i-1} = \frac{d_{i-1}}{6}, \quad B_{i,i} = \frac{d_{i-1}+d_i}{3}, \quad B_{i,i+1} = \frac{d_i}{6}, \tag{10}$$

for $2 \leq i \leq n-1$, as well as $B_{1,1} = d_1/3$, $B_{1,2} = d_1/6$, $B_{n,n-1} = d_{n-1}/6$, and $B_{n,n} = d_{n-1}/3$.

For the Galerkin approximation, we obtain

$$\left[\langle \psi_i,\, \Delta\widetilde{x} \rangle\right]_{i=1,\ldots,n} = \left[\textstyle\sum_j \langle \psi_i,\, \Delta\psi_j \rangle y_j\right]_{i=1,\ldots,n} = \boldsymbol{H}\boldsymbol{y}, \tag{11}$$

where $\boldsymbol{H}$ is a tridiagonal matrix with non-zero elements

$$H_{i,i-1} = \frac{1}{d_{i-1}}, \quad H_{i,i} = -\left(\frac{1}{d_{i-1}} + \frac{1}{d_i}\right), \quad H_{i,i+1} = \frac{1}{d_i}, \quad 2 \leq i \leq n-1, \tag{12}$$

for $2 \leq i \leq n-1$. Note that rows 1 and n are zero. The requirement that the collection $\left[\langle \psi_i,\, \Delta\widetilde{x} \rangle\right]_{i=1,\ldots,n}$ should have the same distribution as $\left[\langle \psi_i,\, \Delta x \rangle\right]_{i=1,\ldots,n}$ is fulfilled by the random field $\boldsymbol{y}$ with precision matrix $\boldsymbol{Q} = \boldsymbol{H}^T\boldsymbol{B}^{-1}\boldsymbol{H}$, which is a dense matrix. The following result (see the Appendix for a proof) states that this Galerkin approximation is actually an exact solution to the SDE, possibly apart from at the boundaries. For a numerical validation, see Section 3.2.

Theorem 1. *For circular topology, $\boldsymbol{Q} = \boldsymbol{H}^T\boldsymbol{B}^{-1}\boldsymbol{H}$ is the pseudo-inverse of the covariance $\widetilde{\boldsymbol{\Sigma}}$ for the SDE conditionally on $\sum_{i=1}^n x(s_i) = 0$.*

Because of the dense precision matrix, the Galerkin model is computationally expensive. However, by approximating $\boldsymbol{B}$ with a diagonal matrix, $\boldsymbol{A}$, we obtain a sparse precision matrix, and thus a Markov random field. We construct this diagonal matrix by distributing the off-diagonal values of $\boldsymbol{B}$ to the main diagonal, giving $\boldsymbol{A}$ with $A_{11} = d_1/2$, $A_{nn} = d_{n-1}/2$, and $A_{ii} = \frac{d_{i-1}+d_i}{2}$ elsewhere.

Multiplying the factors of $\boldsymbol{Q}$, the non-zero elements of row i are given by

$$Q_{i,i-2} = \frac{2}{d_{i-2}d_{i-1}(d_{i-2}+d_{i-1})}, \qquad Q_{i,i-1} = \frac{-2}{d_{i-1}^2}\left(\frac{1}{d_{i-2}} + \frac{1}{d_i}\right), \tag{13}$$

$$Q_{i,i} \quad = \frac{2}{d_{i-1}^2(d_{i-2}+d_{i-1})} + \frac{2}{d_{i-1}d_i}\left(\frac{1}{d_{i-1}} + \frac{1}{d_i}\right) + \frac{2}{d_i^2(d_i+d_{i+1})}, \tag{14}$$

with $Q_{i,i+1} \equiv Q_{i+1,i}$, and $Q_{i,i+2} \equiv Q_{i+2,i}$ due to symmetry. At the discretisation boundaries, we use the convention that terms with non-existing components are ignored, or, equivalently, $d_{-1} = d_0 = d_n = d_{n+1} = \infty$. This affects only the upper left corner of $\boldsymbol{Q}$, which becomes

$$\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} \frac{2}{d_1^2(d_1+d_2)} & \frac{-2}{d_1^2 d_2} \\ \frac{-2}{d_1^2 d_2} & \frac{2}{d_1 d_2}\left(\frac{1}{d_1} + \frac{1}{d_2}\right) + \frac{2}{d_2^2(d_2+d_3)} \end{bmatrix}, \tag{15}$$

and correspondingly for the lower right corner of $\boldsymbol{Q}$. It is straightforward but tedious to verify that $\boldsymbol{Q}$ has rank $n-2$, with eigenvectors $\left[1,\ldots,1\right]^T$ and $\left[s_1,\ldots,s_n\right]^T$ corresponding to the double eigenvalue 0, which means that the resulting field is invariant to addition of a linear trend. In the special case where all $d_i = 1$, this $\boldsymbol{Q}$-matrix coincides with the precision matrix for the usual second order random walk,

$$\boldsymbol{Q} = \begin{bmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{bmatrix}, \tag{16}$$

which means that the model in (1) can be interpreted as an (approximate) Galerkin approximation of the SDE in (3).

