

# On the Second-Order Random Walk Model for Irregular Locations

FINN LINDGREN

*Mathematical Statistics, Centre for Mathematical Sciences, Lund University*

HÅVARD RUE

*Department of Mathematical Sciences, Norwegian University of Science and Technology*

**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. Here, 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.

*Key words:* Galerkin approximation, integrated Wiener process, intrinsic Gaussian Markov random fields, numerical methods for sparse matrices, second-order random walk

## 1. Introduction

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

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

where  $\mathbf{x} = [x_1, \dots, x_n]^T$ . 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 penalized likelihood are methods used (Green & Silverman, 1994; Fahrmeir & Knorr-Held, 2000; Fahrmeir & Lang, 2001; Fahrmeir & Tutz, 2001) (see also the numerous examples and references in Chapter 1 in Rue & Held 2005).

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

$$\pi(x_i | \mathbf{x}_{-i}) = \pi(x_i | x_{i-2}, x_{i-1}, x_{i+1}, x_{i+2}) \quad (2)$$

for  $2 < i < n - 2$ , and with trivial changes near the boundary. Here,  $\mathbf{x}_{-i}$  denote all elements in  $\mathbf{x}$  except for  $x_i$ . The Markov property allows not only 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 & 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 (2001) and Rue & 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 distances between the  $s_i$ s are not constant. One such example is given in section 5, where (1) represents the effect of a covariate in a generalized linear model and the different values of the covariate are not regularly spaced. 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  $s_1 < s_2 < \dots < s_n$  be the set of (fixed) locations and  $x_i$  be the corresponding response at  $s_i$ , for  $i = 1, \dots, n$ . There are two approaches to construct an 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 & Knorr-Held, 2000). The second is to consider the RW2 model as a discretely observed continuous time process  $x(t)$ , where  $x(t)$  is Gaussian and is the solution of

$$\Delta x(t) = \frac{dW(t)}{dt}, \quad (3)$$

where  $\Delta = d^2/dt^2$  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 & Held, 2005, chapter 3).

The solution of (3) does not have any Markov properties, meaning that  $\pi(x_i | \mathbf{x}_{-i})$  does not simplify; the precision (the inverse covariance) matrix  $\mathbf{Q}$  is dense (Rue & Held, 2005, chapter 2). The solution of (3) does however have a Markov property on an augmented space

$$\tilde{\mathbf{x}} = [x_1, x'_1, \dots, x_n, x'_n]^T$$

where also the derivatives (velocities) at  $\{s_i\}$  are included. For this vector with  $2n$  elements, Jones (1981) and Wecker & 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 & Held, 2005, chapter 3.5 for details). To avoid the increased complexity, it is natural to require an RW2 model for irregular locations such that (2) holds, but where the precision matrix is such that when  $n$  increases, the processes converge (in some sense) to the solution of (3). Here, we give such a formulation using a Galerkin approach (see Thomée, 1984, in essence the common *finite element* method) to solve the stochastic differential equation (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 paper 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, with a derivation of a simple time-series representation of the model. This representation differs from previous models in how the variances depend on the irregular discretization. For example, the models from Fahrmeir & Lang (2001) used by Chib & Jeliazkov (2006) have variances more suited to first-order

random walks, which leads to inconsistencies for irregular locations. For related reasons, the model derived for the second-order random walk variance in Rue & Held (2005, section 3.4.1) is also inconsistent for irregular locations, as illustrated in sections 3 and 4.2. In contrast, the model proposed here is consistent with a continuous time RW2 process. Finally, we present an example application.

**2. Construction**

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

$$\langle \phi, \Delta x \rangle = \langle \phi, dW/dt \rangle \tag{4}$$

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

$$\tilde{x}(t) = \sum_{i=1}^n \psi_i(t) w_i$$

such that the joint distribution of all scalar products  $\langle \psi_i, \Delta \tilde{x} \rangle$  equals the joint distribution of all  $\langle \psi_i, \Delta x \rangle$ . The problem is thus reduced to finding the distribution of the weights  $w = [w_1, \dots, w_n]^T$ .

