

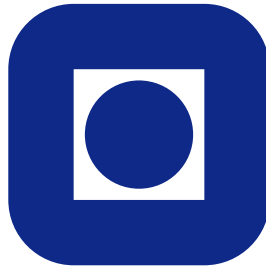
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

**Fitting dynamic models using integrated nested Laplace
approximations – INLA**

by

Ramiro Ruiz-Cárdenas, Elias T. Krainski and Håvard Rue

PREPRINT
STATISTICS NO. 12/2010



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL <http://www.math.ntnu.no/preprint/statistics/2010/S12-2010.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.

Fitting dynamic models using integrated nested Laplace approximations - INLA

Ramiro Ruiz-Cárdenas^{†*}, Elias T. Krainski[§] and Håvard Rue[‡]

First version: September 02, 2010

This version: October 23, 2011

[†] Department of Statistics, Federal University of Minas Gerais – Belo Horizonte, Brazil

[§] Department of Statistics, Federal University of Paraná – Curitiba, Brazil

[‡] Norwegian University for Science and Technology – Trondheim, Norway

Abstract

Inference in state-space models usually relies on recursive forms for filtering and smoothing of the state vectors regarding the temporal structure of the observations, an assumption that is, from our view point, unnecessary if the data set is fixed, that is, completely available before analysis. In this paper we propose a computational framework to perform approximate full Bayesian inference in linear and generalized dynamic linear models based on the Integrated Nested Laplace Approximation (INLA) approach. The proposed framework directly approximates the posterior marginals of interest disregarding the assumption of recursive updating/estimation of the states and hyperparameters in the case of fixed data sets and, therefore, enable us to do fully Bayesian analysis of complex state-space models more easily and in a short computational time. The proposed framework overcomes some limitations of current tools in the dynamic modeling literature and is vastly illustrated with a series of simulated as well as well known real-life examples from the literature, including realistically complex models with correlated error structures and models with more than one state vector, being mutually dependent on each other.

Keywords: Approximate Bayesian inference, state-space models, Laplace approximation, augmented model, spatio-temporal dynamic models

1 Introduction

State space models, also known as *dynamic models* in the Bayesian literature, are a broad class of parametric models with time varying parameters where both, the parameter variation and the available data information are described in a probabilistic way. They find application in the modeling and forecasting of time series data and regression (for a comprehensive treatment see for example West and Harrison, 1997; Durbin and Koopman, 2001). As their static analogues, they can also be generalized to deal with responses belonging to the exponential family of distributions (e.g., West et al., 1985). Inference in these models usually relies on recursive forms for filtering and smoothing of the state vectors regarding the temporal structure of the observations, an assumption that is, from our view point, unnecessary if the data set is fixed,

*Corresponding author. Tel.: 55 31 34095905; fax: 55 31 34095924; e-mail: ramiro@est.ufmg.br

that is, completely available before analysis. In this paper we propose, and illustrate through a series of examples, a computational framework to perform approximate full Bayesian inference in linear and generalized dynamic linear models based on the Integrated Nested Laplace Approximation (INLA) approach. The proposed framework directly approximates the posterior marginals of interest disregarding the assumption of recursive updating/estimation of the states and hyperparameters in the case of fixed data sets and, therefore, enable us to do fully Bayesian analysis of complex state-space models more easily.

INLA is a recent approach proposed by Rue and Martino (2007) and Rue et al. (2009) to perform fast full Bayesian inference through the accurate approximation of the marginal posterior densities of hyperparameters and latent variables in latent Gaussian models. This class of statistical models embraces a wide range of models commonly used in applications, including generalized linear models, generalized additive models, smoothing spline models, semi-parametric regression, spatial and spatio-temporal models, log-Gaussian Cox processes, geostatistical and geoaddivitive models, besides state-space models. Most of these latent models have been successfully fitted using the INLA library, a user friendly interface for using INLA with the R programming language (R Development Core Team, 2010).

In the case of Gaussian observations, the INLA method is actually exact up to integration error. Extensive comparison between INLA and Markov chain Monte Carlo approaches (its simulation-based counterpart) have been performed in Rue et al. (2009) and more recently in diverse contexts, such as posterior predictive checks (Held et al., 2010), spatio-temporal disease mapping models (Schrödle et al., 2011), spatial predictive process models (Eidsvik et al., 2010) and analysis of registry data (Riebler et al., 2011,b), showing the advantages of using the approximate method.

Currently the INLA library provides tools to fit univariate dynamic models with a simple random walk evolution form for the states, such as first order and dynamic regression models, and assuming in the specification of the latent model that the errors are independent. However, for more complex cases, such as growth models and spatio-temporal dynamic models, where there can be more than one state vector, being mutually dependent on each other, and where the error terms can be structured as a matrix with correlated values, the INLA approach does not provide direct ways to perform inference. In this paper we show that, even in these complex cases, it is still possible to formulate specific latent models in a state-space form in order to perform approximate Bayesian inference on them using INLA.

A computational framework to inference is proposed to achieve this goal and illustrated with simulated as well as real-life examples of linear and generalized dynamic linear models. In a first approach existing model options in the INLA library are used to directly model first order random walk evolution and seasonal behavior of simpler state-space models. Furthermore, we build on some functionalities present in the INLA library to manipulate several likelihoods and correlated error structures to develop a framework that enable the formulation and fitting of dynamic models in a more general setting using INLA. This generic approach becomes useful in the case of more complex models, as those mentioned above, and consists in merging the actual observations from the observational equation with “pseudo” observations coming from the evolution (system) equations of the dynamic model in a unique structure and fit this augmented latent model in INLA considering different likelihoods for the observations and states. The combination of the two approaches is also possible. We show how this inference framework enables the fitting of several kinds of dynamic models, including realistically complex spatio-temporal models, in a short computational time and in a user friendly way.

The rest of the paper is organised as follows. In Section 2 we briefly introduce dynamic models and the main computational approaches in the literature to perform inference on this

class of models. Section 3 describes the basics of the INLA computational approach. The proposed framework to fit state-space models using INLA is illustrated in Section 4 through a series of simulated examples, ranging from simple univariate models to realistically complex spatio-temporal dynamic models. In Section 5 some well known worked examples from the literature are considered and their fitting using the INLA library is compared with current computational tools. Concluding remarks and future work are stated in Section 6.

2 Dynamic models

According to Migon et al. (2005), dynamic models can be seen as a generalization of regression models, allowing changes in parameter values throughout time by the introduction of an equation governing the temporal evolution of regression coefficients. In the linear Gaussian case they consist of the couple of equations

$$y_t = F_t'x_t + \nu_t, \quad \nu_t \sim N(0, V_t) \quad (1)$$

$$x_t = G_t x_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t), \quad (2)$$

where y_t is a time sequence of scalar observations and x_t is a sequence of state (latent) parameters describing locally the system. It is assumed that y_t is conditionally independent given x_t . F_t is a vector of explanatory variables, while G_t represents a matrix describing the states evolution. Both, F_t and G_t are usually defined by the modeler according to model design principles (see West and Harrison, 1997). The disturbances ν_t and ω_t are assumed to be both, serially independent and also independent of each other although it is not a necessary condition. Therefore, the model is completely specified by the quadruple $\{F_t; G_t; V_t; W_t\}$. When these quantities are known, inference on the states x_t can be performed analytically through an iterative procedure using the Kalman filter algorithm (for details see for example West and Harrison, 1997).

On the other hand, if either of the variances V_t and W_t is unknown, inference in dynamic linear models is not available analytically. In order to circumvent this problem, several proposals to perform approximate inference in DLMS have appeared in the literature, including approaches based on the extended Kalman filter (Anderson and Moore, 1979), Gaussian quadratures (Pole and West, 1990), data augmentation (Frühwirth-Schnatter, 1994), Laplace approximations (Ehlers and Gamerman, 1996) and assumed density approaches (Zoeter and Heskes, 2006). In recent years, attention has been mostly concentrated in simulation-based approaches, such as sequential Monte Carlo also known as particle filters (e.g., Gordon et al., 1993; Storvik, 2002; Doucet and Tadić, 2003) and Markov chain Monte Carlo (MCMC) methods (e.g., Carter and Kohn, 1994, 1996; Gamerman, 1998; Reis et al., 2006) or still in a combination of these two methods (e.g., Andrieu et al., 2010; Whiteley et al., 2010). MCMC is currently the most common approach to inference in dynamic models due to its generality and capability to obtain samples from the posterior distribution of all unknown model parameters in an efficient way. However, MCMC implementation is more involved and it suffers from a series of well known problems that have hindered its wider utilization in applied settings. For example, convergence can be quite difficult to diagnose and the computational cost may become prohibitively high for complex models, as is the case of spatio-temporal dynamic models. The Monte Carlo errors are also intrinsically large and strong correlation among parameters is common making necessary longer runs of the MCMC algorithm.

A sort of computational tools to fit state-space models using some of the inference methods mentioned above have also appeared in the literature to help end users to benefit from methodological developments. The first of them was the Bats software (West et al., 1988; Pole et

al., 1994), a package for time series analysis and forecasting using Bayesian dynamic modeling, developed in the late 80's by the "Bayesian Forecasting Group" of Warwick University. It deals with univariate time series and dynamic regression models. It performs sequential estimation and uses a discount factor approach to model the unknown variances.

The `SsfPack` library (Koopman et al., 1999), a module for the programming language Ox, provides functions for likelihood evaluation and signal extraction of linear Gaussian state-space models, with support for estimating some non-Gaussian and nonlinear models using importance sampling and MCMC methods.

More recently some R packages and functions to fit linear and generalized linear dynamic models have been developed. The function `StructTS` written by Bryan Ripley (see Ripley, 2002) fits linear Gaussian state-space models (also called structural models) for univariate time series by maximum likelihood, by decomposing the series in trend and/or seasonal components. The `dlm` package (Petris, 2010), performs maximum likelihood estimation, Kalman filtering and smoothing, and Bayesian analysis (through a Gibbs sampler) of Gaussian linear dynamic models. The algorithms used for Kalman filtering, likelihood evaluation, and sampling from the state vectors are based on the singular value decomposition of the relevant variance matrices.

The `sspir` package (Dethlefsen and Lundbye-Christensen, 2006) includes functions for Kalman filtering and smoothing of linear and generalized dynamic linear models. Estimation of variance matrices can be performed using the EM algorithm in the Gaussian case, but it requires that the variance matrices to be estimated are constant. Non-Gaussian state space models are approximated to a Gaussian state-space model through an iterated extended Kalman filtering approach. The `KFAS` package (Helske, 2010), also provides functions for Kalman filtering and smoothing of univariate exponential family state space models. Yet another implementation is given in the `FKF` package (Luethi et al., 2010), which implements a fast Kalman filter for fitting high-dimensional linear state-space models to large datasets.

These computational tools have contributed to a wider use of dynamic models in applied contexts. However, support remains incomplete in some particular modeling aspects. For example, the above approaches, in general, allow missing values just in the observation vector, but not in the covariates. As we leave the Gaussian univariate case, estimation of hyperparameters and its uncertainty is not straightforward to obtain. Estimation of more complex dynamic models as is the case of the spatio-temporal ones is also not possible with these tools. All these gaps can however be filled using the INLA approach.

Another concern is related to the fact that parameter estimation with the above software tools is performed, in general, using a "dynamic" (recursive) algorithm, for example, following Kalman's ideas, where filtering and smoothing steps are used in order to estimate the hyperparameters and states, considering the temporal order of the observations. Their inference procedure is based on the (not always needed) assumption that sequential updating/estimation is required. This seems to be also built in stone in reading books on the subject. This thinking is justified in problems where on-line (i.e., real time) estimation and prediction is required every time that a new observation arrives, both from the point of view of storage costs as well as for rapid adaptation to changing signal characteristics. Examples of applied settings where this situation is common include computer vision, economics and financial data analysis, feed-back control systems, mobile communications, radar surveillance systems, etc. In that cases, a recursive filter is a convenient solution and several proposals have appeared in the literature to efficiently solve the problem based, for example, on sequential Monte Carlo algorithms (for a review see Andrieu et al., 2004; Cappé et al., 2007, and references therein). However, when the data set is fixed, in the sense that all observations were already measured, and the interest is

in the estimation of parameters and states using just this information, there is no reason why the procedure for inference should also be “dynamic”. This is the idea pursued in the INLA approach, where the posteriors of interest are directly approximated avoiding look (at least computationally) at the temporal structure of the data. From our view point it seems more natural and makes full Bayesian analysis (that is, assessment of uncertainty for the hyperparameters, predictive marginals, etc.) possible in an easy way, even for complex state-space models.

Yet another advantage of the INLA approach is its suitability to perform model comparison. Marginal likelihoods, for example, which can be used as a basis to compare competing models through the Bayes factor, can be easily computed with INLA. Additionally, the deviance information criterion (Spiegelhalter et al., 2002) and two predictive measures, the conditional predictive ordinate and the probability integral transform, used to validate and compare models and as a tool to detect atypical observations, are available from the INLA output (see Martino and Rue, 2010, for implementation details).

In the next two sections we show, through a series of examples, how to take advantage of the INLA features to obtain an improved inference in linear and generalized dynamic linear models in a simple, yet flexible way.

3 The Integrated Nested Laplace Approximation (INLA) approach

INLA is a computational approach, recently introduced by Rue and Martino (2007) and Rue et al. (2009), to perform fast Bayesian inference in the broad class of latent Gaussian models, that is, models of an outcome variable y_i that assume independence conditional on some underlying (unknown) latent field $\boldsymbol{\xi}$ and a vector of hyperparameters $\boldsymbol{\theta}$. It was proposed as an alternative to the usually time consuming MCMC methods. Unlike MCMC where posterior inference is sample-based, the INLA computational approach directly approximates the posteriors of interest with a closed form expression. Therefore, problems of convergence and mixing, inherent to MCMC runs, are not an issue. The main aim of the INLA approach is to approximate the marginal posteriors for the latent variables as well as for the hyperparameters of the Gaussian latent model, given by

$$\pi(\xi_i | \mathbf{y}) = \int \pi(\xi_i | \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (3)$$

$$\pi(\theta_j | \mathbf{y}) = \int \pi(\boldsymbol{\theta} | \mathbf{y}) d\theta_{-j}. \quad (4)$$

This approximation is based on an efficient combination of (analytical Gaussian) Laplace approximations to the full conditionals $\pi(\boldsymbol{\theta} | \mathbf{y})$ and $\pi(\xi_i | \boldsymbol{\theta}, \mathbf{y})$, $i = 1, \dots, n$, and numerical integration routines to integrate out the hyperparameters $\boldsymbol{\theta}$.

The INLA approach as proposed in Rue et al. (2009) includes three main approximation steps to obtain the marginal posteriors in (3) and (4). The first step consists in approximate the full posterior $\pi(\boldsymbol{\theta} | \mathbf{y})$. To achieve this, firstly and approximation to the full conditional distribution of $\boldsymbol{\xi}$, $\pi(\boldsymbol{\xi} | \mathbf{y}, \boldsymbol{\theta})$, is obtained using a multivariate Gaussian density $\tilde{\pi}_G(\boldsymbol{\xi} | \mathbf{y}, \boldsymbol{\theta})$ (for details see Rue and Held, 2005) and evaluated at its mode. Then the posterior density of $\boldsymbol{\theta}$ is approximated by using the Laplace approximation

$$\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \propto \frac{\pi(\boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\boldsymbol{\xi} | \boldsymbol{\theta}, \mathbf{y})} \Big|_{\boldsymbol{\xi}=\boldsymbol{\xi}^*(\boldsymbol{\theta})},$$

where $\boldsymbol{\xi}^*(\boldsymbol{\theta})$ is the mode of the full conditional of $\boldsymbol{\xi}$ for a given $\boldsymbol{\theta}$. Since no exact closed form is available for $\boldsymbol{\xi}^*(\boldsymbol{\theta})$, an optimization scheme is necessary. Rue et al. (2009) computes this mode using a Newton-Raphson algorithm. The posterior $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ will be used later to integrate out the uncertainty with respect to $\boldsymbol{\theta}$ when approximating the posterior marginal of ξ_i .

The second step computes the Laplace approximation of the full conditionals $\pi(\xi_i | \mathbf{y}, \boldsymbol{\theta})$ for selected values of $\boldsymbol{\theta}$. These values of $\boldsymbol{\theta}$ must be carefully chosen, as explained below, as they will also be used as evaluation points in the numerical integration applied to obtain the posterior marginals of ξ_i in (3). The density $\pi(\xi_i | \boldsymbol{\theta}, \mathbf{y})$ is approximated using the Laplace approximation defined by:

$$\tilde{\pi}_{LA}(\xi_i | \boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\boldsymbol{\xi}_{-i} | \xi_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\boldsymbol{\xi}_{-i} = \boldsymbol{\xi}_{-i}^*(\xi_i, \boldsymbol{\theta})}, \quad (5)$$

where $\boldsymbol{\xi}_{-i}$ denotes the vector $\boldsymbol{\xi}$ with the i th component omitted, $\tilde{\pi}_G(\boldsymbol{\xi}_{-i} | \xi_i, \boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation of $\pi(\boldsymbol{\xi}_{-i} | \xi_i, \boldsymbol{\theta}, \mathbf{y})$, treating ξ_i as fixed (observed) and $\boldsymbol{\xi}_{-i}^*(\xi_i, \boldsymbol{\theta})$ is the mode of $\pi(\boldsymbol{\xi}_{-i} | \xi_i, \boldsymbol{\theta}, \mathbf{y})$.

The approximation of $\pi(\xi_i | \boldsymbol{\theta}, \mathbf{y})$ using (5) can be quite expensive, since $\tilde{\pi}_G(\boldsymbol{\xi}_{-i} | \xi_i, \boldsymbol{\theta}, \mathbf{y})$ must be recomputed for each value of ξ_i and $\boldsymbol{\theta}$. Two alternatives are proposed in Rue et al. (2009) to obtain these full conditionals in a cheaper way. The first one is just the Gaussian approximation $\tilde{\pi}_G(\xi_i | \boldsymbol{\theta}, \mathbf{y})$, which provides reasonable results in short computational time. However, according to Rue and Martino (2007), its accuracy can be affected by errors in the location and/or errors due to the lack of skewness. These weaknesses can be corrected at a moderate extra cost, using a simplified version of the Laplace approximation, defined as the series expansion of $\tilde{\pi}_{LA}(\xi_i | \boldsymbol{\theta}, \mathbf{y})$ around $\xi_i = \mu_i(\boldsymbol{\theta})$, the mean of $\tilde{\pi}_G(\xi_i | \boldsymbol{\theta}, \mathbf{y})$ (for details see Rue et al., 2009).

Finally, in the third step the full posteriors obtained in the previous two approximation steps are combined and the marginal posterior densities of ξ_i and θ_j are obtained by integrating out the irrelevant terms. The approximation for the marginal of the latent variables can be obtained by the expression

$$\pi(\xi_i | \mathbf{y}) = \int \pi(\xi_i | \mathbf{y}, \boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \approx \sum_k \tilde{\pi}(\xi_i | \boldsymbol{\theta}_k, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}_k | \mathbf{y}) \Delta_k, \quad (6)$$

which is evaluated using numerical integration on a set of grid points for $\boldsymbol{\theta}$, with area weights Δ_k for $k = 1, 2, \dots, K$. According to Rue et al. (2009), since these integration points are selected in a regular grid, it is feasible to take all the area weights Δ_k to be equal. A similar numerical integration procedure is used for the evaluation of the marginals $\pi(\theta_j | \mathbf{y})$. Since the dimension of $\boldsymbol{\theta}$ is assumed small (i.e., ≤ 7), these numerical routines are efficient in returning a discretized representation of the marginal posteriors.