3 Convergence

3.1 Theoretical discussion

The covariance properties of the approximating GMRF converges to the continuous process as the density of the discretisation time-points increases. This can be shown in the same manner as for the corresponding random field on the sphere, see Lindgren and Rue (2004), via the following observation.

The matrix $\boldsymbol{H}$ can be factorised as $\boldsymbol{H} = \boldsymbol{A}\boldsymbol{W}$, where $\boldsymbol{A}$ is the diagonal matrix defined above, and where each row of $\boldsymbol{W}$ contain the coefficients for calculating local approximations of the Laplacian, with exact results for quadratic polynomials. Then, the precision matrix can be written $\boldsymbol{Q} = \boldsymbol{W}^T\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{W} = \boldsymbol{W}^T\boldsymbol{A}\boldsymbol{W}$, which is the same form used in Lindgren and Rue (2004), and the elements of $\boldsymbol{A}$ can be interpreted as integration weights. This form is closely related to the method of *defining* a solution to the SDE through Stratonovich integration, i.e. as the limit of a trapezoidal integration scheme (see Arnold, 1974).

3.2 Convergence in practice

In order to evaluate the intrinsic stationarity and variance of the constructed intrinsic GMRF, we restrict the field to a circle, looking only at periodic realisations. The variogram $v(u, v)$ for a process $x(t)$ on the circle, $0 \leq t \leq \pi$, $x(0) = x(t)$, is defined as the variance $\mathsf{V}(x(u) - x(v))$. For stationary processes, the variogram depends only on the distance (on the circle) between u and v, $\tau = |u-v| \in [0, l]$, where l is the circumference.

The theoretical variogram for the SDE restricted to the unit circle can also be calculated. The eigenvalues of the Laplacian with respect to the orthonormal basis functions $\cos(kt)/\sqrt{\pi}$, $\sin(kt)/\sqrt{\pi}$, for $k = 0, \ldots, \infty$ is $(\lambda_{\cos,k}, \lambda_{\sin,k}) = (-k^2, -k^2)$, which gives the spectrum $\lambda_k = k^{-4}$, $k = 1, \ldots, \infty$, for the solutions to the SDE. Through the spectral representation of the solutions, the variogram can be calculated as

$$v(\tau) = 2\sum_{k=1}^{\infty} \frac{1 - \cos(k\tau)}{\pi k^4} = \frac{1}{24\pi}\tau^2(2\pi - \tau)^2, \quad 0 \leq \tau \leq 2\pi, \tag{17}$$

where the infinite series can be found in Gradshteyn and Ryzhik (1994).

The variogram for the sparse Galerkin intrinsic GMRF can be obtained for all pairs (s_i, s_j) by computing $V_{ij} = C_{ii} + C_{jj} - 2C_{ij}$, where $\boldsymbol{C}$ is the pseudo-inverse of $\boldsymbol{Q}$. A comparison between the true variogram and the model variogram is shown in Figure 1, for $n = 10$, 20, and 40, and for both regular and irregular spacing. With regular spacing, the variograms are practically indistinguishable already for $n = 20$, whereas the variogram deviation in the case of uniformly distributed discretisation points is larger, and about 3 times as many time-points are needed for the same maximal error as for regular spacing. The bulk of the error appears to be a global scaling factor, and should only have a negligible impact in practical applications.

The variogram for the complete Galerkin approximation turns out to be numerically equivalent to the true solution, as anticipated by Theorem 1; The maximal variogram deviation for $n \leq 40$ is less than 10^{-11} for regular locations and $< 10^{-6}$ for irregular locations. Additionally, this is smaller than for the exact discretisation using an augmented state space mentioned in Section 1, which can typically have a maximal deviation of 10^{-3} for irregular locations, due to numerical issues.