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

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

These basis functions are continuous, and have piecewise constant derivatives. By using generalized functions, the second-order derivatives of  $\psi_3, \dots, \psi_{n-2}$  are nonetheless well defined, and can be expressed as

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

where  $\delta_s(t)$  is a Dirac's delta function at  $s$ . Similar expressions hold for  $\psi_1, \psi_2, \psi_{n-1}$  and  $\psi_n$ , with the exception that there are no terms containing  $\delta_{s_{-1}}, \delta_{s_1}, \delta_{s_n}$  or  $\delta_{s_{n+1}}$ .

Using the identity (4) for each basis function  $\psi_i$ , we obtain

$$[\langle \psi_i, \Delta x \rangle]_{i=1,\dots,n} = [\langle \psi_i, dW/dt \rangle]_{i=1,\dots,n},$$

where the right-hand side has a Gaussian distribution, with expectation 0 and tridiagonal covariance matrix  $B = [\langle \psi_i, \psi_j \rangle]_{i,j=1,\dots,n}$  with interior elements

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

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

$$[\langle \psi_i, \Delta \tilde{x} \rangle]_{i=1, \dots, n} = \left[ \sum_j \langle \psi_i, \Delta \psi_j \rangle w_j \right]_{i=1, \dots, n} = \mathbf{H} \mathbf{w},$$

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

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

for  $2 \leq i \leq n-1$ . Note that rows 1 and  $n$  are zero. The requirement that the collection  $[\langle \psi_i, \Delta \tilde{x} \rangle]_{i=1, \dots, n}$  should have the same distribution as  $[\langle \psi_i, \Delta x \rangle]_{i=1, \dots, n}$  is fulfilled by the random vector  $\mathbf{w}$  with precision matrix  $\mathbf{Q} = \mathbf{H}^T \mathbf{B}^{-1} \mathbf{H}$ , 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, see section 4.2 for a numerical validation. We make use of circular topology, i.e. wrapping a finite interval of length  $l$  onto a circle, which for the SDE corresponds to conditioning on  $x(0) = x(l)$  and  $x'(0) = x'(l)$ . For the discrete versions, the indices are taken modulo  $n$ .

**Theorem 1**

For circular topology,  $\mathbf{Q} = \mathbf{H}^T \mathbf{B}^{-1} \mathbf{H}$  is the pseudo-inverse of the covariance  $\tilde{\Sigma}$  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  $\mathbf{B}$  with a diagonal matrix,  $\mathbf{A}$ , we obtain a sparse precision matrix, and thus a Markov random field. We construct this diagonal matrix by distributing the off-diagonal values of  $\mathbf{B}$  to the main diagonal, giving  $\mathbf{A}$  with  $A_{11} = d_1/2$ ,  $A_{nn} = d_{n-1}/2$  and  $A_{ii} = \frac{d_{i-1} + d_i}{2}$  elsewhere. The matrix  $\mathbf{A}$  is a lower-order approximation of the trapezoidal integration scheme given by  $\mathbf{B}$ . The effect of the approximation can be interpreted as if the discrete process is driven by uncorrelated noise with variance matrix  $\mathbf{A}$  instead of correlated with covariance  $\mathbf{B}$ . However, the integrated noise sequences are similar in the sense that sums of consecutive noise terms will have only slightly higher variance in the approximative model. As seen in the numerical evaluation in the next section, the long-term effect of the approximation is small.