A good choice of the set of evaluation points is crucial to the accuracy of the above numerical integration steps. In order to do that, Rue et al. (2009) suggest to compute the negative Hessian matrix S at the mode, $\boldsymbol{\theta}^*$, of $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ and to consider its spectral value decomposition, $S^{-1} = Q\Lambda Q^T$. Then $\boldsymbol{\theta}$ is defined via a standardized variable, z , as

$$\boldsymbol{\theta}(z) = \boldsymbol{\theta}^* + Q\Lambda^{1/2}z \quad \text{or} \quad z = Q^T\Lambda^{-1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

and a collection, Z , of z values is found, such that the corresponding $\boldsymbol{\theta}(z)$ points are located around the mode $\boldsymbol{\theta}^*$. Starting from $z = 0$ ($\boldsymbol{\theta} = \boldsymbol{\theta}^*$), each component entry of z is searched in the positive and negative directions in step sizes of η_z . All z -points satisfying

$$\log \tilde{\pi}(\boldsymbol{\theta}(0) | \mathbf{y}) - \log \tilde{\pi}(\boldsymbol{\theta}(z) | \mathbf{y}) < \eta_\pi$$

are taken to be in Z . The set of evaluation points is finally based on the values in Z . An appropriate tuning of the η_z and η_π values should be performed in order to produce accurate approximations.

An efficient computational implementation of the procedures needed by the INLA approach was built on the open source library GMRFLib (Rue and Follstad, 2002), a C-library for fast and exact simulation of Gaussian Markov random fields. It particularly benefits from sparse matrix algorithms available in the GMRFLib library to quickly model the Gaussian latent field. Its user friendly R interface is available from the web page <http://www.r-inla.org/>, which also includes many examples of fitting for most of the latent models mentioned in the introduction.

The following two sections show, through a series of examples, how the INLA approach can be extended to deal with inference in dynamic models in an easy way using the INLA library.

4 INLA for state-space models

In this section we illustrate, through a series of simulated data sets, the proposed computational framework to formulate and fit the most common types of dynamic models using the INLA approach. Firstly, the steps to perform fast Bayesian inference on these models using the INLA library are described in detail using a simple univariate dynamic linear model. Next, we consider examples of state-space models that can be directly fitted using model options that already exist in the INLA library, that is, models with a simple random walk evolution form and with independently distributed error terms. Finally, the proposed framework is applied on models that could not be fitted using INLA's standard tools, as is the case of models with several state vectors dependent among them as well as models whose error terms are correlated. The results evidence the capability of INLA to fit realistically complex state-space models.

All the examples in this section were fitted using different log-gamma priors for the log-precisions of observations and states specified as follows: an informative prior, with mean equal to the real value and coefficient of variation equal to 0.5, a vague prior also centered on the true simulated value but with coefficient of variation equal to 10 and the default INLA log-gamma prior set as $\text{log-gamma}(1, 0.00005)$. As little sensitivity to prior specification was, in general, observed in all the cases, marginal posterior densities for the hyperparameters with these priors are only detailed for some of the examples.

The analyses of all the examples in this paper were carried out in an Intel Core 2 Duo model laptop at 2.40GHz and 4GB RAM running under a Windows 7 operating system. Computer time is reported for the most representative cases. Main parts of the R code, considered as relevant to a better understanding of the proposed approach are included in this section. The full code to simulate and fit all the examples considered in this paper is available from the R script accompanying this report.

4.1 A toy example

We begin with a very simple simulated example of a first order univariate dynamic linear model in order to gain insight into the specification of dynamic models for use within INLA. The model has the following observational and system equations:

$$y_t = x_t + \nu_t, \quad \nu_t \sim N(0, V), \quad t = 1, \dots, n \quad (7)$$

$$x_t = x_{t-1} + \omega_t, \quad \omega_t \sim N(0, W), \quad t = 2, \dots, n \quad (8)$$

That is, we are assuming that $F_t = G_t = 1$, $V_t = V$ and $W_t = W$, for all t . Therefore, the model has $n + 2$ unknown parameters. Following the notation in section 3, the vector of hyperparameters is given by $\theta = \{V, W\}$, while the latent field corresponds to $\xi = \{x_1, \dots, x_n\}$. Since the evolution of states in this simple model follows a first order random walk process, it could just be fitted with the INLA library using existing model options according to the following code:

```
i <- 1:n          # indices for x_t
formula <- y ~ f(i, model="rw1", constr=F) -1
r <- inla(formula, data = data.frame(i,y))
```

where the option “constr=F” relaxes the sum to zero constraint on the x vector in the RW1 model, allowing the recovering of the posterior mean for the states. However, we will use this simple example to propose and illustrate a generic approach to fit dynamic models with INLA based on an augmented model structure, which enables the fitting of complex state-space models as will be shown later on in this section. The key feature of this generic approach consists in equating to zero the system equation, that is, we re-write (8) as

$$0 = x_t - x_{t-1} - \omega_t, \quad \omega_t \sim N(0, W), \quad t = 2, \dots, n \quad (9)$$

and then we build an augmented model with dimension $n + (n - 1)$ merging these “faked zero observations” from system equation (9) with the actual observations from Eq. (7) in a unique structure, as shown in Diagram 1. The first column of this structure is associated to the actual observations, $y_t = \{y_1, \dots, y_n\}$, which occupy the first n elements. The second column is associated to the system equation, whose last $n - 1$ elements, corresponding to the number of state parameters in Eq. (9), are forced to be zero. All the other entries in the structure are filled in with NA’s. This augmented structure will have additional columns when the state-space model contains more than one evolution equation.

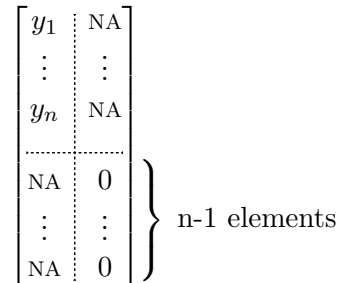


Diagram 1. Schematic representation of the data structure for the augmented model.

Inference in this augmented model using the INLA approach is performed considering two different likelihoods, one for each column of the augmented structure. Hence, given the states, x_t , the actual data points in the first column of the augmented structure are assumed to follow a Gaussian distribution with unknown precision V^{-1} . The artificial observations in the second column are deterministically known (i.e., with zero variance) conditioned on the values of x_t , x_{t-1} and ω_t in (9). This condition is represented in INLA by assuming that these “faked” observations follow a Gaussian distribution with a high and fixed precision.

In order to estimate the states, x_t in this model, it is necessary just to know the perturbations ω_t , $t = \{2, \dots, T\}$, which are the only stochastic terms in the system equation (8). Thus, we are considering that there is no information about x_t beyond its temporal evolution form. In the formulation of the model this information is induced through vectors of indices, say i , j and l , associated to x_t , x_{t-1} and ω_t terms, respectively. The x_{t-1} terms are modeled as a copy

of the x_t terms, making the value at position $k + 1$ in index vector j equal to the value at position k in index vector i ; this lag induces the temporal evolution of x_t . Since we are not considering stochasticity in the states, for the formulation of the model we must leave the terms x_t free to assume values in any region of the parametric space. This is achieved defining \mathbf{x} as a vector of independent and Gaussian distributed random variables with a fixed and low precision (high variance). At $t = 1$, we have that x_t comes directly from this uninformative distribution, whereas from $t = 2, \dots, n$, x_t follows Eq (8). Let \mathbf{y} be a $n \times 1$ vector of observations simulated from the dynamic linear model described above. The following R code implements the augmented approach to this model using the INLA library:

```
## building the augmented model
m <- n-1
Y <- matrix(NA, n+m, 2)
Y[1:n, 1] <- y # actual observations are stored in the 1st column
Y[1:m + n, 2] <- 0 # faked observations (assuming zero values) are in the 2nd column
## indices for the INLA library
i <- c(1:n, 2:n) # indices for the states x_t
j <- c(rep(NA,n), 2:n -1) # indices for x_{t-1}
w1 <- c(rep(NA,n), rep(-1,m)) # weights for x_{t-1}
l <- c(rep(NA,n), 2:n) # indices for the perturbation terms w_t
w2 <- c(rep(NA,n), rep(-1,m)) # weights for w_t
## formulating the model
formula <- Y ~ f(i, model="iid", initial=-10, fixed=T) +
            f(j, w1, copy="i") + f(l, w2, model ="iid") -1
## call to fit the model
require(INLA)
r <- inla(formula, data = data.frame(i,j,w1,w2,l),
          family = c("gaussian", "gaussian"),
          control.data = list(list(), list(initial=10, fixed=T)))
```

Note that two different Gaussian likelihoods are defined in the call to fit the augmented model (for the actual and faked observations). The list in "control.data" statement defines the parameters for the log-precisions of the actual and faked observations, respectively. Default values are assumed in the first case, whereas a high and fixed value is defined for the log-precision in the second case. Weights w_1 and w_2 are needed in the formulation of the lagged states, x_{t-1} , and perturbations, w_t , as these terms are negative in Eq (9).

To be properly considered by the INLA library, each term on the right side of the observational and system equations must be indexed, according to its corresponding time index, in the dataframe to be passed to INLA. For example, the state vector, x_t , appears in Eq. (7) at times $t = 1, 2, \dots, n$ as well as in Eq. (8) at times $t = 2, 3, \dots, n$. Therefore, the index vector, "i" for this term is specified in the dataframe as $[1, 2, \dots, n, 2, 3, \dots, n]'$. Note that each column in the dataframe passed to INLA must be a vector of dimension equal to the number of rows of the augmented structure, which is $n + (n - 1)$ in this example. When a term is present in just one part of the augmented structure, the remaining elements of the index vector for that term in the dataframe are filled in with NA's. This is the case of term x_{t-1} , which appears just in Eq. (8). Its index vector, "j", then is given by $[NA, \dots, NA, 1, 2, \dots, n - 1]'$, where the first n elements are NA's.

The series of simulated and estimated values for the vectors of observations and states, respectively, are illustrated in Figure 1. Precisions of the perturbation terms were well estimated in all cases, as can be seen in Figure 2, with credibility intervals always including the true simulated values.

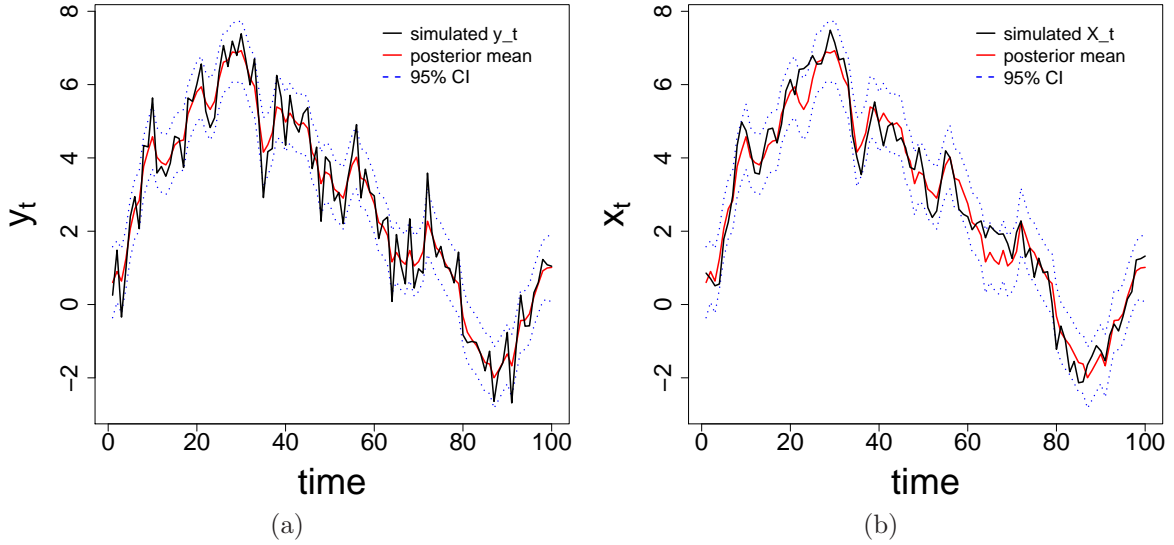


Figure 1: Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and states (b) in the toy example.

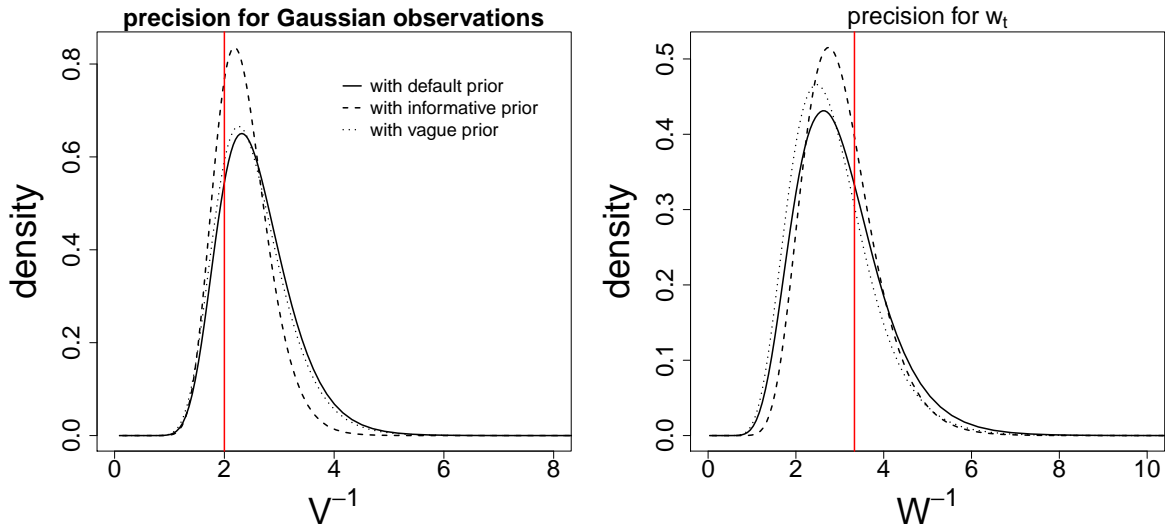


Figure 2: Posterior densities for the hyperparameters in the toy example. Red lines indicate true simulated values.

4.1.1 Sensitivity of the fixed precisions

A sensitivity analysis was performed to check the influence of the initial values for both the fixed log-precision of the “faked” observations (p_f) and the fixed log-precision of the states (p_s) on the posterior estimates of the hyperparameters $\tau_V = V^{-1}$ and $\tau_W = W^{-1}$ as well as on the predictive performance of the model. To achieved this, 1000 data sets of size $n = 100$ were simulated from the model in (7) and (8), with variances of the error terms for the measurement and state equations at each time series being randomly chosen from the intervals $[0.01, 2]$ and $[0.01, 1]$, respectively. Each data set was fitted 225 times with different combinations of initial values for p_f and p_s from a 15×15 grid, where $p_f \in \{1, 2, \dots, 14, 15\}$ and $p_s \in \{-1, -2, \dots, -14, -15\}$. As sensitivity criteria we adopted the mean absolute error (MAE) of the posterior mean of τ_V and τ_W and two model selection measures available from the INLA output, the deviance

information criterion, DIC, (Spiegelhalter et al., 2002) and the conditional predictive ordinate, CPO, (Pettit, 1990). Following Roos and Held (2011) we used the mean logarithmic CPO, which measures the predictive quality of a model and is defined as $\overline{LCPO} = -\frac{1}{n} \sum_{i=1}^n \log(CPO_i)$, where $CPO_i = \pi(y_i | \mathbf{y}_{-i})$ and the subscript $-i$ indicates that element i of vector of observations \mathbf{y} was removed. Lower values of MAE, DIC and \overline{LCPO} indicate a better model.

It should be remembered that inference here is being performed in an augmented model framework. However, to be able to compare the DIC and CPO predictive measures across models fitted with different choices of p_f and p_s , these measures must be constructed considering just the actual n observations, since the additional $n - 1$ “faked” observations of the augmented model were created just as an artefact to allow the modeling of the states but do not constitute real observations itself. These quantities are easily obtained as a sum of the local DIC and CPO measures for the first n observations, as they are readily available from the standard INLA output. This will also be necessary when these measures are used for model comparison between models using the augmented model framework with those using other inference procedures, as well as when model choice criteria based on the marginal likelihood were considered. However, the INLA library currently does not provide local marginal likelihood statistics.

The results of this analysis are summarized in the box plots of Figure 3. According to all sensitivity criteria the worst results were found when the model was fitted with the higher initial values for the fixed log-precision of the states. Predictive performance of the model and estimation error of the hyperparameters were improved as this initial value decreased up to -5, remaining stable from this value onwards. On the other hand, different initial values of the fixed log-precision for the “faked” observations had no influence on the predictive criteria for this model, whereas the MAE of τ_V and τ_W became lower with increases in p_f 's initial values up to 5, after of which it remained virtually constant. Based on these results we chose -10 and 10 as “safe” initial values for p_s and p_f , respectively. These values worked fine for all the examples in this paper. However, for particular models a sensitivity analysis to address the choice of these initial values, as described here, might be desirable.

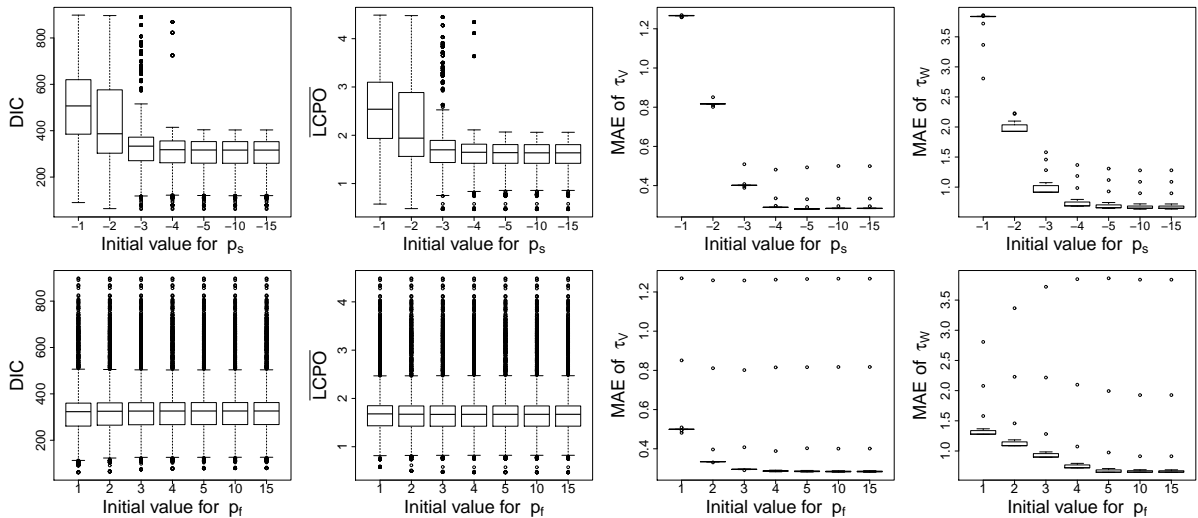


Figure 3: Boxplots of the Deviance information criterion (DIC), the mean logarithmic CPO (\overline{LCPO}) and the mean absolute error (MAE) of the posterior mean of hyperparameters τ_V and τ_W for different initial values of the fixed log-precision of the states (upper frame) and “faked” observations (lower frame) in the toy example.