4 Time-series representation

In addition to the random field representation derived in Section 2, where the full conditionals are specified, it is sometimes convenient to have a time-series representation of the random walk model, specifying the distribution of $y_i|y_{i-2}, y_{i-1}$. This can be obtained by moving the future time-points, s_{i+1} and s_{i+2}, toward infinity ($d_i \to \infty$), in the limit removing their influence on y_i. Evaluating the limits of the conditional mean

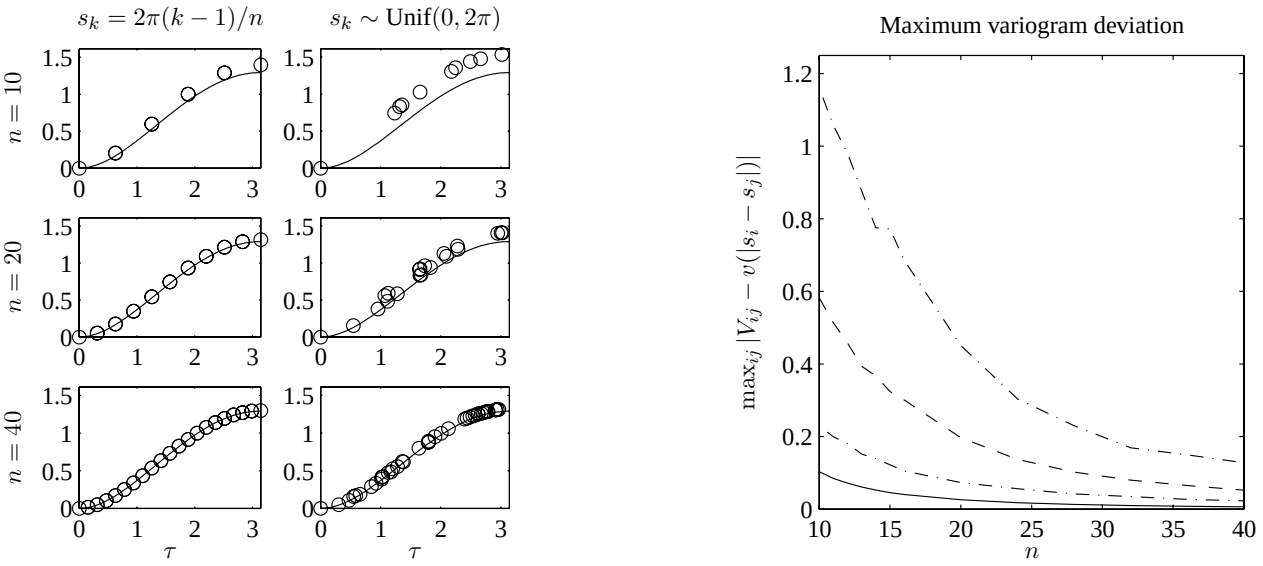


Figure 1: Left: The variogram $V_{1,j}$ for the sparse Galerkin model (circles) compared to the exact SDE variogram (solid line) on the unit circle. Right: The maximum variogram deviation across the entire domain with regular spacing (solid line) for $10 \leq n \leq 40$. The mean maximum deviation for uniformly random spacing (dashed line) was estimated using 1000 replicates for each n. The dash-dotted lines show the 5%- and 95% quantiles.

and variance as $d_i \to \infty$, yields

$$\mathsf{E}(y_i|y_{i-1}, y_{i-2}) = \left(\frac{d_{i-1}}{d_{i-2}} + 1\right) y_{i-1} - \frac{d_{i-1}}{d_{i-2}} y_{i-2}, \qquad \mathsf{V}(y_i|y_{i-1}, y_{i-2}) = \frac{d_{i-1}^2(d_{i-2} + d_{i-1})}{2}. \tag{18}$$

The conditional expectation is the same as given by Fahrmeir and Knorr-Held (2000), but the variance is different.

References

Arnold, L. (1974). *Stochastic Differential Equations; Theory and Applications*. John Wiley & Sons.

Fahrmeir, L. and Knorr-Held, L. (2000). Dynamic and semiparametric models. In Schimek, M. G., editor, *Smoothing and Regression: Approaches, Computation, and Application*, pages 513–544. John Wiley & Sons, Ltd., New-York.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society, Series C*, 50(2):201–220.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, Berlin, 2nd edition.

Gradshteyn, I. and Ryzhik, I. (1994). *Table of Intergrals, Series, and Products*. Academic Press, Fifth edition.

Green, P. J. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Jones, R. H. (1981). Fitting a continous time autoregression to discrete data. In *Applied Time Series Analysis, II (Tulsa, Okla., 1980)*, pages 651–680. Academic Press, New York.

Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*. Lecture Notes in Statistics no. 116. Springer-Verlag, New York.

Lindgren, F. and Rue, H. (2004). Intrinsic Gaussian Markov random fields on triangulated spheres. *Preprints in Mathematical Sciences, Lund University*, 2004(25).

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Thomée, V. (1984). *Galerkin Finite Element Methods for Parabolic Problems*, volume 1054 of *Lecture notes in mathematics*. Springer-Verlag.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B*, 40(3):364–372.

Wecker, W. E. and Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78(381):81–89.