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

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

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

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 discretization 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 and lower right corner of  $\mathbf{Q}$ . For example, the former becomes

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

It is straightforward but tedious to verify that  $\mathbf{Q}$  has rank  $n-2$ , with eigenvectors  $[1, \dots, 1]^T$  and  $[s_1, \dots, s_n]^T$  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  $\mathbf{Q}$ -matrix coincides with the precision matrix for the usual second-order random walk,



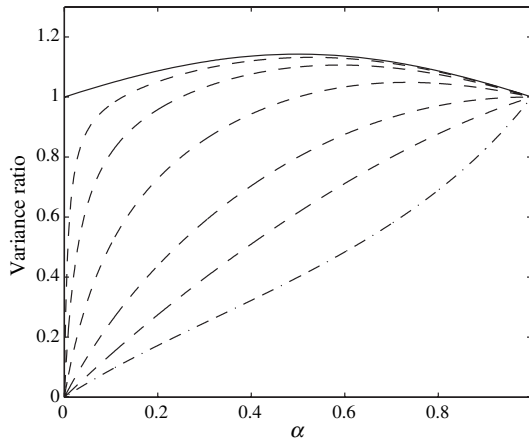


Fig. 1. The ratio between the one-step and two-step variances for the Galerkin model (solid), R&H (dashed), and F&KH (dash dotted) for  $0 < \alpha < 1$ . Only the ratio for R&H depends on  $\Delta$ , and is shown for  $\Delta = 0.5, 1, 2, 4, 8$  increasing towards the ratio for the Galerkin model as  $\Delta \rightarrow \infty$ .

As expected, the F&KH model, which uses a second-order expectation but a first-order random walk variance, has a low consistency, with high two-step variances for almost all  $\alpha$ . The variance ratio for Galerkin model does not depend on  $\Delta$ , but still shows a small deviation from the intuitively desired value 1 for  $\alpha$  away from 0 or 1. The R&H model approaches the Galerkin model for large  $\Delta$ , but the variance ratio depends on the absolute scale of the time-distances, with high inconsistency for small  $\alpha$  even when  $\Delta$  is large. This appears to be an undesired side-effect of the principle used to derive the R&H model, leading to too large variances for irregularly spaced time-points. In section 4.2, these local consistency properties are seen to hold also for the global models.

## 4. Convergence

### 4.1. Theoretical discussion

The covariance properties of the approximating GMRF converges to the continuous process as the density of the discretization time-points increases. This can be shown in the same manner as for the corresponding random field on the sphere, see Lindgren & Rue (2004), via the following observation. The matrix  $\mathbf{H}$  can be factorised as  $\mathbf{H} = \mathbf{A}\mathbf{W}$ , where  $\mathbf{A}$  is the diagonal matrix defined in section 2, and where each row of  $\mathbf{W}$  contains the coefficients for calculating local approximations of the Laplacian, with exact results for quadratic polynomials. Then, the precision matrix can be written  $\mathbf{Q} = \mathbf{W}^T \mathbf{A} \mathbf{A}^{-1} \mathbf{A} \mathbf{W} = \mathbf{W}^T \mathbf{A} \mathbf{W}$ , which is the same form used in Lindgren & Rue (2004), and the elements of  $\mathbf{A}$  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).

### 4.2. 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 realizations. For a process  $x(t)$  on the circle,  $0 \leq t \leq l$ ,  $x(0) = x(l)$ , the variogram  $v(s, t)$  is defined as the variance

$V(x(s) - x(t))$ ,  $0 \leq s, t \leq l$ . For stationary processes, the variogram depends only on the distance (on the circle) between  $s$  and  $t$ ,  $\tau = |s - t| \in [0, l]$ , 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  $k=0, \dots, \infty$  are  $(\lambda_{\cos,k}, \lambda_{\sin,k}) = (-k^2, -k^2)$ , which gives the spectrum  $\lambda_k = k^{-4}$ ,  $k=1, \dots, \infty$ , for the solutions to the SDE. Through the spectral representation of the solutions, the variogram for  $l=2\pi$  can be calculated as

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

where the infinite series can be found in Gradshteyn & Ryzhik (1994). For general  $l$ , the variogram is

$$v(\tau) = \frac{1}{12l} \tau^2 (l - \tau)^2, \quad 0 \leq \tau \leq l.$$

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  $C$  is the pseudo-inverse of  $Q$ . A comparison between the true variogram and the model variogram is shown in Fig. 2 (left), for  $n=10, 20$  and  $40$ , and for both regular and uniform irregular spacing. The corresponding results for the inconsistent Rue & Held (2005) model (marked with  $\times$ ) are included for reference.