4.1.2 Comparison with MCMC

The same 1000 simulated data sets generated in subsection 4.1.1 were also used to compare the performance between INLA and MCMC approaches under different prior assumptions for the hyperparameters. To achieve this, the model in (7) and (8) was fitted with this two approaches for each simulated time series under informative, vague and default log-gamma priors for τ_V and τ_W , as specified at the beginning of this section. Following results in the previous subsection fixed precisions in the INLA approach were initialized as $p_s = -10$ and $p_f = 10$ and initial values for the hyperparameters were specified at their default settings. The MCMC algorithm was implemented in R. Inverse Gamma prior distributions for V and W were specified to match the values of the log-gamma priors for $\log(\tau_V)$ and $\log(\tau_W)$ in the INLA approach. The Gaussian prior for x_0 was set as $N(0,10)$. The hyperparameters were sampled through their full conditionals whereas the states, \mathbf{x} , were sampled through FFBS (forward filtering backward sampling – Frühwirth-Schnatter, 1994; Carter and Kohn, 1994). From a unique MCMC chain 55000 posterior samples were drawn by using a thinning of 20 and a burn-in of 5000. Hence, 2500 effective samples were left for an estimation of the posterior quantities. The CODA package (Plummer et al., 2006) was used to diagnose convergence through the Geweke and Raftery–Lewis diagnostic tests. The comparison was carried out considering the mean absolute error (MAE) and the root mean squared error (RMSE) of the posterior mean of τ_V and τ_W as well as the 95% coverage percent of the credibility intervals for these hyperparameters.

Results summarized in Table 1 show a very similar performance with the two approaches in terms of both (absolute and squared) errors and percent coverage for the two hyperparameters when informative priors are used. On the other hand, results were clearly better with INLA when the default prior was considered. Errors in this scenario were more than 50% and more than 100% higher with MCMC for the observation and state variances, respectively. Percent coverage was also better (more than 10% higher) with INLA for both hyperparameters. Results in the case of vague priors were mixed, with error measures showing similar results in both approaches but with percent coverage being a bit better with the MCMC approach. The mean computer time to run each one of the 3000 models (1000 data sets \times 3 types of priors) was vastly favorable to the INLA approach, with 2.1 seconds against 260.3 seconds spent by the MCMC algorithm.

Table 1: Mean absolute error (MAE), root mean squared error (RMSE) and percent coverage of the 95% credibility interval (cover95) for the state (W) and observation (V) variances of a first order univariate dynamic linear model based on 1000 runs of the INLA and MCMC approaches under three different prior assumptions (see text for details).

parameter	measure	type of prior (INLA)			type of prior (MCMC)		
		default	informative	vague	default	informative	vague
V	MAE	0.2087	0.1436	0.1946	0.3188	0.1436	0.1962
	RMSE	0.2810	0.1956	0.2596	0.4491	0.1980	0.2633
	cover95	87.3	96.4	91.5	75.4	96.8	94.3
W	MAE	0.1711	0.0852	0.1526	0.3980	0.0902	0.1665
	RMSE	0.2398	0.1221	0.2156	0.7041	0.1324	0.2414
	cover95	82.4	98.3	89.7	72.9	98.9	92.5

4.1.3 Impact of the augmented data framework on inference

The proposed augmented model framework slightly modifies the standard structure of the state-space model introducing artificial observations following additional likelihoods (just one in our toy example), which allow the modeling of the states within INLA. Fixed precisions, associated to these new “faked” observations and to the states, are also added to the model. Here we used the first order univariate dynamic linear model of subsection 4.1, which can be fitted in INLA with both the standard and the augmented approaches, to exemplify how these new variables impact the inference of the joint predictive distribution of the observations given the hyperparameters when compared with that obtained under the standard modeling approach.

Predictive distribution under the standard approach

First we will obtain the joint distribution of \mathbf{y} given the precision parameters when a first order random walk (RW1) prior is assumed for the states \mathbf{x} . This is a default option in the INLA approach. Rewriting equations 7 and 8 we can express the state-space model in the toy example in matricial form as

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\nu} \quad (10)$$

$$\mathbf{x}_{-1} = \mathbf{x}_{-n} - \boldsymbol{\omega} \quad (11)$$

with $\mathbf{y} = (y_1, \dots, y_n)$ and $\mathbf{x} = (x_1, \dots, x_n)$ denoting the observations and states vectors, respectively. $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)$ and $\boldsymbol{\omega} = (\omega_2, \dots, \omega_n)$ are the perturbation vectors for the observation and state equations, respectively, and \mathbf{x}_{-i} denotes the state vector \mathbf{x} without the i th entry. The likelihood for this model is

$$\mathbf{y}|\mathbf{x}, \tau_V \propto \exp \left\{ -\frac{\tau_V}{2} (\mathbf{y} - \mathbf{x})' (\mathbf{y} - \mathbf{x}) \right\}$$

and using the RW1 prior for \mathbf{x} , we obtain the posterior full conditional $\mathbf{x}|\tau_W \propto \exp \left\{ -\frac{\tau_W}{2} \mathbf{x}' \mathbf{R} \mathbf{x} \right\}$, with structure matrix

$$\mathbf{R} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{pmatrix}.$$

Then the joint distribution

$$\mathbf{y}|\tau_V, \tau_W \propto \int \exp \left\{ -\frac{\tau_V}{2} (\mathbf{y} - \mathbf{x})' (\mathbf{y} - \mathbf{x}) - \frac{\tau_W}{2} \mathbf{x}' \mathbf{R} \mathbf{x} \right\} d\mathbf{x}$$

after some calculations can be expressed as

$$\mathbf{y}|\tau_V, \tau_W \propto \exp \left\{ -\frac{1}{2} \mathbf{y}' (\tau_V \mathbf{I}_n - \tau_V^2 \mathbf{Q}_X^{-1}) \mathbf{y} \right\}, \quad (12)$$

with \mathbf{I}_n being an $n \times n$ identity matrix and

$$\mathbf{Q}_X = (\tau_V \mathbf{I}_n + \tau_W \mathbf{R}). \quad (13)$$

Predictive distribution under the augmented model

Following the framework at the beginning of subsection 4.1 we equate to zero the system equation (11), which can be rewritten in matricial form as

$$\mathbf{0} = \mathbf{d} - \boldsymbol{\omega} \text{ ,}$$

where $\mathbf{d} = (x_2 - x_1, x_3 - x_2, \dots, x_n - x_{n-1})$, $\mathbf{0}$ is an $n - 1$ dimensional vector of zeros (the “faked” observations) with an additional likelihood given by

$$\mathbf{0} | \mathbf{d}, \boldsymbol{\omega} \propto \exp \left\{ -\frac{\tau_0}{2} (\boldsymbol{\omega} - \mathbf{d})' (\boldsymbol{\omega} - \mathbf{d}) \right\} \text{ ,}$$

and τ_0 is a large fixed precision for the difference between \mathbf{d} and $\boldsymbol{\omega}$. Gaussian independent priors are assumed for \mathbf{x} and $\boldsymbol{\omega}$ as

$$\mathbf{x} | \tau_x \propto \exp \left\{ -\frac{\tau_x}{2} \mathbf{x}' \mathbf{x} \right\} \quad \text{and} \quad \boldsymbol{\omega} | \tau_W \propto \exp \left\{ -\frac{\tau_W}{2} \boldsymbol{\omega}' \boldsymbol{\omega} \right\} \text{ ,}$$

with τ_x being a large fixed precision for the states \mathbf{x} .

To obtain the joint predictive distribution of \mathbf{y} , given the hyperparameters we integrate out \mathbf{x} and $\boldsymbol{\omega}$ yielding

$$\begin{aligned} \mathbf{y} | \tau_V, \tau_W, \tau_0, \tau_x &\propto \int \int \exp \left\{ -\frac{\tau_V}{2} (\mathbf{y} - \mathbf{x})' (\mathbf{y} - \mathbf{x}) - \frac{\tau_0}{2} (\boldsymbol{\omega} - \mathbf{d})' (\boldsymbol{\omega} - \mathbf{d}) \right\} \times \\ &\quad \exp \left\{ -\frac{\tau_x}{2} \mathbf{x}' \mathbf{x} - \frac{\tau_W}{2} \boldsymbol{\omega}' \boldsymbol{\omega} \right\} d\boldsymbol{\omega} d\mathbf{x} \\ &\propto \exp \left\{ -\frac{\tau_V}{2} \mathbf{y}' \mathbf{y} \right\} \int \exp \left\{ -\frac{\tau_V}{2} (\mathbf{x}' \mathbf{x} - 2\mathbf{y}' \mathbf{x}) - \frac{\tau_0}{2} \mathbf{d}' \mathbf{d} - \frac{\tau_x}{2} \mathbf{x}' \mathbf{x} \right\} \times \\ &\quad \underbrace{\int \exp \left\{ -\frac{\tau_0}{2} (\boldsymbol{\omega}' \boldsymbol{\omega} - 2\mathbf{d}' \boldsymbol{\omega}) - \frac{\tau_W}{2} \boldsymbol{\omega}' \boldsymbol{\omega} \right\} d\boldsymbol{\omega}}_{\text{(II)}} d\mathbf{x} \text{ ,} \end{aligned}$$

where (II) $\propto \exp \left\{ \frac{\tau_0^2}{2(\tau_0 + \tau_W)} \mathbf{d}' \mathbf{d} \right\}$ and $\mathbf{d}' \mathbf{d} = \mathbf{x}' \mathbf{R} \mathbf{x}$. Additionally, we have that

$$\tau_0 - \frac{\tau_0^2}{\tau_0 + \tau_W} = \frac{\tau_0}{\tau_0 + \tau_W} \tau_W \text{ , implying that}$$

$$\begin{aligned} \mathbf{y} | \tau_V, \tau_W, \tau_0, \tau_x &\propto \exp \left\{ -\frac{\tau_V}{2} \mathbf{y}' \mathbf{y} \right\} \times \\ &\quad \int \exp \left\{ -\frac{1}{2} \left[\mathbf{x}' \left((\tau_V + \tau_x) \mathbf{I}_n + \tau_0 \mathbf{R} - \frac{\tau_0^2}{\tau_0 + \tau_W} \mathbf{R} \right) \mathbf{x} - 2\tau_V \mathbf{y}' \mathbf{x} \right] \right\} d\mathbf{x} \\ &\propto \exp \left\{ -\frac{\tau_V}{2} \mathbf{y}' \mathbf{y} \right\} \int \exp \left\{ -\frac{1}{2} (\mathbf{x}' \mathbf{Q}_{X^*} \mathbf{x} - 2\mathbf{y}' \mathbf{x}) \right\} d\mathbf{x} \\ &\propto \exp \left\{ -\frac{1}{2} (\tau_V \mathbf{y}' \mathbf{y} - \tau_V^2 \mathbf{y}' \mathbf{Q}_{X^*}^{-1} \mathbf{y}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \mathbf{y}' (\tau_V \mathbf{I}_n - \tau_V^2 \mathbf{Q}_{X^*}^{-1}) \mathbf{y} \right\} \text{ ,} \end{aligned} \tag{14}$$

with,

$$\mathbf{Q}_{X^*} = (\tau_V + \tau_x) \mathbf{I}_n + \left(\frac{\tau_0}{\tau_0 + \tau_W} \tau_W \right) \mathbf{R} \text{ .} \tag{15}$$

Comparing equations (12), (13) with (14), (15) we note that the impact of the augmented model framework on the joint distribution of \mathbf{y} given the hyperparameters is due to two factors

in matrix \mathbf{Q}_{X^*} . The first one is the shift of τ_V coming from the fixed precision τ_x . The second one is the $\tau_0/(\tau_0 + \tau_W)$ reduction of τ_W due to the fixed precision τ_0 assumed for the “faked” observations. However, it is easy to see that \mathbf{Q}_{X^*} converges to \mathbf{Q}_X as $\tau_0 \rightarrow \infty$ and $\tau_x \rightarrow 0$, which is achieved in practice letting τ_0 and τ_x to assume a priori large and small fixed values, respectively, as shown in subsection 4.1.1.

Further, we can establish criteria to set the values of these fixed precisions based on its associated factors in matrix \mathbf{Q}_{X^*} . For instance, we could desire that τ_x be less than r , for a certain $r \approx 0$. On the other hand, factor $0 < \tau_0/(\tau_0 + \tau_W) < 1$ depends on τ_W . Therefore, for a certain $p \approx 1$, $\tau_0/(\tau_0 + \tau_W) > p$ means that τ_0 should be greater than $\frac{p}{1-p}\tau_W$, and a first guess for τ_W could be obtained from a preliminary run of the model using the default fixed log-precision values for $\log(\tau_0)$ and $\log(\tau_x)$ suggested in subsection 4.1.1.

4.2 Applying current INLA modeling options

The INLA approach currently offers some functionalities to deal with first and second order random walk and with seasonal models that can be directly used to fit simpler state-space models. This is illustrated in this subsection through two examples where a generalized dynamic regression model and a dynamic seasonal model are fitted using existing model options from the INLA library. What both of these examples have in common is the independence among the error terms and the simple evolution form for the states. Examples of more complex models, where these assumptions are not valid and that are unable to be fitted using INLA’s standard tools, are considered in the next subsection.

Example 1: A dynamic seasonal model

In this example we simulate a monthly time series with an annual seasonal pattern by using a cosine form. The response y_t is assumed to be normal. The DLM formulation is then defined as

$$y_t = a_t \cos\left(\frac{\pi(t-1)}{6}\right) + \nu_t, \quad \nu_t \sim N(0, V), \quad t = 1, \dots, n$$

$$a_t = a_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \quad t = 2, \dots, n$$

This model can easily be fitted just using a first order random walk model in INLA, as follows:

```
i <- t <- 1:n
cosw <- cos(pi*(t-1)/6)
formula <- y ~ f(i, cosw, model="rw1", constr=F) -1
r <- inla(formula, data = data.frame(i,y))
```

The results are summarized in Figures 4 and 5:

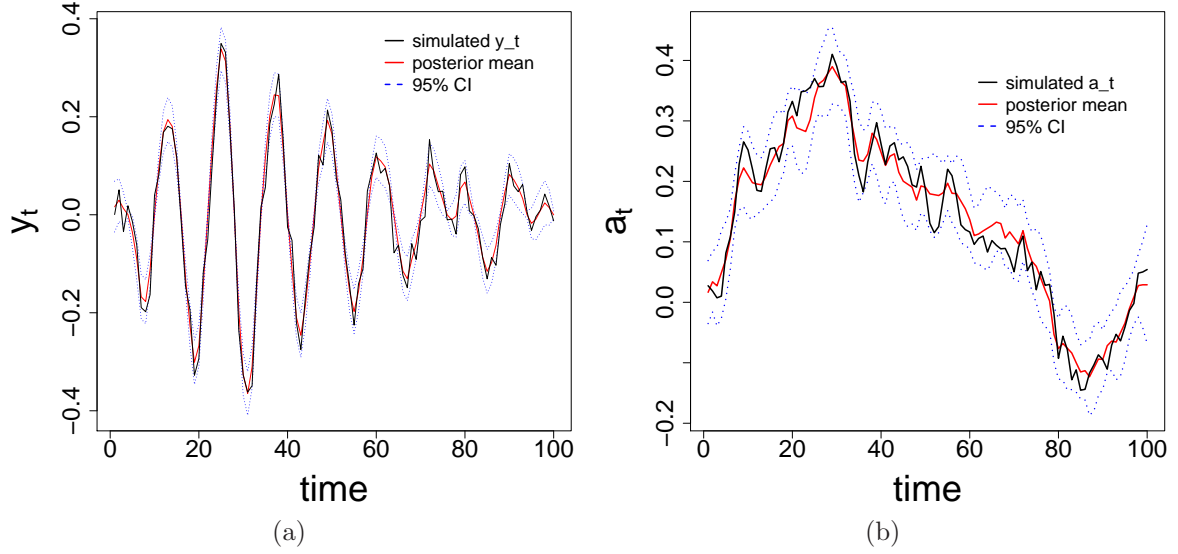


Figure 4: Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and for a_t state vector (b) in the dynamic seasonal model.

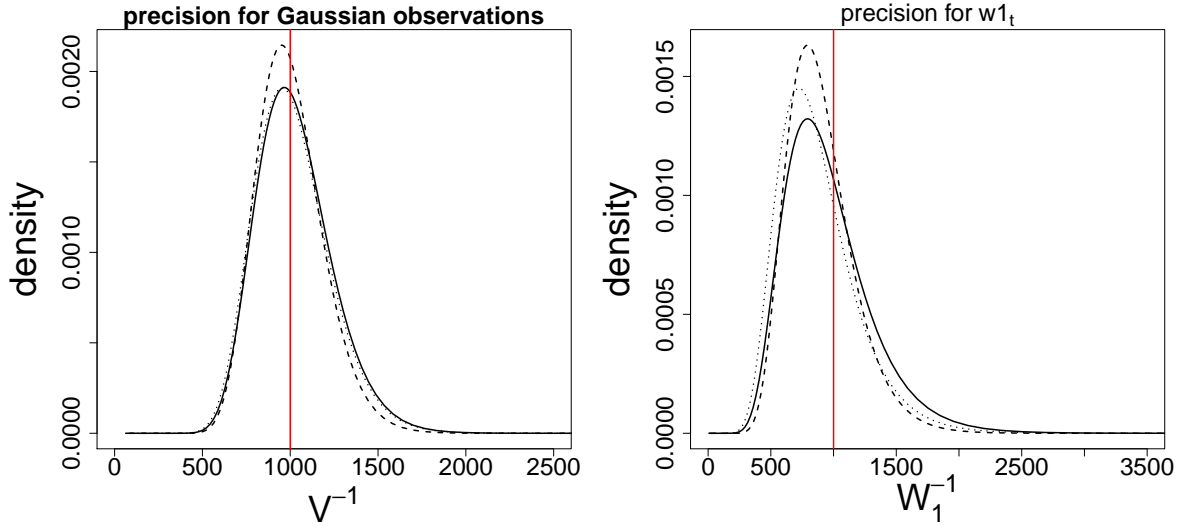


Figure 5: Posterior densities for the hyperparameters in the dynamic seasonal model. Red lines indicate true simulated values.

Example 2: A generalized dynamic regression model

Here we simulate data from a multiple Poisson regression model with two regressors, Z_{1t} and Z_{2t} . Thus, the linear predictor is given by $\lambda_t = \mathbf{F}_t \mathbf{x}_t$, where $\mathbf{F}_t = (1, Z_{1t}, Z_{2t})$ and the regression coefficients $\mathbf{x}_t = (\beta_{0t}, \beta_{1t}, \beta_{2t})$ follow a simple random walk evolution. The model has the following observational and system equations:

$$\begin{aligned}
 (y_t | \mu_t) &\sim \text{Poisson}(\mu_t) \\
 \log(\mu_t) &= \lambda_t = \beta_{0t} + \beta_{1t}Z_1 + \beta_{2t}Z_2 & t = 1, \dots, n \\
 \beta_{0t} &= \beta_{0,t-1} + \omega_{0t}, & \omega_{0t} &\sim N(0, W_0), & t = 2, \dots, n \\
 \beta_{1t} &= \beta_{1,t-1} + \omega_{1t}, & \omega_{1t} &\sim N(0, W_1), & t = 2, \dots, n \\
 \beta_{2t} &= \beta_{2,t-1} + \omega_{2t}, & \omega_{2t} &\sim N(0, W_2), & t = 2, \dots, n
 \end{aligned}$$

The interest is to obtain full Bayes estimates of the dynamic regression coefficients, $\beta_{it} = \{\beta_{i1}, \dots, \beta_{in}\}$, and the variance parameters, W_i , $i \in \{0, 1, 2\}$.