Appendix: Proof of Theorem 1

Let $\boldsymbol{\Sigma}$ be the covariance of the SDE sampled at $s_1, \ldots, s_n$ with circular topology of length l, conditionally on $\int_0^l x(t)\,\mathrm{d}t = 0$. From the spectral representation of the SDE, $\Sigma_{ij} = l^3/720 - \tau_{ij}^2(l-\tau_{ij})^2/(24l)$, where $\tau_{ij} = (s_j - s_i) \mod l$. Using the intrinsic property of the SDE, the covariance conditionally on $\sum_{i=1}^n x(s_i) = 0$ is given by $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{J}\boldsymbol{\Sigma}\boldsymbol{J}$, where $\boldsymbol{J} = \boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/n$. Recall that $\boldsymbol{Q} = \boldsymbol{H}^T\boldsymbol{B}^{-1}\boldsymbol{H}$.

First, we need to show that $\boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{Q} = \boldsymbol{Q}$. For any circular topology, $\boldsymbol{H}$ and $\boldsymbol{B}$ are symmetric circulant matrices, and direct calculation shows that

$$-24l\left[\boldsymbol{\Sigma}\boldsymbol{H}\right]_{ij} = \begin{cases} d_{j-1}(l-d_{j-1})^2 + d_j(l-d_j)^2, & i = j, \\ d_{j-1}((2\tau_{ij} - l - d_j)^2 - 2\tau_{ij}(l-\tau_{ij})) & \\ \quad + d_j((2\tau_{ij} - l + d_j)^2 - 2\tau_{ij}(l-\tau_{ij})), & i \neq j. \end{cases} \tag{19}$$

and $\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H} = \boldsymbol{B} - \boldsymbol{d}\boldsymbol{d}^T/l$, where $\boldsymbol{d} = \boldsymbol{B}\boldsymbol{1}$, so that

$$\begin{aligned} \boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{Q} &= \boldsymbol{Q}\boldsymbol{J}\boldsymbol{\Sigma}\boldsymbol{J}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q} && (20)\\ &= \boldsymbol{H}\boldsymbol{B}^{-1}(\boldsymbol{B} - \boldsymbol{d}\boldsymbol{d}^T/l)\boldsymbol{B}^{-1}\boldsymbol{H} = \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{H} - \boldsymbol{H}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{H}/l && (21)\\ &= \boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{H} = \boldsymbol{Q}, && (22) \end{aligned}$$

where we use that $\boldsymbol{1}$ is an eigenvector of both $\boldsymbol{H}$ and $\boldsymbol{Q}$, with eigenvalue 0.

The second part is more difficult. We need to show that $\widetilde{\boldsymbol{\Sigma}}\boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{\Sigma}}$. Expanding and simplifying the left- and right-hand sides yields $\boldsymbol{J}\boldsymbol{\Sigma}\boldsymbol{H}\boldsymbol{B}^{-1}\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{J} = \boldsymbol{J}\boldsymbol{\Sigma}\boldsymbol{J}$. Define $\widetilde{\boldsymbol{F}} = \left[\tau_{ij}(l-\tau_{ij})/(2l)\right]_{i,j=1,\ldots,n}$ and $\boldsymbol{F} = \boldsymbol{1}\boldsymbol{a} + \widetilde{\boldsymbol{F}}$, for some row vector $\boldsymbol{a}$. We will verify that $\boldsymbol{F} = \boldsymbol{\Sigma}\boldsymbol{H}\boldsymbol{B}^{-1}$ by comparing $\boldsymbol{\Sigma}\boldsymbol{H}$, calculated above, to $\boldsymbol{F}\boldsymbol{B}$. The reader may not want to verify by direct calculation that $\left[\boldsymbol{\Sigma}\boldsymbol{H} - \widetilde{\boldsymbol{F}}\boldsymbol{B}\right]_{ij} = \boldsymbol{1}\boldsymbol{b}$, where $b_j = \left(d_{j-1}(l^2 - d_{j-1}^2) + d_j(l^2 - d_j^2)\right)/(-24l)$. Thus, $\boldsymbol{\Sigma}\boldsymbol{H} = \boldsymbol{F}\boldsymbol{B}$ is fulfilled for $\boldsymbol{a} = \boldsymbol{b}\boldsymbol{B}^{-1}$.

Finally, direct calculation of $\widetilde{\boldsymbol{F}}\boldsymbol{H}$ yields $\widetilde{\boldsymbol{F}}\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{1}\boldsymbol{d}^T/l$, so that

$$\boldsymbol{J}\boldsymbol{F}\boldsymbol{H} = \boldsymbol{J}(\boldsymbol{1}\boldsymbol{a} + \widetilde{\boldsymbol{F}})\boldsymbol{H} = \boldsymbol{J}\widetilde{\boldsymbol{F}}\boldsymbol{H} = \boldsymbol{J}(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{d}^T/l) = \boldsymbol{J} \tag{23}$$

and we are done.