With regular spacing, the variograms are practically indistinguishable already for  $n=20$ , whereas the variogram deviation in the case of uniformly distributed discretization points is larger, requiring about three times as many time-points as for regular spacing for the same maximal error. The bulk of the error appears to be a global scaling factor, and should only have a negligible impact in practical applications. The maximal error behaviour between  $n=10$  and  $n=100$  is shown in Fig. 2 (right), in log-log scale, for both regular (solid line)

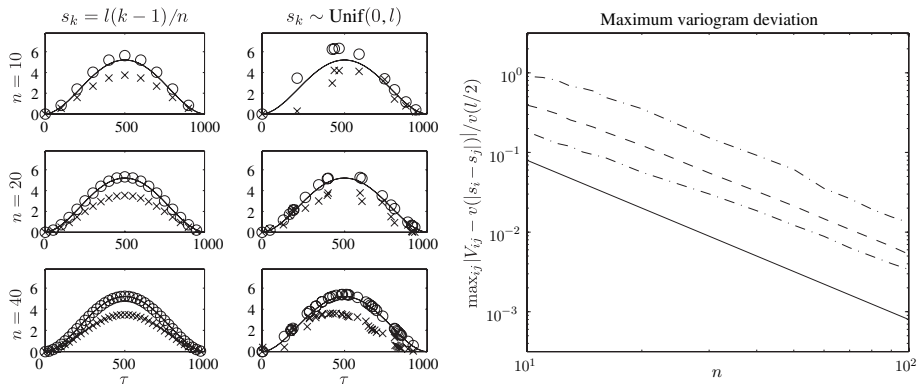


Fig. 2. Left: The variogram  $V_{1,j}$  (rescaled by a factor  $10^{-6}$ ) for the model in Rue & Held (2005, section 3.4.1) (crosses) and for the sparse Galerkin model (circles) compared to the exact SDE variogram (solid line). The length  $l$  is set to 1000, since for small  $l/n$ , the Rue & Held (2005) model has a too different global variance to allow meaningful comparison. Note that even for regular locations, the R&H model has a scaling error of a factor 2/3. Right: The maximum variogram deviation relative to the theoretical value of  $v(l/2)$  for the sparse Galerkin method across the entire domain with regular spacing (solid line) for  $10 \leq n \leq 100$ . The maximum deviation for uniformly random spacing (dashed line) was estimated using 1000 replicates for each  $n$ . The dashed line is the median, and the dash-dotted lines show the 5% and 95% quantiles.

and irregular (dashed lines) spacing. For irregular spacing, the median and the 5% and 95% quantiles are shown, estimated from 1000 simulated spacings for each  $n$ . The relative maximal error in both situations is  $\mathcal{O}(n^{-2})$ . The Rue & Held (2005) model does not converge to the continuous process, as indicated by the local analysis in section 3. Even for regular locations, the overall scale is different, with near equality only for  $d_i \equiv 1$ .

The variogram for the complete Galerkin approximation turns out to be numerically equivalent to the true solution, as anticipated in theorem 1; the maximal relative variogram deviation for  $n=100$  is less than  $10^{-3}$  for regular locations and  $<10^{-2}$  for irregular locations. The exact discretization using an augmented state space, mentioned in section 1, can typically have a maximal relative deviation of  $10^{-4}$  for irregular locations ( $n=100$ ), but numerical issues increase the error in practice for large  $n$ .

## 5. An example

In this example, we will analyse data on the rent of various apartments in Munich in 2003 (see Fahrmeir & Tutz, 2001, example 8.7; or Rue & Held, 2005, section 4.2.2 for more background on this example). The newly constructed RW2 model will enter in representing semi-parametrically the effect of covariates *size* (of the flat) and *year of construction*. The ability to allow for irregular locations in the RW2 model is convenient, as the observed values of the covariates are not regularly spaced. The model is as follows. For each of the 2034 different apartments,  $i$ , the rent per square meter  $y_i$ , is assumed to be Gaussian distributed with mean