Note that existing well known approaches to perform sequential analysis of generalized dynamic linear models (GDLMs), like that in West et al. (1985) using linear Bayes estimation or those available in the R packages `sspir` and `KFAS`, are not able to fit this multiple model or just provide punctual estimates of the parameters of interest. In contrast, the INLA library provides full Bayes estimates of the parameters for this and other models in this class in an easy and fast way (3.7 seconds in this example), using its default first order random walk option to model the evolution of the regression coefficients. The comparison between simulated and predicted regression coefficients (posterior mean and 95% credibility interval) is shown in Figure 6. Figure 7 presents the marginal posterior densities of the hyperparameters.

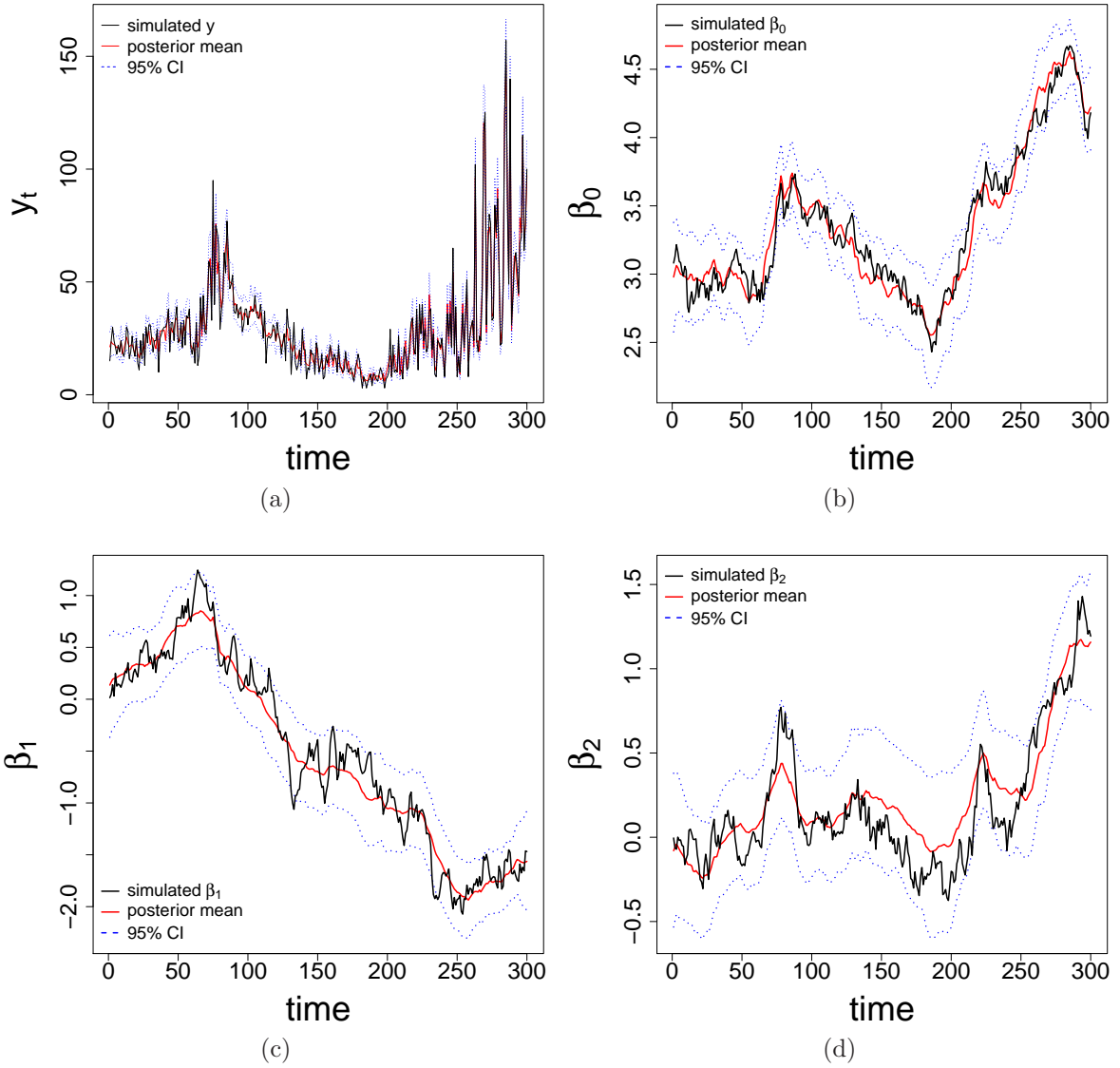


Figure 6: Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and regression coefficients, β_0 , β_1 and β_2 (b-d) in the generalized dynamic regression example.

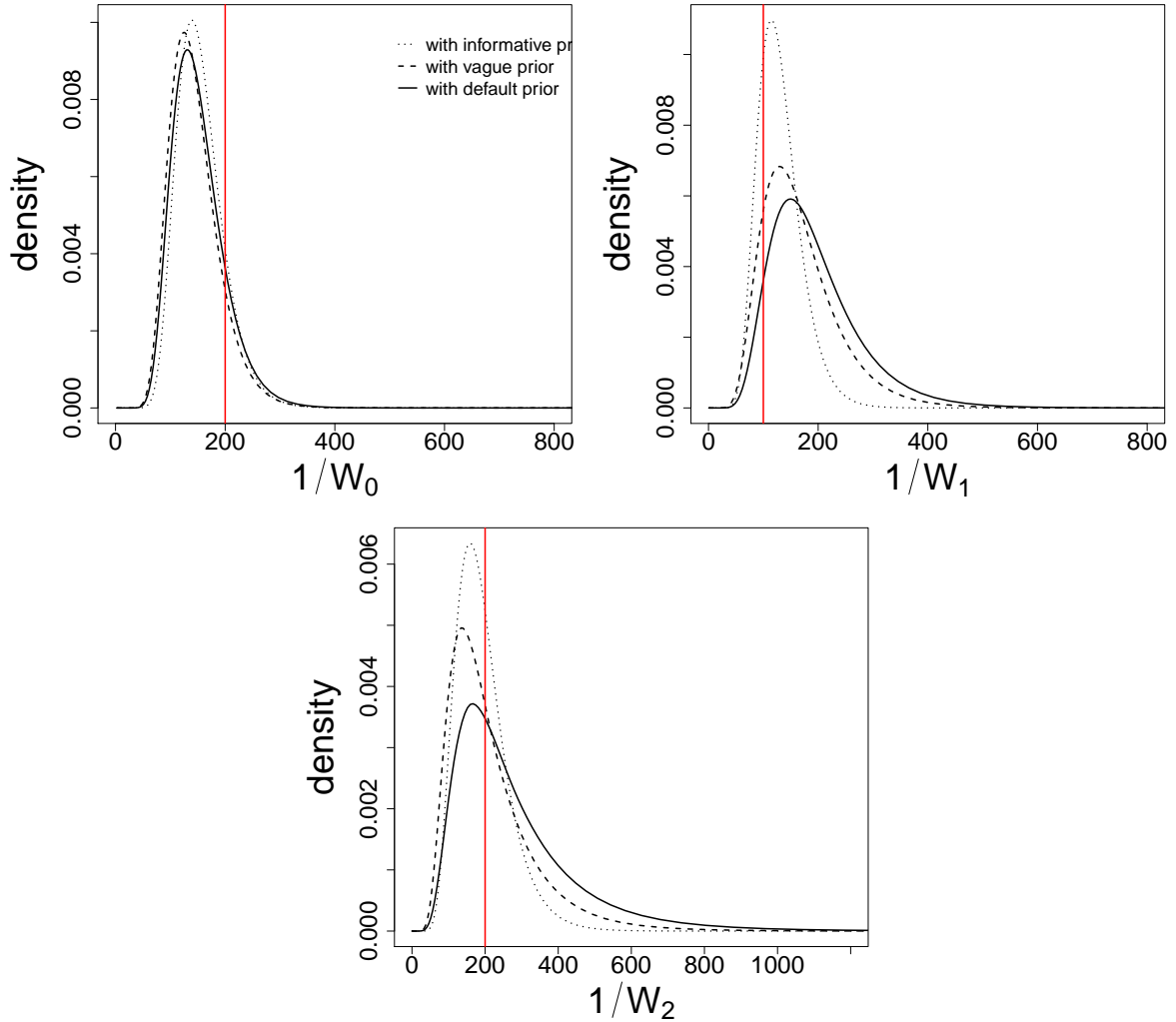


Figure 7: Marginal posterior densities for the hyperparameters in the generalized dynamic regression example. Red lines indicate true simulated values.

4.3 Extending the INLA modeling capabilities

As seen in the above subsection, the class of state-space models that can be fitted under the INLA approach using current tools to fit latent Gaussian models, is restricted to models with a simple temporal evolution form for the states and with an independence assumption for the error terms. This is not the case, for example, of growth models and spatio-temporal dynamic models, where there can be more than one state vector, being mutually dependent on each other, and where the error terms can be structured as a matrix with correlated values. The proposed framework for inference described in subsection 4.1 extends the INLA approach in order to perform approximate Bayesian inference in these more complex cases. The examples in this subsection illustrate how to apply the proposed framework in the above mentioned situations.

4.3.1 Models with different evolution structure for the states

Here a second order polynomial dynamic model and a dynamic seasonal model with two harmonics are considered to exemplify how to deal, under the INLA approach, with state-space models when the temporal evolution structure for the states involve more than one state vector.

Example 3: A second order polynomial dynamic model

This is a growth model with a state vector that comprises two elements, $x_t = (x_{1t}, x_{2t})$, the first representing the current level and the second representing the current rate of change in the level. The observational and system equations for this model are simulated as

$$y_t = x_{1t} + \nu_t, \quad \nu_t \sim N(0, V), \quad t = 1, \dots, n \quad (16)$$

$$x_{1t} = x_{1,t-1} + x_{2,t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \quad t = 2, \dots, n \quad (17)$$

$$x_{2t} = x_{2,t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), \quad t = 2, \dots, n \quad (18)$$

Note that the temporal evolution for x_{1t} in (17) depends on both, $x_{1,t-1}$ and $x_{2,t-1}$. This evolution structure can not be directly accounted for the INLA library, therefore, we made use here of the augmented model structure described in subsection 4.1, equating (17) and (18) to zero and then merging these “pseudo” observations with the actual observations from Eq. (16). The augmented structure has dimension $n+2(n-1)$ and three different likelihoods (see Diagram 2), being the elements in the first column Gaussian distributed with unknown precision V^{-1} , while the artificial observations in the other two columns are modeled as Gaussian with a high and fixed precision (see the toy example for details).

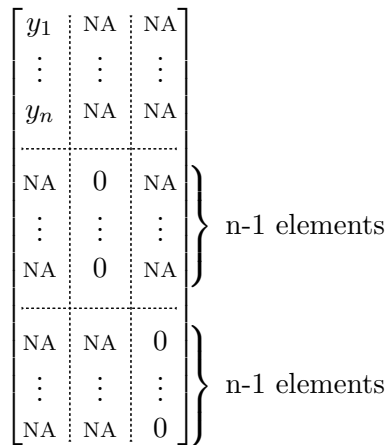


Diagram 2. Schematic representation of the data structure for the augmented model in the second order DLM.

Simulated and predicted values for observations and states are presented in Figure 8. The posterior densities for the three precision parameters are also shown in Figure 9.

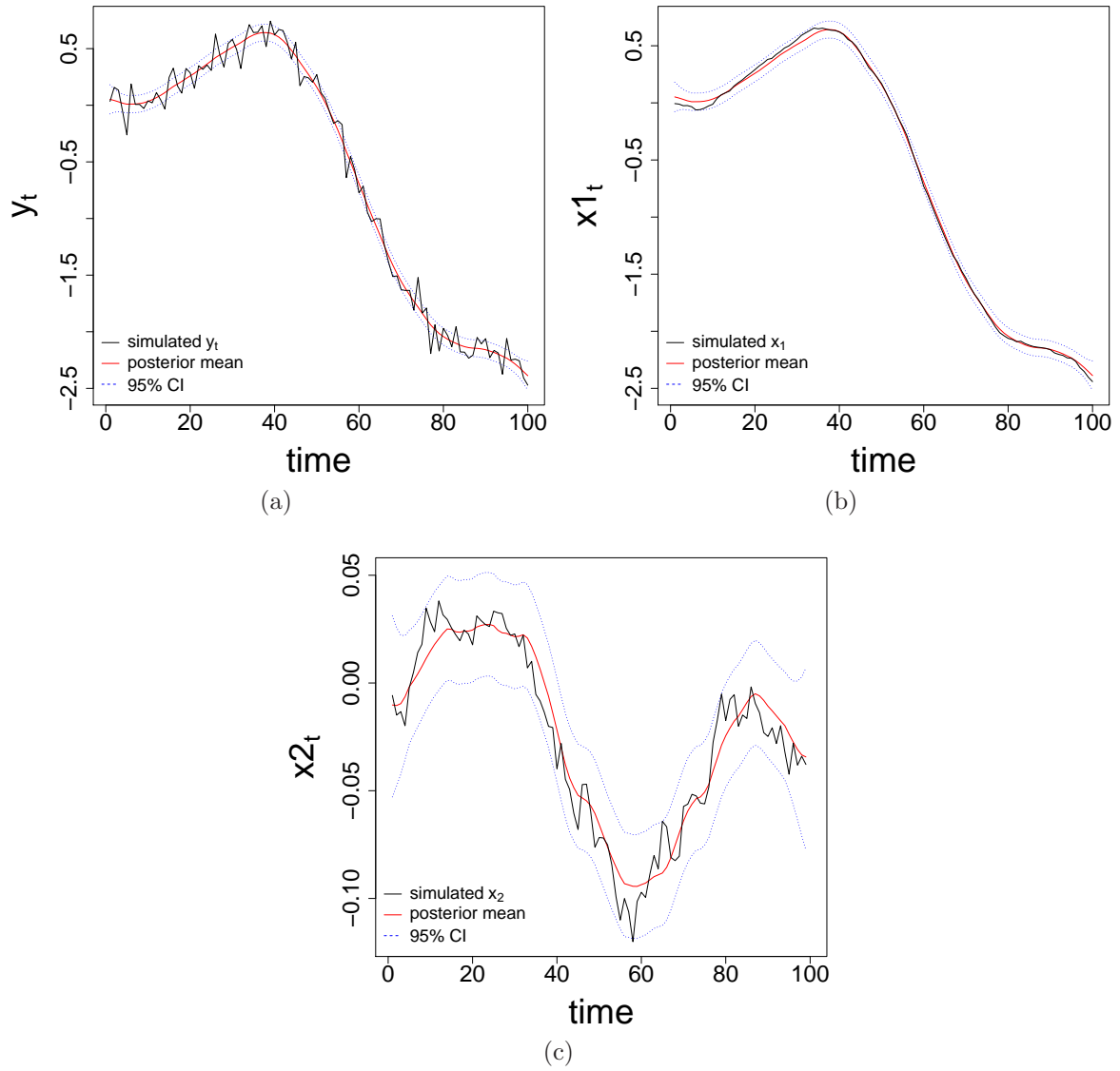


Figure 8: Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and for x_{1t} (b) and x_{2t} (c) state vectors in the second order polynomial dynamic model.

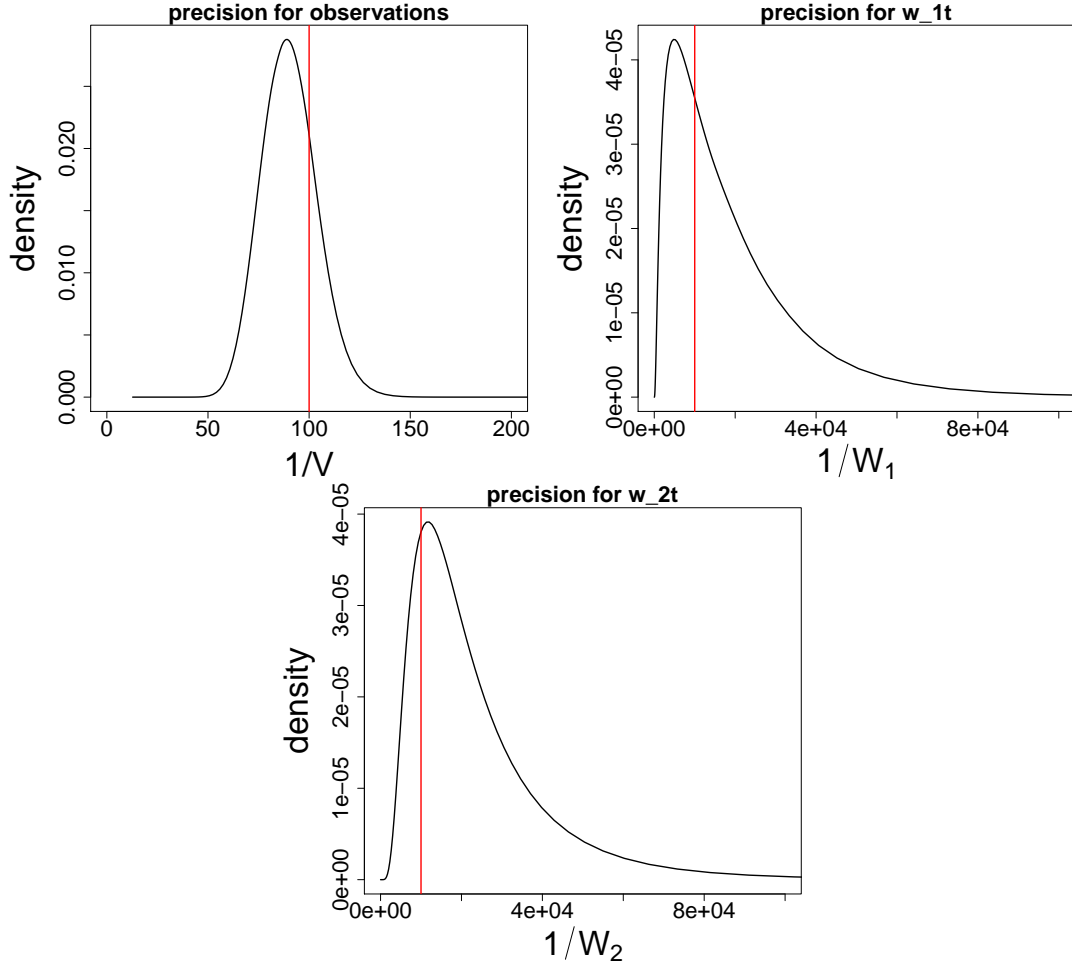


Figure 9: Marginal posterior densities for the hyperparameters in the second order polynomial dynamic model. Red lines indicate true simulated values.

Example 4: A seasonal dynamic model with harmonics

In this example we simulated a monthly time series with an annual seasonal pattern by using a sum of sine and cosine terms with the same frequency. This model has a two parameters state $x_t = (a_t, b_t)$ and can be defined as:

$$x_t = \begin{pmatrix} a_t \\ b_t \end{pmatrix}; \quad F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \quad G_t = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}; \quad \text{where } \phi = \frac{\pi}{6},$$

with observational and system equations given by

$$y_t = a_t + \nu_t, \quad \nu_t \sim N(0, V), \quad t = 1, \dots, n \quad (19)$$

$$a_t = \cos(\phi)a_{t-1} + \sin(\phi)b_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \quad t = 2, \dots, n \quad (20)$$

$$b_t = -\sin(\phi)a_{t-1} + \cos(\phi)b_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), \quad t = 2, \dots, n \quad (21)$$

Note that the temporal evolution for a_t in (20) depends on both, a_{t-1} and b_{t-1} . This evolution structure can not be directly accounted for the INLA library, therefore, we made use here of the augmented model structure described in subsection 4.1, equating (20) and (21) to zero and then merging these “pseudo” observations with the actual observations from Eq. (19). The augmented structure has dimension $n + 2(n - 1)$ and three different likelihoods,

being the elements in the first column Gaussian distributed with unknown precision V^{-1} , while the artificial observations in the other two columns are modeled as Gaussian with a high and fixed precision (see the toy example for details). The length of the simulated time series in this example was 110 months, but model fitting was performed using just the first 100 months, leaving out the last 10 ones to compare the simulated values with those forecasted using our approach. Estimates agreed quite well with simulated values, even in the forecast period. Computer time for this example was 2.4 seconds. Posterior means for the observations and states, including the forecast period, and its 90% credibility interval are shown in Figure 10. Marginal posterior densities of the hyperparameters can be found at Figure 11.

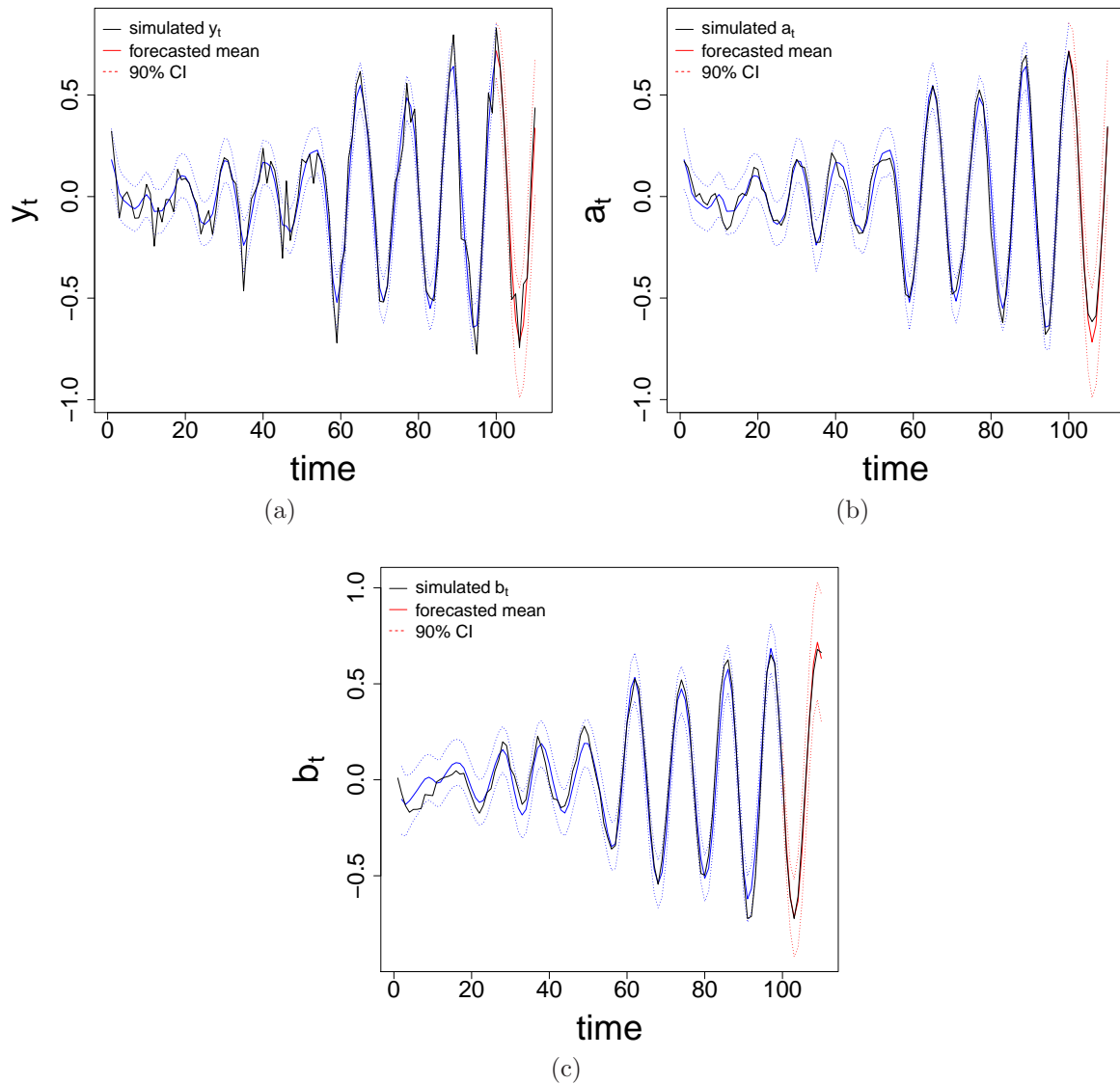


Figure 10: Simulated and predicted values (posterior mean and 90% credibility interval) for the observations (a) and for a_t (b) and b_t (c) state vectors in the harmonic seasonal model. Red lines are for the 10 steps ahead forecast.

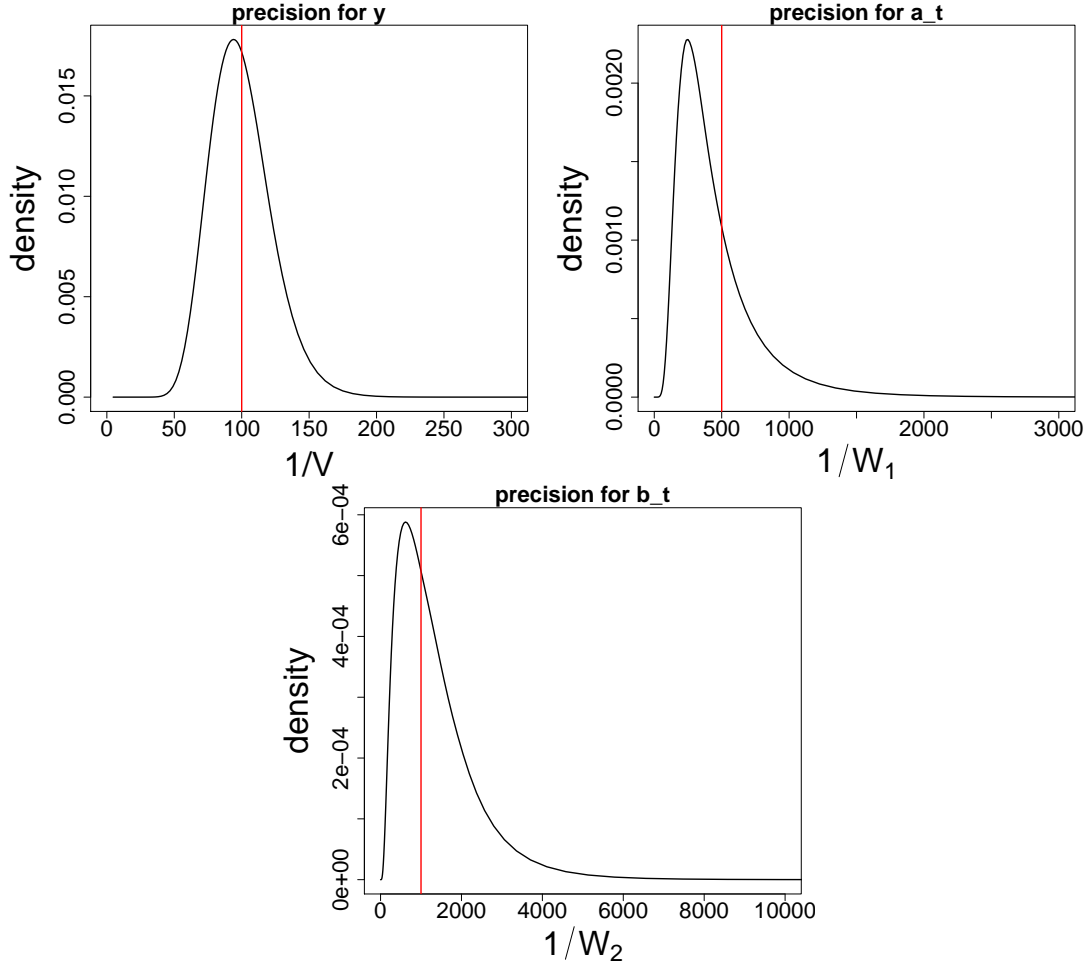


Figure 11: Marginal posterior densities for the hyperparameters in the harmonic seasonal model. Red lines indicate true simulated values.

4.3.2 Models with correlated error structure

In the following examples we simulate data from two different versions of a Gaussian spatio-temporal dynamic model for aerial data (Vivar and Ferreira, 2009) in order to demonstrate how the INLA library also can easily deal with correlated error structures in complex spatio-temporal dynamic models.

Example 5: A first order spatio-temporal dynamic model with covariates

We begin with a non-stationary first-order Gaussian model with one spatially structured covariate, where for each time t and area s , $t = \{1, \dots, T\}$; $s = \{1, \dots, S\}$, the response y_{ts} is specified as:

$$\mathbf{y}_t = \mathbf{F}'_t \mathbf{x}_t + \mathbf{\Upsilon}'_t \mathbf{z}_t + \boldsymbol{\omega}_{1t}, \quad \boldsymbol{\omega}_{1t} \sim \text{PGMRF}(\mathbf{0}_s, \mathbf{W}_1^{-1}) \quad (22)$$

$$\mathbf{x}_t = \mathbf{G}_t \mathbf{x}_{t-1} + \boldsymbol{\omega}_{2t}, \quad \boldsymbol{\omega}_{2t} \sim \text{PGMRF}(\mathbf{0}_s, \mathbf{W}_2^{-1}) \quad (23)$$

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \boldsymbol{\omega}_{3t}, \quad \boldsymbol{\omega}_{3t} \sim \text{PGMRF}(\mathbf{0}_s, \mathbf{W}_3^{-1}) \quad (24)$$

with $\mathbf{y}_t = (y_{t1}, \dots, y_{tS})'$ and $\mathbf{\Upsilon}_t = (\Upsilon_{t1}, \dots, \Upsilon_{tS})'$ denoting the observed field and the vector of covariates at time t , respectively. $\mathbf{F}_t = \mathbf{I}_s$, $\mathbf{G}_t = \rho \mathbf{I}_s$, $\mathbf{0}_s$ is an $S \times S$ null matrix, and \mathbf{I}_s is

the $S \times S$ identity matrix. The vectors of errors $\boldsymbol{\omega}_{i1}, \dots, \boldsymbol{\omega}_{iT}$, $i = \{1, 2, 3\}$, are assumed to be independent and modeled as proper Gaussian Markov random fields (PGMRF). Matrices \mathbf{W}_i describe the spatial covariance structure of $\boldsymbol{\omega}_{it} = (\omega_{it1}, \dots, \omega_{itS})'$. Precision matrices \mathbf{W}_i^{-1} are modeled as $\mathbf{W}_i^{-1} = \tau_i \left(\mathbf{I}_S - \frac{\phi_i}{\lambda_{max}} \mathbf{C} \right)$, with \mathbf{C} being a structure matrix defined as

$$C_{k,l} = \begin{cases} 0 & \text{if } k = l, \\ h_{k,l} & \text{if } k \in d_l, \\ 0 & \text{otherwise,} \end{cases}$$

d_l is the set of neighbors of area l , $h_{k,l} > 0$ is a measure of similarity between areas k and l (here we assume that $h_{k,l} = 1$). λ_{max} is the maximum eigenvalue of matrix \mathbf{C} ; τ_i are scale parameters and $0 \leq \phi_i < 1$ control the degree of spatial correlation.

We simulated a time series of 100 times for each of the 26 counties in the map of Eire (that is, $S = 26$ and $T = 100$). This map is available in R from `spdep` package (Bivand, 2010). Inference is performed for the state vectors \boldsymbol{x}_t and \boldsymbol{z}_t , as well as for the scale and correlation parameters, τ_i , ϕ_i , $i = \{1, 2, 3\}$, but not for ρ , whose value was fixed in one before analysis, leading to a non-stationary process.

For the implementation of this model using the INLA library, it is necessary to specify the precision matrices \mathbf{W}_1 and \mathbf{W}_2 through a generic model. This can be done using option `model='generic1'` in the formula to be called by the INLA library. This option requires that the structure matrix \mathbf{C} be passed as a file containing only the non-zero entries of the matrix. The file must contain three columns, where the first two ones contain the row and column indices of the non-zero entries of matrix \mathbf{C} , and the third column contain the corresponding non-zero values of structure matrix \mathbf{C} . The code in the appendix shows how this matrix can be built in R. For further details of how to specify structure matrices in INLA for use in the fitting of spatio-temporal models see Schrödle and Held (2009).

Predicted values closely followed the simulated series of observations and states in all cases, as illustrated in Figures 12 and 13 for the 18th area and its neighbors. A comparison between simulated and predicted observations and states for all areas at some instant times can also be seen in the maps of Figures 15 to 17. Precision and correlation parameters were also well estimated even when default INLA values for hyperprior parameters and initial values were specified (see Figure 14).

Using our framework it was possible to perform full Bayesian inference in this complex model with just a few lines of commands in R and spending less than five minutes on a modest laptop. Currently inference in models like this involves computationally expensive MCMC algorithms, which are also hard to tune.

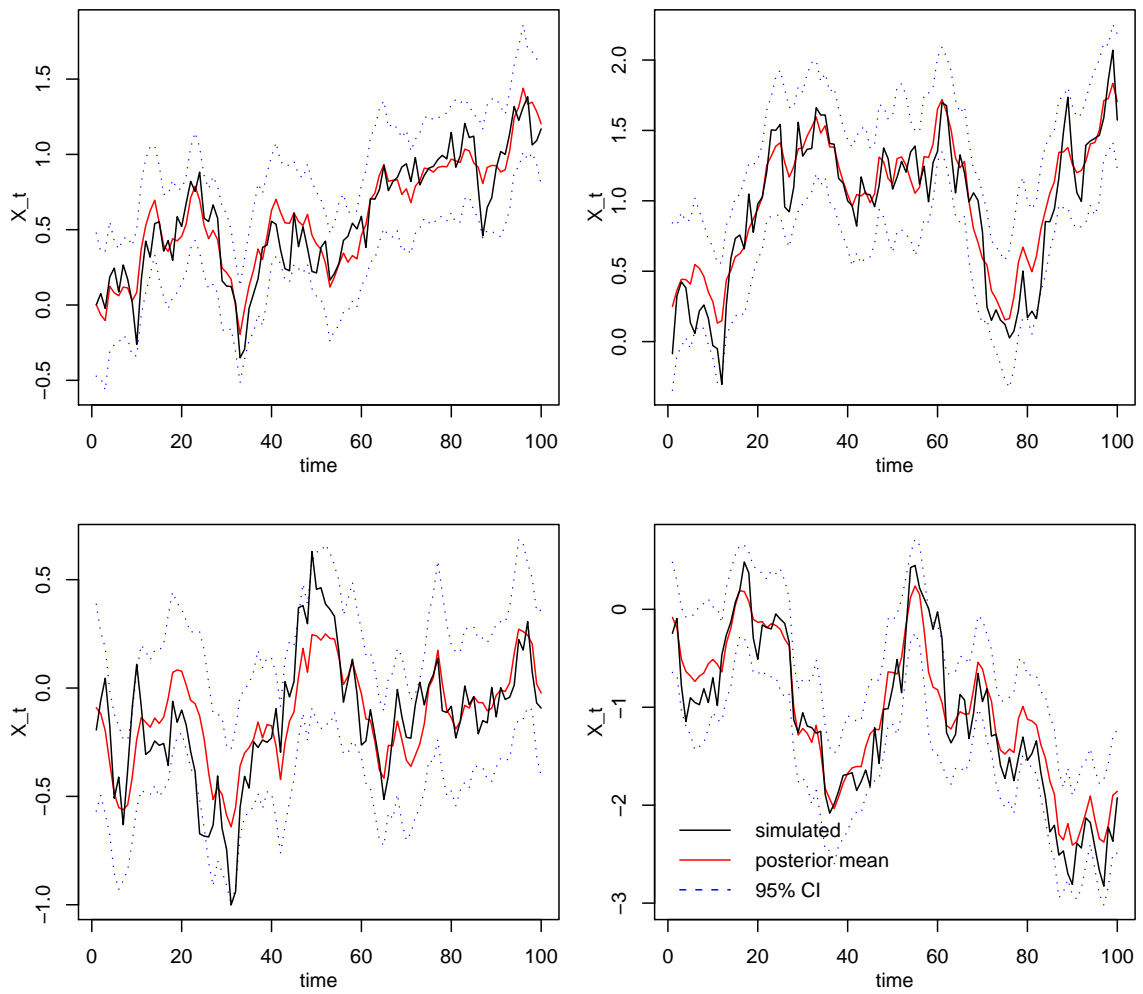


Figure 12: Simulated and predicted values (posterior mean and 95% credibility interval) for the states \mathbf{x}_t in the 18th area (a) and in its neighbors (b-d) in the first order spatio-temporal dynamic model.

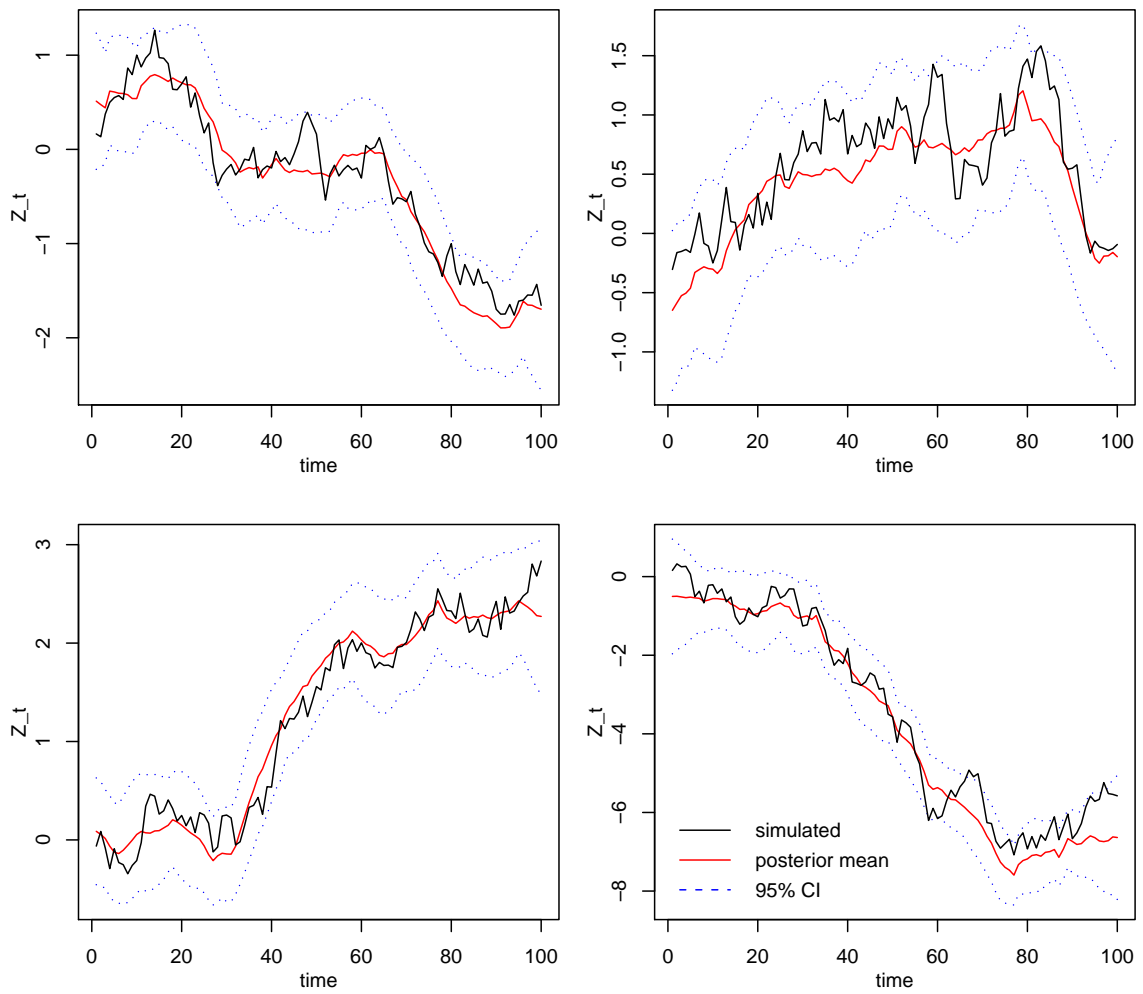


Figure 13: Simulated and predicted values (posterior mean and 95% credibility interval) for the states z_t in the 18th area (a) and in its neighbors (b-d) in the first order spatio-temporal dynamic model.