$$\mu_i = \mu + x^{\text{Size}}(s_i^{\text{Size}}) + x^{\text{Year}}(s_i^{\text{Year}}) + x^{\text{Spatial}}(s_i^{\text{Spatial}}) + \mathbf{z}_i^T \boldsymbol{\beta} \quad (7)$$

and constant variance. The terms in (7) are as follows;  $\mu$  is the overall mean,  $x^{\text{Size}}(s_i^{\text{Size}})$  is the effect of the *size* ( $s_i^{\text{Size}}$ ) of the apartment,  $x^{\text{Year}}(s_i^{\text{Year}})$  is the effect of the *year of construction*,  $x^{\text{Spatial}}(s_i^{\text{Spatial}})$  is a spatial effect and  $\boldsymbol{\beta}$  is fixed effects of 14 binary covariates  $\mathbf{z}_i$  like central heating, balcony, special bathroom interior, etc.

The effect of covariate *size*, is assumed to follow an RW2 model with irregular locations: the observed values of *size*. Not all values of *size* are observed, hence the need for a model with irregular locations. A similar approach is taken for the effect covariate *year of construction*. The rationale for a such a semi-parametric approach, is to avoid strict assumptions on the parametric form of the covariates, and rather let the ‘data speak for themselves’.

We estimated the parameters using integrated nested Laplace approximations (Rue *et al.*, 2007), and the results are displayed in Fig. 3. The posterior mean is the central line with approximate 95% pointwise credibility interval and observed values of the covariates are indicated by a dot centred at the posterior mean. The effects of the *size* and *year of construction* covariate appears to be non-linear with a quite natural interpretation; very small apartments are relatively more expensive than large ones, and newer apartments are more expensive than older ones. A slight exception is apartments in the oldest buildings, which are often nicely located and renovated.

## 6. Conclusions

We have presented a new GMRF model that approximates a continuous time integrated random walk. The model is consistent even for irregular locations, in contrast to some previously

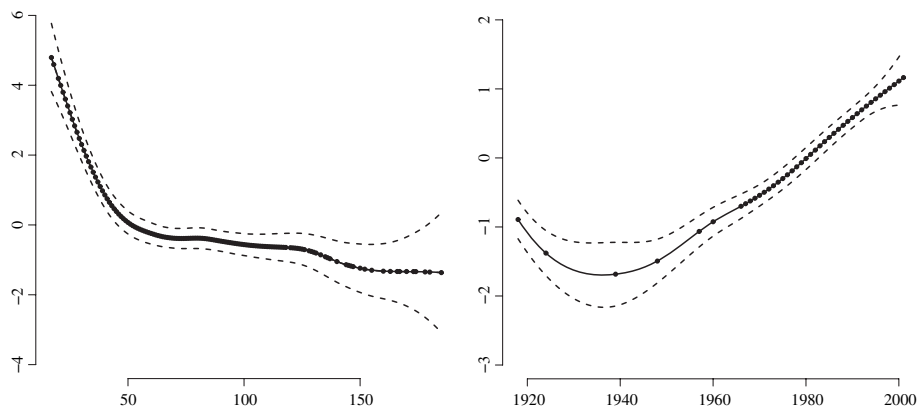


Fig. 3. Left: The estimated effect on the size of the apartment in square metres. Right: The estimated effect of the year of construction. In both plots, the central solid line is a (spline interpolated) posterior mean and approximate 95% point-wise credibility intervals are indicated with dashed line. The observed covariates irregular spacing is indicated by dots on the posterior mean curve.

proposed models. The sparse nature of the precision matrix is appealing for computations in practical applications.

The finite element method provides a straightforward approach to construction of GMRF models approximating simple stochastic partial differential equations (SPDE:s) in time or space. Irregular locations in space can be handled in a similar manner to this work. For one class of SPDE:s, Lindgren & Rue (2007) provide explicit calculations for triangulated 2-manifolds.

### 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. The authors would especially like to thank the two referees and the associate editor for many helpful comments and suggestions.