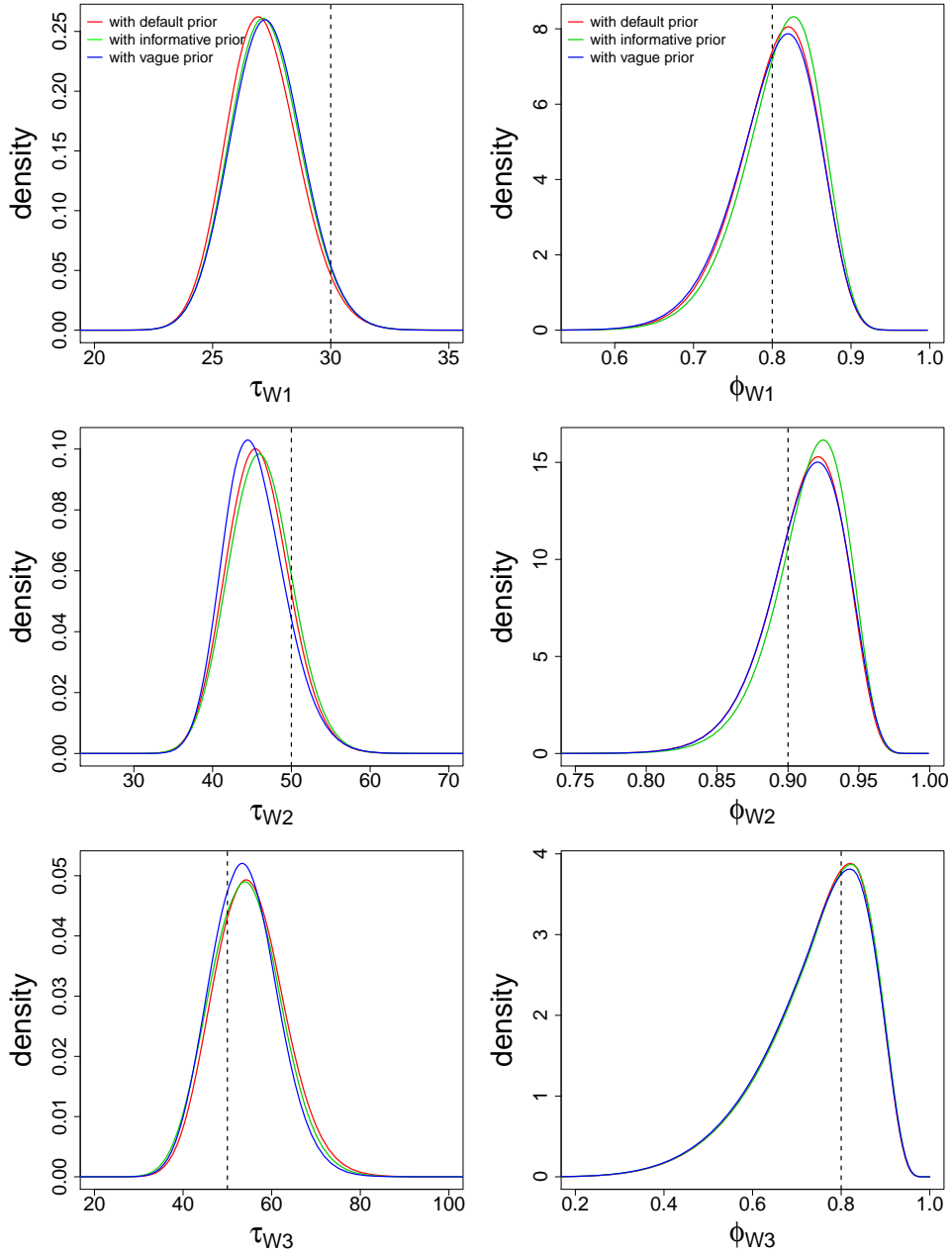


Figure 14: Marginal posterior densities for precision (left) and correlation (right) parameters in the first order spatio-temporal dynamic model using default, vague and informative priors for the precision parameters (see text for details). Vertical dashed lines indicate true simulated values.

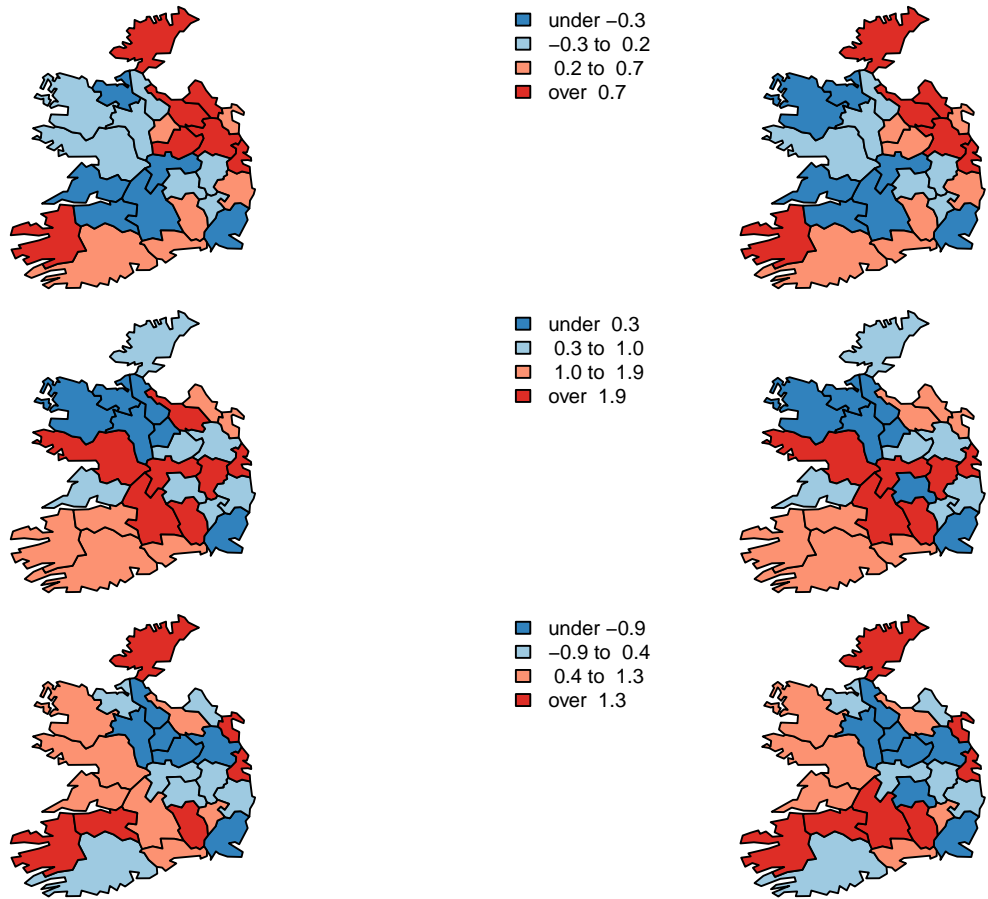


Figure 15: Maps of simulated (left) and predicted values (right) for observations, y_t , at times 18, 51 and 77 in example 5.

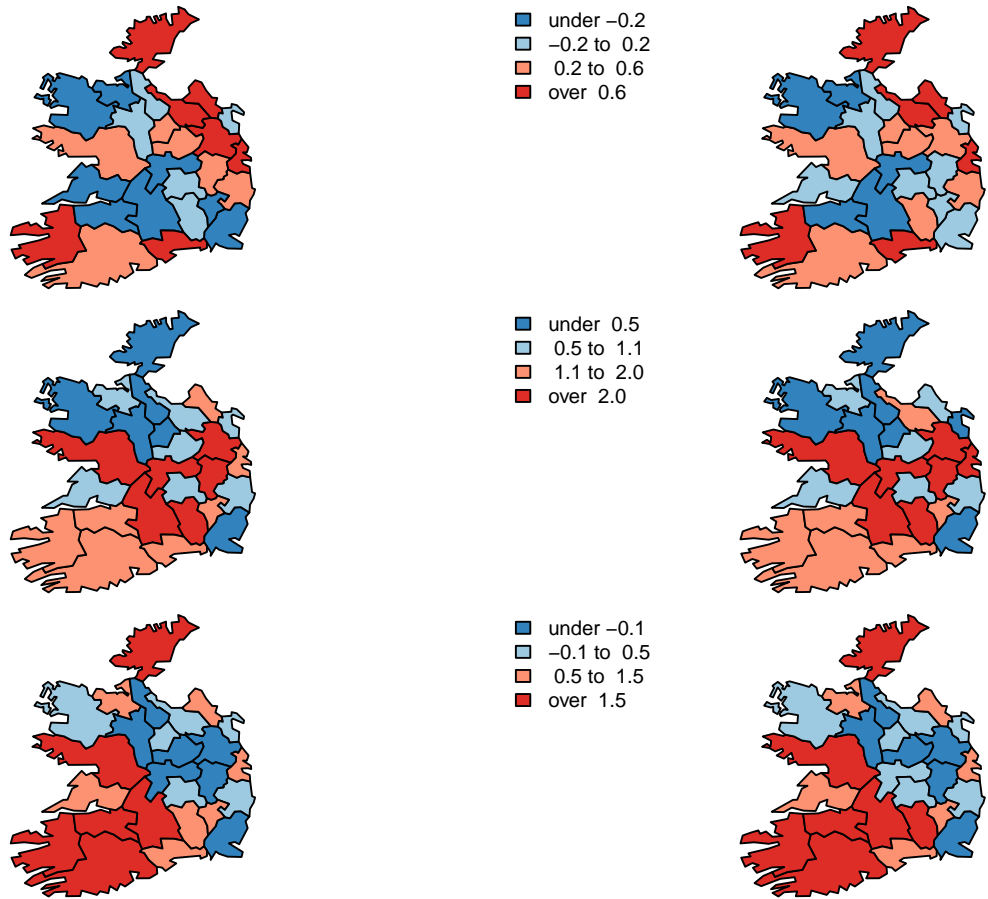


Figure 16: Maps of simulated (left) and predicted values (right) for state vector, \mathbf{x}_t , at times 18, 51 and 77 in example 5.

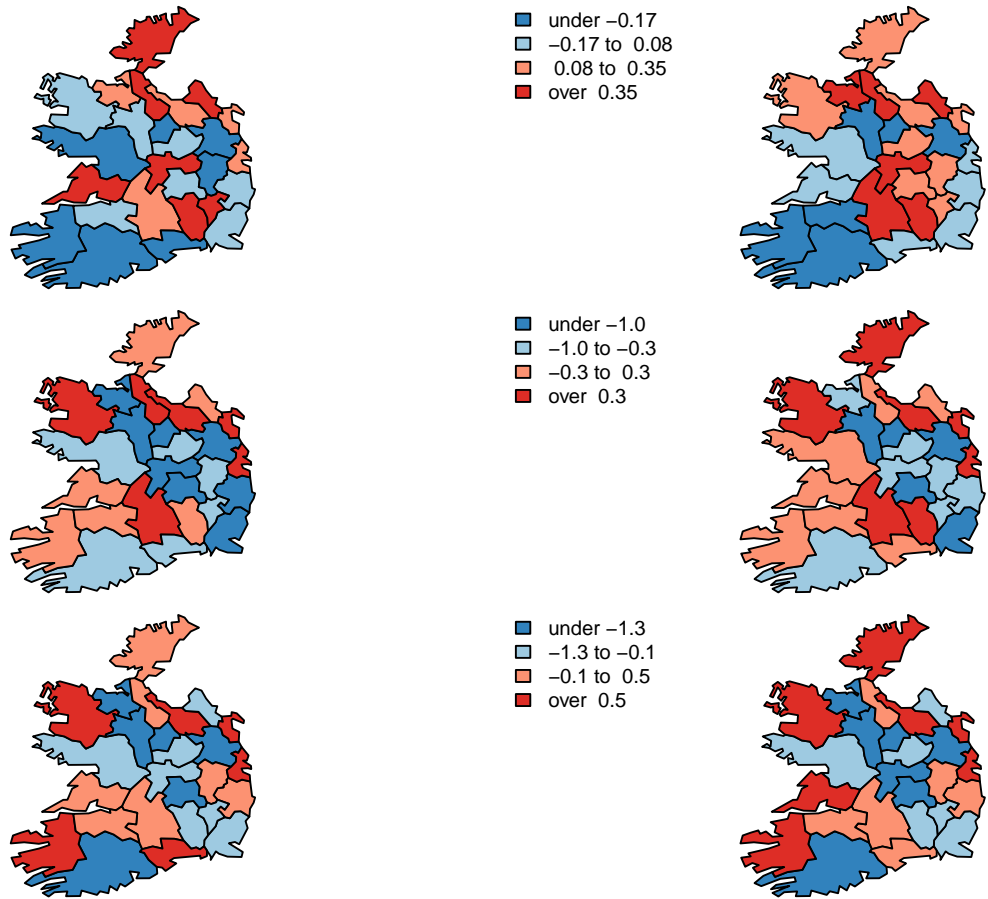


Figure 17: Maps of simulated (left) and predicted values (right) for state vector, z_t , at times 18, 51 and 77 in example 5.

Example 6: A second order spatio-temporal dynamic model

Now we will simulate data from a non-stationary second-order Gaussian model without covariates (Vivar and Ferreira, 2009). In this example both, the correlated error structure and the growth structure for the evolution of the states are present. The model is specified as:

$$\mathbf{y}_t = \mathbf{F}'_t \mathbf{x}_t + \boldsymbol{\omega}_{1t}, \quad \boldsymbol{\omega}_{1t} \sim \text{PGMRF}(\mathbf{0}_s, \mathbf{W}_1^{-1}) \quad (25)$$

$$\mathbf{x}_t = \mathbf{G}_t \mathbf{x}_{t-1} + \boldsymbol{\omega}_{23t}, \quad \boldsymbol{\omega}_{23t} \sim \text{PGMRF}(\mathbf{0}_s, \mathbf{W}_{23}^{-1}) \quad (26)$$

where,

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{pmatrix}, \quad \mathbf{F}_t = \begin{pmatrix} \mathbf{I}_s \\ \mathbf{0}_s \end{pmatrix}, \quad \mathbf{G}_t = \begin{pmatrix} \rho_1 \mathbf{I}_s & \rho_1 \mathbf{I}_s \\ \mathbf{0}_s & \rho_2 \mathbf{I}_s \end{pmatrix}, \quad \boldsymbol{\omega}_{23t} = \begin{pmatrix} \boldsymbol{\omega}_{2t} \\ \boldsymbol{\omega}_{3t} \end{pmatrix} \quad \text{and} \quad \mathbf{W}_{23}^{-1} = \begin{pmatrix} \mathbf{W}_2^{-1} & \mathbf{0}_s \\ \mathbf{0}_s & \mathbf{W}_3^{-1} \end{pmatrix},$$

and the rest of notation follows example 5.

The simulated time series here are for the 100 counties in the map of North Carolina at 30 instant times. Inference is performed for the state vectors \mathbf{x}_1 and \mathbf{x}_2 and for the scale and correlation parameters, $\tau_j, \phi_j, j = \{1, 2, 3\}$. A non-stationary process is defined assuming that $\rho_1 = \rho_2 = 1$. Appendix A shows some parts of the R code used to fit this model with INLA. Further details can be found in the R script accompanying this paper.

As in example 5, for the implementation of this model in INLA it is necessary to specify the precision matrices \mathbf{W}_j , associated to the error vectors $\boldsymbol{\omega}_{jt}$, through a generic model using option `model='generic1'` in the formula to be called by the INLA library.

Note that precision parameter for observations in this and the above example is declared as fixed in the call to fit the model (first element of the list in `control.data`), because it is also specified through a generic model in the first line of the formula.

Predicted values here also closely agree with the simulated series of observations and states in all counties. Figures 18 and 19 exemplify this for the 20th county and its neighbors. Maps comparing simulated and predicted observations and states for all areas at some instant times are also shown in Figures 15 and 16. This model, which is the most complex one considered in this work, took about eight minutes to run on the simple machine specified at the beginning of this section.

In order to check how this model responds to changes in initial values of the hyperparameters as well as to different prior hyperparameter specification, we performed a small sensitivity analysis. Firstly, the model was fitted with informative, vague and default log-gamma priors for $\log(\tau_i), i = \{1, 2, 3\}$, as specified at the beginning of this section. In this case, priors for the correlation parameters, ϕ_i , and initial values for all the hyperparameters were specified at their default settings. Furthermore, we consider default prior specification for all the hyperparameters and the model was fitted under two situations: (i) varying the initial values for $\log(\tau_i)$ in the set $\{\log(10), \log(20), \dots, \log(90), \log(100)\}$ keeping the default initial values for the correlation parameters, and (ii) varying the initial values for ϕ_i in the set $\{0.1, 0.2, \dots, 0.8, 0.9\}$ keeping the default initial values for the log-precision parameters. At each time the three log-precision parameters or the three correlation parameters of the model were initialized with the same value from the sets above. The results are graphically presented in Figures 20 and 21. In general terms, the model was not sensitive to prior specification of the log-precisions, $\log(\tau_i)$, with marginal posterior densities overlapping to a large extent. In contrast, the choice of the initial

values for $\log(\tau_i)$ had a direct impact on the approximation of the marginal posterior densities for these hyperparameters, being the best results found for initial values in the range $\log(40)$ to $\log(60)$, which includes the default initial value set by the INLA library. Posterior results for the correlation parameters were less sensitive to the choice of their initial values.

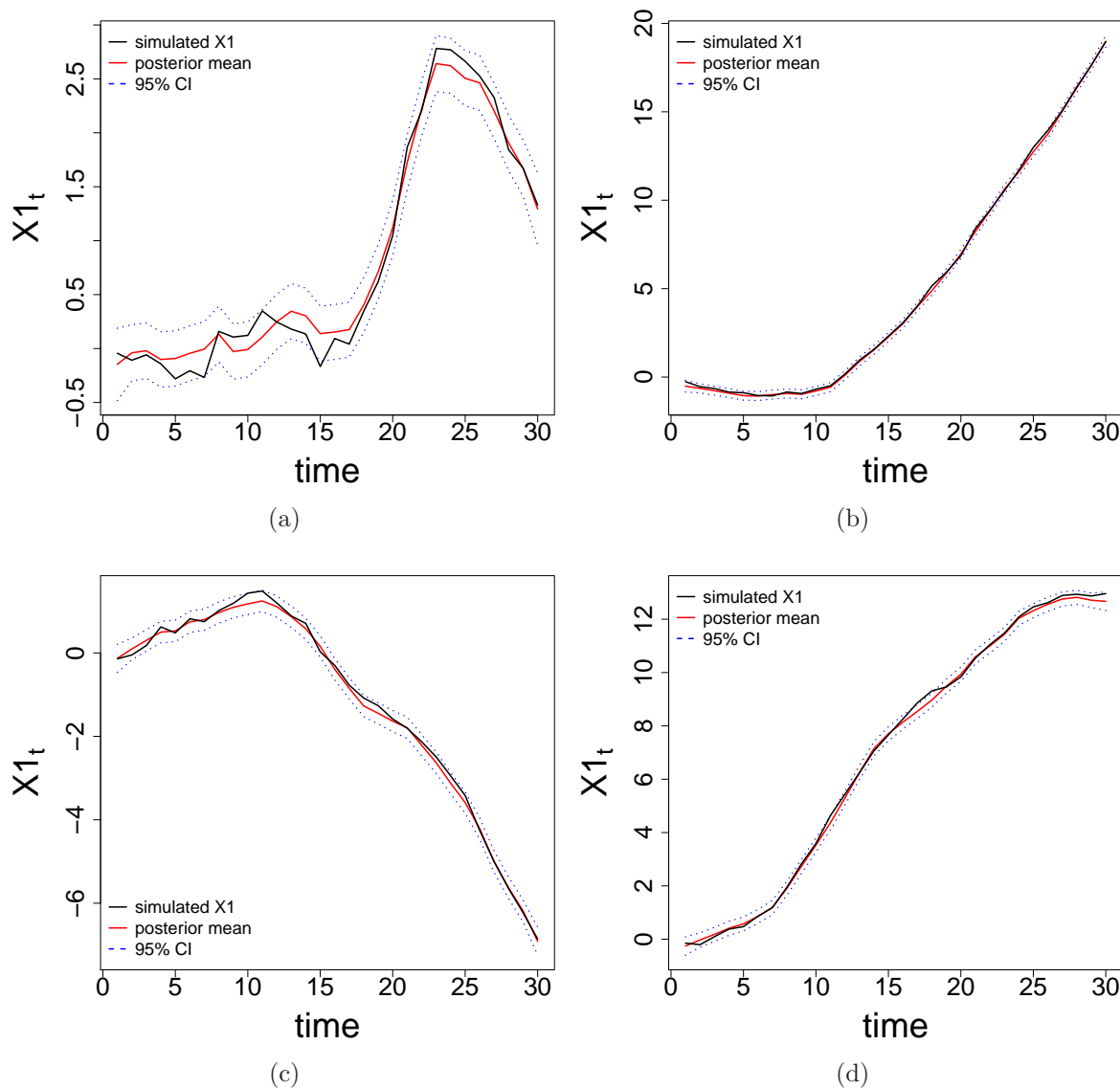


Figure 18: Simulated and predicted values (posterior mean and 95% credibility interval) for the states $x_{1,t}$ in the 20th area (a) and in its neighbors (b-d) in the second order spatio-temporal dynamic model.

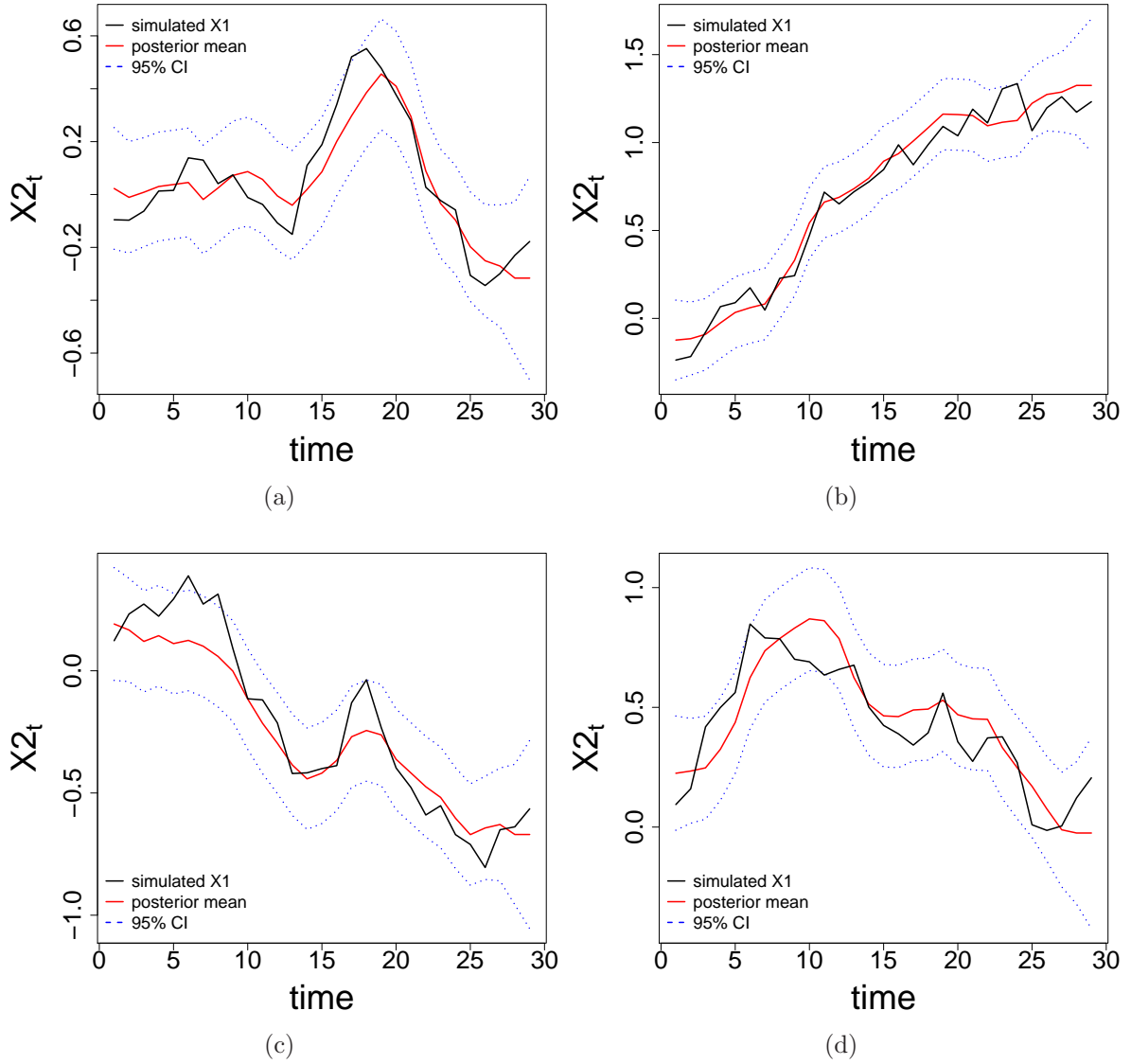


Figure 19: Simulated and predicted values (posterior mean and 95% credibility interval) for the states $x_{2,t}$ in the 20th area (a) and in its neighbors (b-d) in the second order spatio-temporal dynamic model.

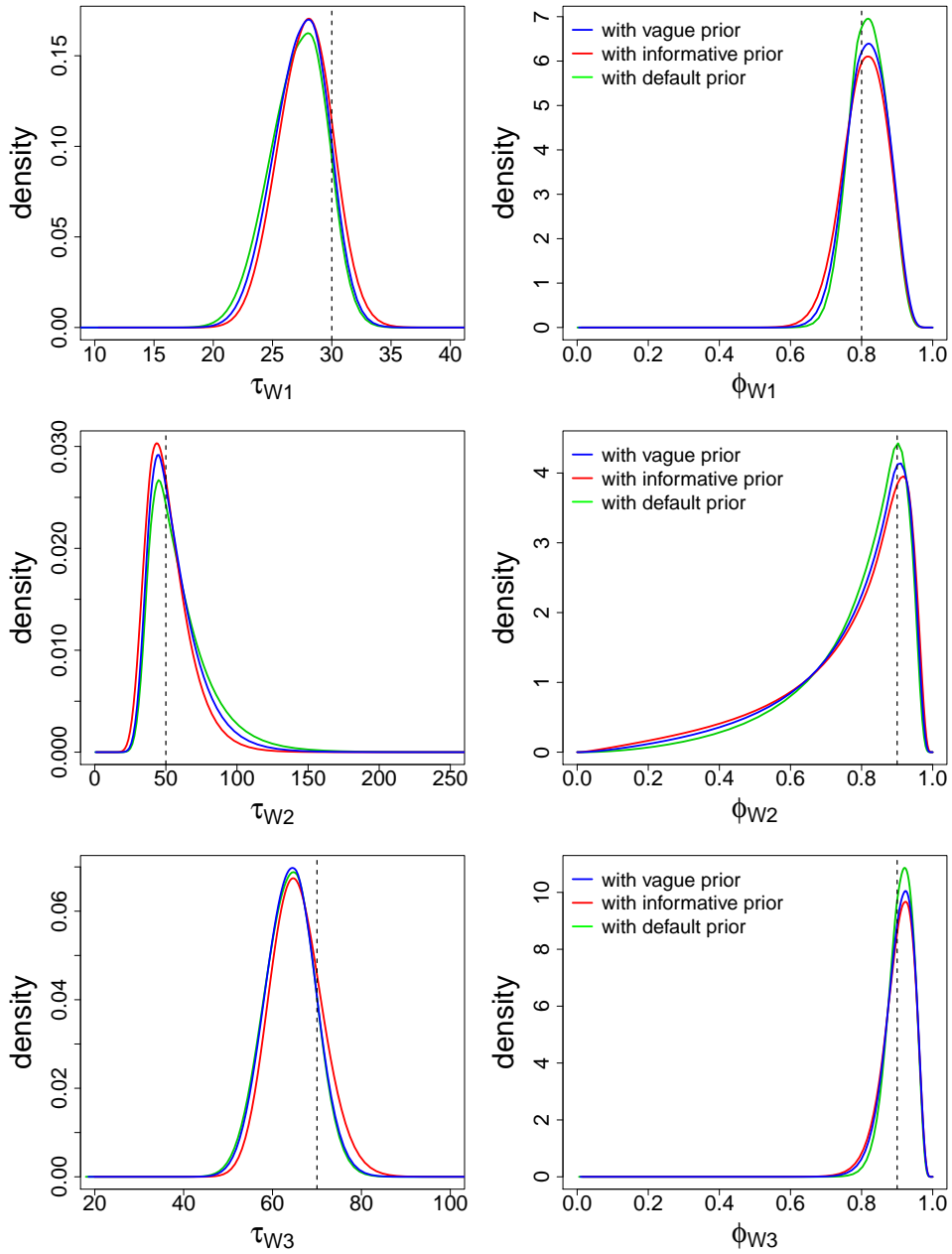


Figure 20: Marginal posterior densities for precision (left) and correlation (right) parameters in the second order spatio-temporal dynamic model using default, vague and informative priors for the precision parameters (see text for details). Vertical dashed lines indicate true simulated values.

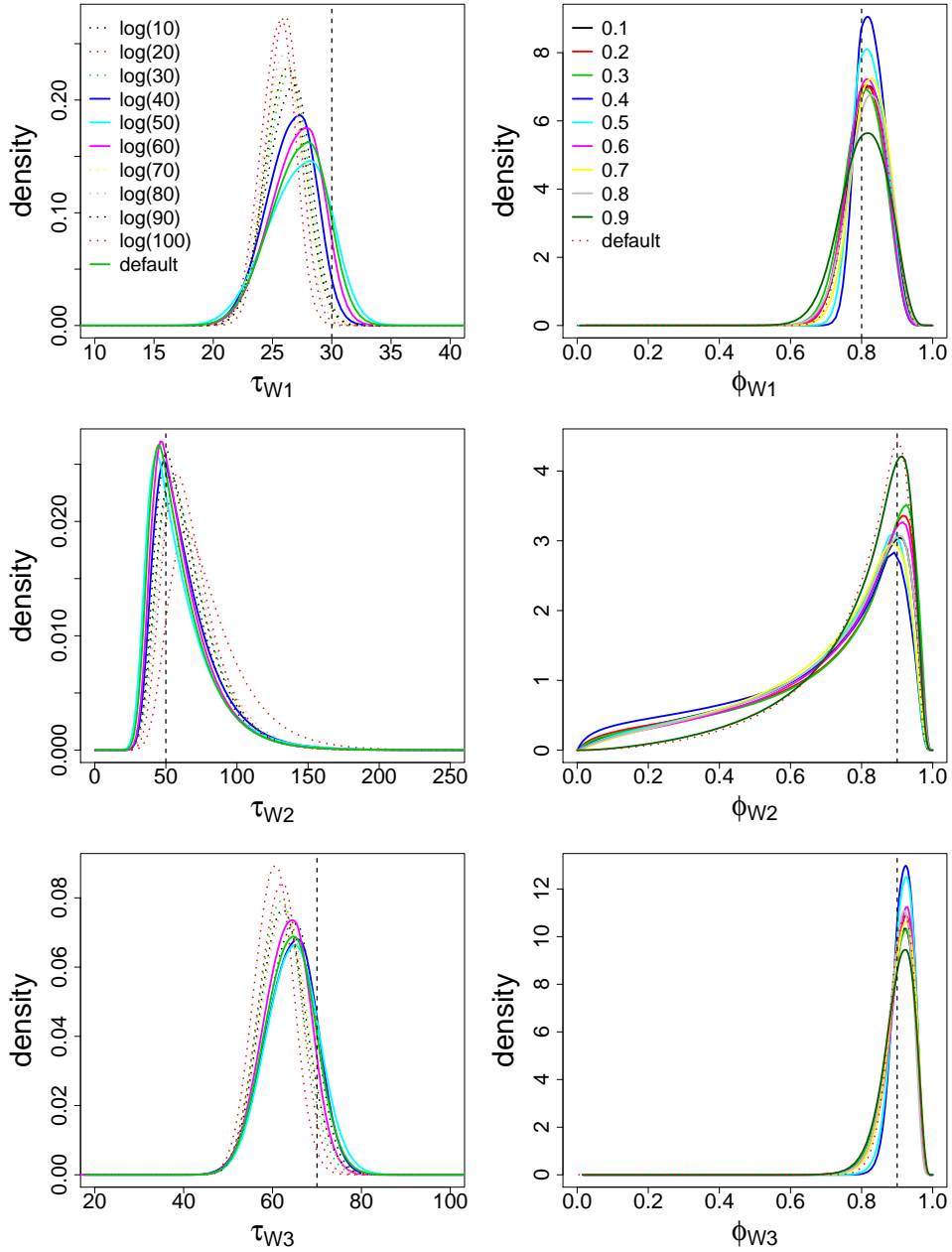


Figure 21: Marginal posterior densities for precision, τ_i , (left) and correlation, ϕ_i , (right) parameters initialized with different values for $\log(\tau)$ and ϕ in the second order spatio-temporal dynamic model. Priors for all the hyperparameters were specified at their default settings. Vertical dashed lines indicate true simulated values.

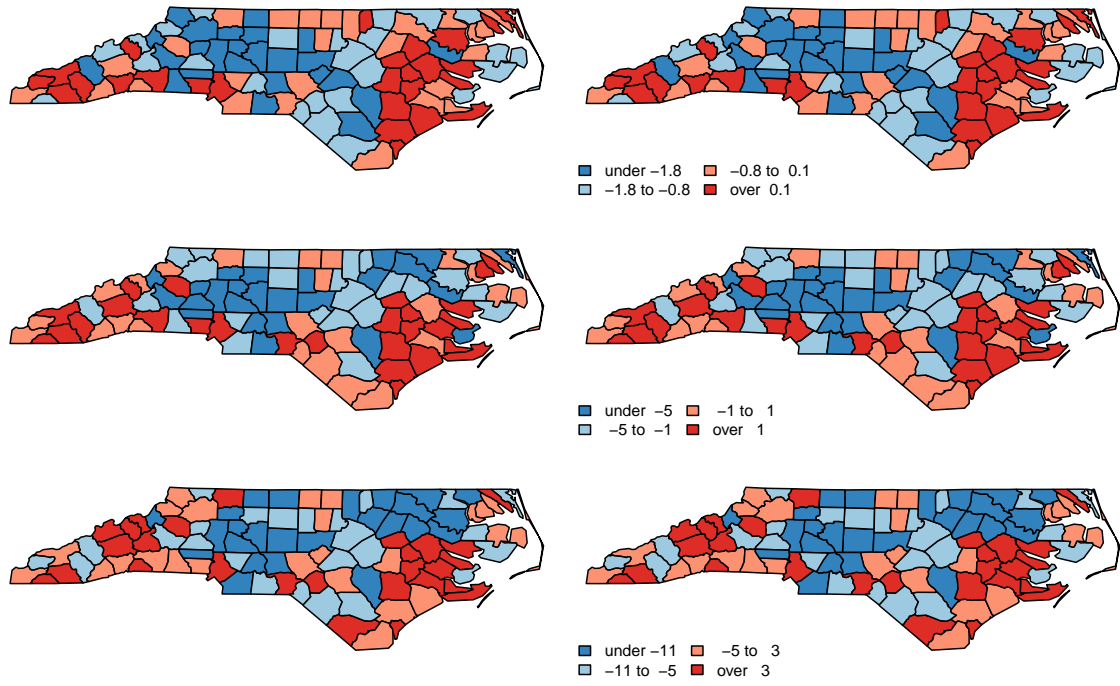


Figure 22: Maps of simulated (left) and predicted values (right) for observations, y_t , at times 7, 15 and 29 in example 6.

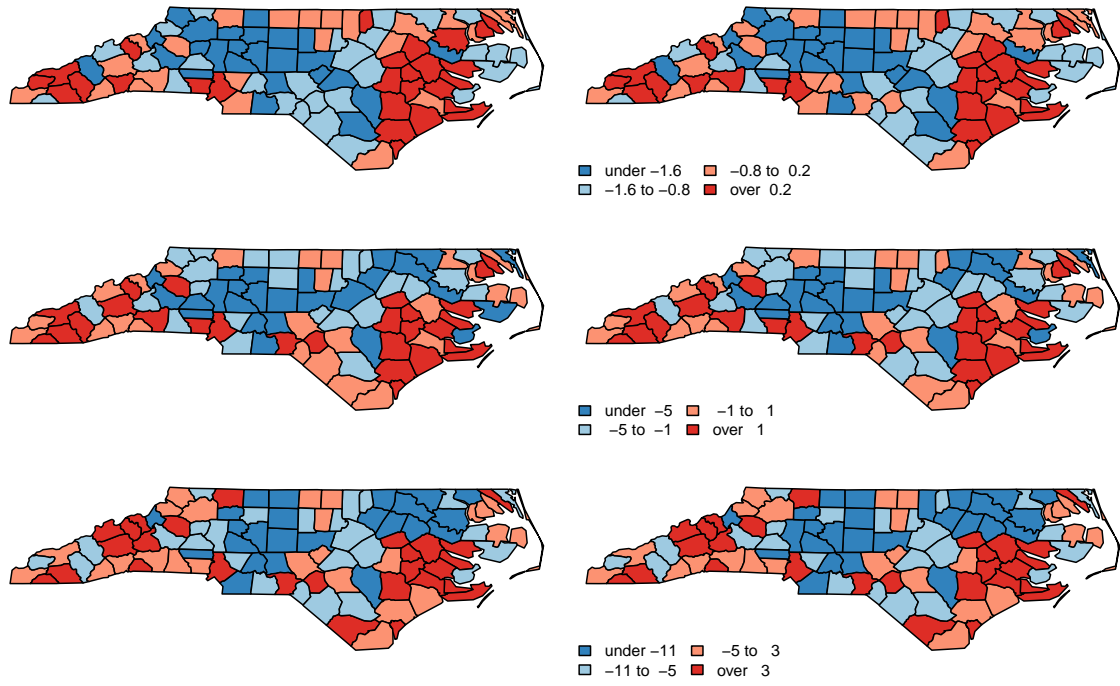


Figure 23: Maps of simulated (left) and predicted values (right) for X_1 state vector at times 7, 15 and 29 in example 6.