### References

- Arnold, L. (1974). *Stochastic differential equations; theory and applications*. Wiley, New York.
- Chib, S. & Jeliazkov, I. (2006). Inference in semiparametric dynamic models for binary longitudinal data. *J. Amer. Statist. Assoc.* **101**, 685–700.
- Fahrmeir, L. & Knorr-Held, L. (2000). Dynamic and semiparametric models. In *Smoothing and regression: approaches, computation, and application* (ed. M. G. Schimek), 513–544. Wiley, New York.
- Fahrmeir, L. & Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *J. Roy. Statist. Soc. Ser. C* **50**, 201–220.
- Fahrmeir, L. & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*, 2nd edn. Springer-Verlag, Berlin.
- Gradshteyn, I. & Ryzhik, I. (1994). *Table of integrals, series, and products*, 5th edn. Academic Press, New York.
- Green, P. J. & Silverman, B. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman & Hall, London.
- Jones, R. H. (1981). Fitting a continuous time autoregression to discrete data. In *Applied time series analysis II* (ed. D. F. Findley), 651–680. Academic Press, New York.

Kitagawa, G. & Gersch, W. (1996). *Smoothness priors analysis of time series*. Lecture Notes in Statistics no. 116. Springer-Verlag, New York.

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

Lindgren, F. & Rue, H. (2007). Explicit construction of GMRF approximations to generalised Matérn fields on irregular grids. Technical Report 12, Preprints in Mathematical Sciences, Lund University.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *J. Roy. Statist. Soc. Ser. B* **63**, 325–338.

Rue, H. & Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman & Hall/CRC Press, London.

Rue, H., Martino, S. & Chopin, N. (2007). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. Technical Report No. 1, Department of mathematical sciences, Norwegian University of Science and Technology.

Thomée, V. (1984). *Galerkin finite element methods for parabolic problems*. Lecture Notes in Mathematics No. 1054. Springer-Verlag, Berlin.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40**, 364–372.

Wecker, W. E. & Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *J. Amer. Statist. Assoc.* **78**, 81–89.

Received November 2006, in final form February 2008

Finn Lindgren, Mathematical Statistics, Centre for Mathematical Sciences, Box 118, SE-221 00, Lund, Sweden.  
E-mail: finn@maths.lth.se

**Appendix: Proof of Theorem 1**

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

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

$$-24l[\Sigma\mathbf{H}]_{ij} = \begin{cases} d_{j-1}(l - d_{j-1})^2 + d_j(l - d_j)^2, & i = j, \\ d_{j-1}((2\tau_{ij} - l - d_j)^2 - 2\tau_{ij}(l - \tau_{ij})) \\ + d_j((2\tau_{ij} - l + d_j)^2 - 2\tau_{ij}(l - \tau_{ij})), & i \neq j, \end{cases}$$

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

$$\begin{aligned} \mathbf{Q}\tilde{\Sigma}\mathbf{Q} &= \mathbf{Q}\mathbf{J}\Sigma\mathbf{J}\mathbf{Q} = \mathbf{Q}\Sigma\mathbf{Q} \\ &= \mathbf{H}\mathbf{B}^{-1}(\mathbf{B} - d\mathbf{d}^T/l)\mathbf{B}^{-1}\mathbf{H} = \mathbf{H}\mathbf{B}^{-1}\mathbf{H} - \mathbf{H}\mathbf{1}\mathbf{1}^T\mathbf{H}/l \\ &= \mathbf{H}\mathbf{B}^{-1}\mathbf{H} = \mathbf{Q}, \end{aligned}$$

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

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

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

$$\mathbf{J}\tilde{\mathbf{F}}\mathbf{H} = \mathbf{J}(\tilde{\mathbf{F}} + \mathbf{1}\mathbf{a})\mathbf{H} = \mathbf{J}\tilde{\mathbf{F}}\mathbf{H} = (\mathbf{J}\tilde{\mathbf{F}} + \mathbf{0})\mathbf{H} = \mathbf{J}(\mathbf{I} - \mathbf{1}\mathbf{d}^T/l) = \mathbf{J}.$$