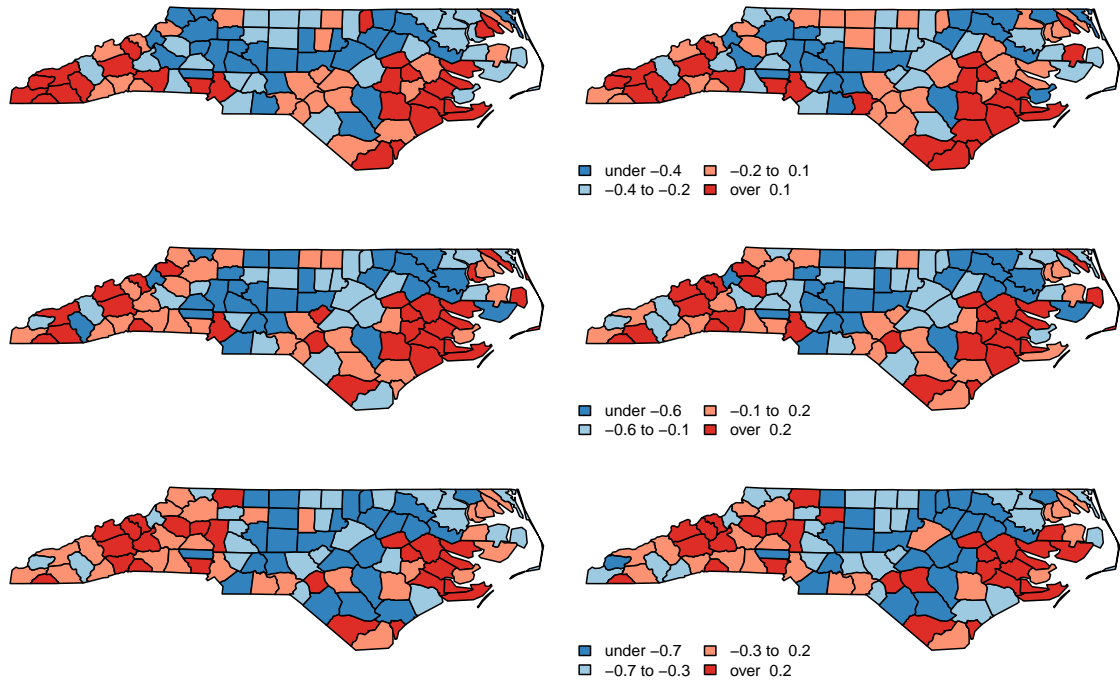


Figure 24: Maps of simulated (left) and predicted values (right) for X_2 state vector at times 7, 15 and 29.

5 Case studies

In this section we use some well known examples from the literature to illustrate how a relevant data analysis with DLMS can be performed using the INLA approach. The examples include dynamic models with Gaussian and Poisson observation densities, temporal trend and seasonality components as well as external covariates. When possible, comparison with results from the literature using other inference methods is provided. In order to facilitate understanding, relevant parts of the R code used in model fitting, for some representative examples, were included in an appendix at the end of this paper. The full code and data sets needed to fit all the examples in this section is also provided from the INLA web page (see the Supplementary material).

Example 7: UK Gas consumption

The first worked example to be analyzed corresponds to the quarterly UK gas consumption from 1960 to 1986, in millions of therms. Details on this dataset can be found in Durbin and Koopman (2001, p. 233). Following Dethlefsen and Lundbye-Christensen (2006) here we use the (base 10) logarithm of the UK gas consumption as response, which is assumed to be normally distributed and we fit a model with a first order polynomial trend (T_t) with time-varying coefficients and an unstructured seasonal component (S_t), also varying over time. Therefore, the observational and system equations are given by

$$y_t = \log_{10}(UKgas)_t = T_t + S_t + \nu_t, \quad \nu_t \sim N(0, V), \quad t = 1, \dots, n \quad (27)$$

$$T_t = T_{t-1} + \beta_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \quad t = 2, \dots, n \quad (28)$$

$$\beta_t = \beta_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), \quad t = 2, \dots, n \quad (29)$$

$$S_t = -(S_{t-1} + S_{t-2} + S_{t-3}) + \omega_{3t}, \quad \omega_{3t} \sim N(0, W_3), \quad t = 4, \dots, n \quad (30)$$

Approximate Bayesian inference in this case was performed using a mixed approach. Firstly, we utilized an augmented structure merging the observational equation (27) with the polynomial trend (28). The slope and seasonal terms in equations (29) and (30), were modeled with standard model options from the INLA library. Specifically, for β_t we used a first order random walk model, while the INLA's seasonal model was used for S_t term.

We also carried out an MCMC fit for the same data set using using a Gibbs sampling algorithm available in the `d1m` package. In order to obtain suitable (precise) estimates of the parameters of interest for a comparison between MCMC and INLA approaches, 1010000 posterior samples were drawn from a unique MCMC chain by using a thinning of 20 and a burn-in of 10000. Hence, 50000 effective samples were left for an estimation of the posterior quantities. Priors for the precision parameters of the observation and evolution errors followed those suggested in Petris (2010) for the same data set. The `CODA` package (Plummer et al., 2006) was used to diagnose convergence through the Geweke and Raftery–Lewis diagnostic tests. Figure 25 compares the marginal posterior densities of the hyperparameters obtained with the MCMC and INLA approaches. Forecasts for the next 12 quarterly periods of the UK gas consumption time series were also performed with the two methods (see Figure 26). However, we must be aware that, unlike INLA, forecasting with the `d1m` package is not fully Bayesian. Instead it is recursively obtained considering the ergodic means of the variance parameters (computed from the output of the MCMC) as the known variances at the last observed time, n , and then the predictive distributions are sequentially updated k steps ahead.

According to Figures 25 and 26, the performance obtained with the two approaches was, in general, very close both in terms of parameter estimation and forecasting. Differences, however,

arise in computer time spent by the two methods. The `d1m` package used 43629.75 seconds to run the MCMC algorithm in R against 3.43 seconds spent by the INLA library (almost 13000 times faster). At first glance this comparison of computational time would not seem totally fair since INLA is heavily optimized whereas the `d1m` package is entirely R based and hence slower than an MCMC algorithm written in a low level language. However, it makes sense if we think in terms of users without high-level computing skills, given that both approaches run easily in R using a familiar formula structure, avoiding the need of writing any code.

An extended Kalman filtering approach was also implemented with the `sspir` package following Dethlefsen and Lundbye-Christensen (2006). This approach, although very fast (0.23 seconds), only provides punctual estimates of the parameters of interest. The estimated variances are also externally obtained from a maximum likelihood algorithm provided by the R function `StructTS` (Ripley, 2002) and then “plugged” into the Kalman filter. No capabilities for forecasting are available with this tool. A figure showing the decomposition of the time series in trend, slope and seasonal components obtained with the three approaches (INLA, MCMC and extended Kalman filter) is available at the online supplementary material. The amplitude of the seasonal term remains virtually constant from 1960–1971, then it increases during the period 1971–1979 and finally it stabilizes again. It is important to highlight that in spite of the similarities in the results from the three methods, only the direct approach offered by INLA provides fully Bayesian inference and forecasting in an efficient (fast) way and is, in this sense, superior to the other approaches.

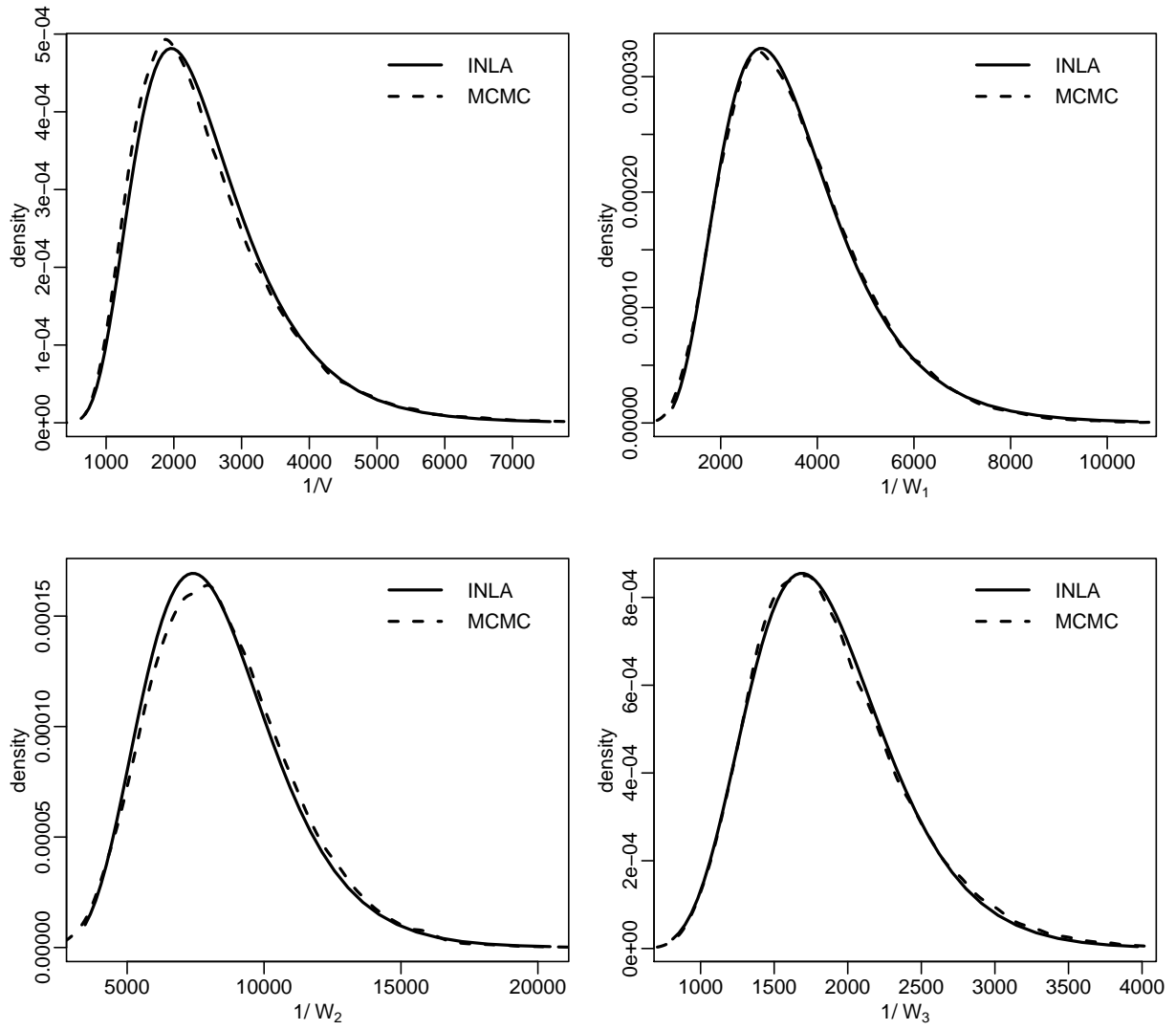


Figure 25: Marginal posterior densities for the precision parameters obtained with INLA and MCMC in the UK gas consumption example.

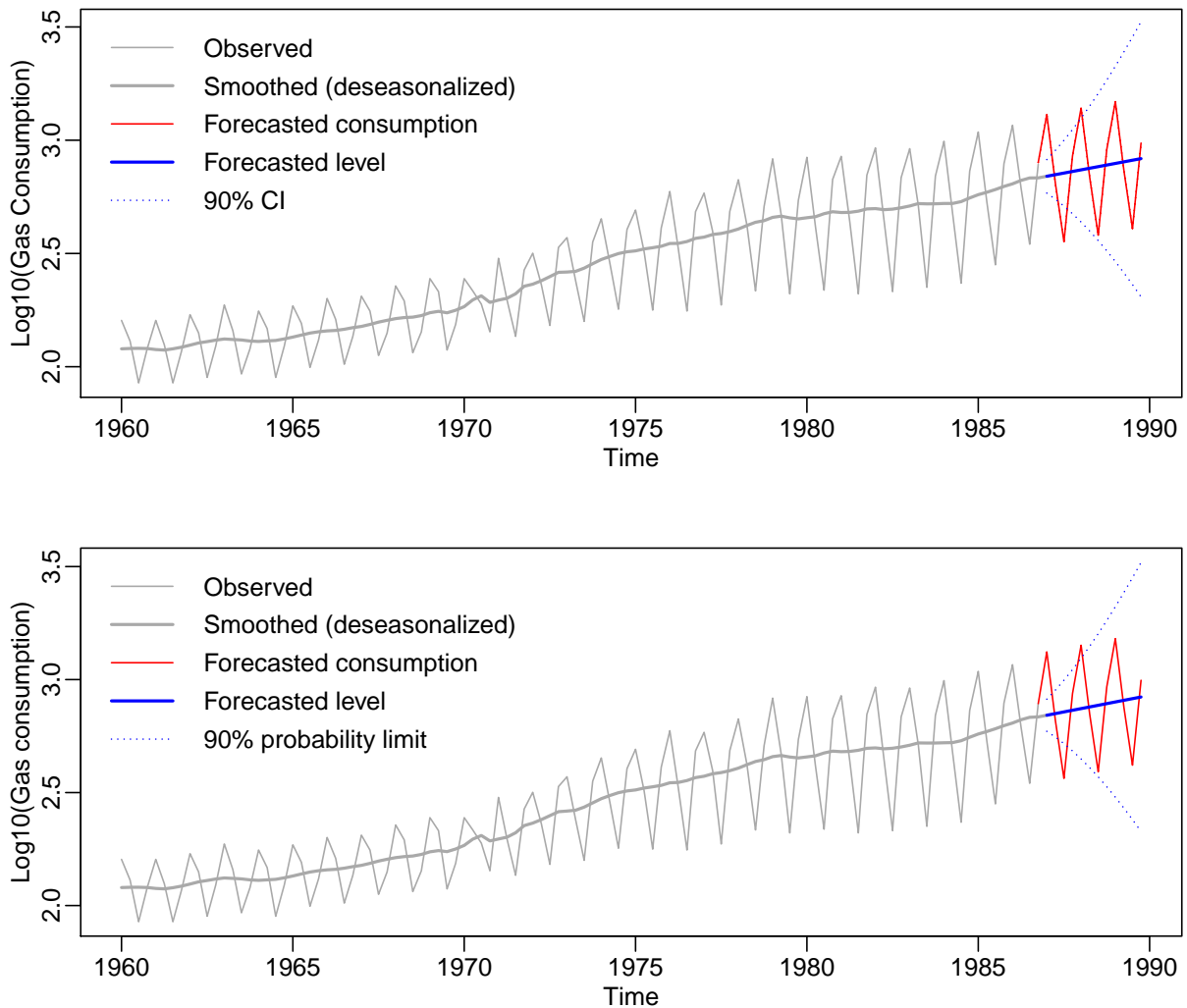


Figure 26: UK gas consumption forecasts for the next 12 quarterly periods obtained with INLA (upper frame) and with the `dlm` package (lower frame).

Example 8: Van drivers

This is a classical example of a generalized dynamic linear model. Here the response y_t corresponds to the monthly numbers of light goods van drivers killed in road accidents in Great Britain, from January 1969 to December 1984 (192 observations). A seat belt law (intervention) was introduced on January 31st, 1983. The interest is in quantifying the effect of the seat belt legislation law on the number of deaths. For further information about the data set see Harvey and Durbin (1986) and Durbin and Koopman (2000).

This dataset has been previously analysed with INLA in a time series setting, assuming that the squared root of the counts, y_t , follows a Gaussian distribution. For details of this implementation see Martino and Rue (2010). Following Dethlefsen and Lundbye-Christensen (2006), here we use a generalized dynamic linear model for Poisson data with a 13-dimensional latent process, consisting of an intervention parameter, seat belt, changing value from zero to one in February 1983, a constant monthly seasonal term (S_t), and a temporal trend (T_t), modeled as a random walk. The observational and system equations for this model are as

follows

$$y_t \sim \text{Poisson}(\mu_t)$$

$$\log(\mu_t) = \lambda_t = T_t + \alpha * \text{seatbelt}_t + S_t, \quad t = 1, \dots, n \quad (31)$$

$$T_t = T_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W), \quad t = 2, \dots, n \quad (32)$$

$$S_t = -(S_{t-1} + \dots + S_{t-11}), \quad t = 12, \dots, n \quad (33)$$

The trend and seasonal terms in linear predictor (31) can be directly modeled using existing first order random walk and seasonal model options from the INLA library, as shown in the following code:

```
i <- j <- 1:n          # indices for T_t and S_t
formula <- y ~ belt + f(i, model="rw1", param=c(1,0.0005), constr=F) +
           f(j, model="seasonal", season.length=12) -1
r <- inla(formula, data = data.frame(belt, i, j, y), family = "poisson")
```

The estimated trend and the effect of the seat belt intervention, as well as its comparison with results obtained with the `sspir` package (Dethlefsen and Lundbye-Christensen, 2006) can be displayed in Figure 27.

The posterior mean for α parameter in Eq. (31), which represents the effect of the seat belt law on the number of deaths, was -0.284 with a posterior standard deviation of 0.152 ; this corresponds to a reduction in the number of deaths of 24.63% . This result agrees with the corresponding value reported by Durbin and Koopman (2000) for this parameter, which was -0.280 with posterior standard deviation of 0.126 . Similarly Dethlefsen and Lundbye-Christensen (2006), using an extended Kalman filter, obtained $\alpha = -0.285$.

Example 9: Mumps

The response y_t in this example is the monthly number of registered cases of mumps in New York City from January 1928 to June 1972. This data set was previously studied by Hipel and McLeod (1994). According to Dethlefsen and Lundbye-Christensen (2006), the incidence of mumps is known to show seasonal behavior and a variation in trend during the study period. For this data set we used a generalized dynamic linear model for Poisson observations, where the mumps incidence was modeled with a first order polynomial trend (T_t) with time-varying coefficients and a time-varying harmonic seasonal component (H_t) as suggested in Dethlefsen and Lundbye-Christensen (2006). The observational and system equations for this model are as follows

$$y_t \sim \text{Poisson}(\mu_t)$$

$$\log(\mu_t) = \lambda_t = T_t + H_t, \quad t = 1, \dots, n$$

$$T_t = T_{t-1} + \beta_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \quad t = 2, \dots, n \quad (34)$$

$$\beta_t = \beta_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), \quad t = 2, \dots, n \quad (35)$$

$$H_t = a_t \cos\left(\frac{2\pi}{12}t\right) + b_t \sin\left(\frac{2\pi}{12}t\right), \quad t = 1, \dots, n \quad (36)$$

$$a_t = a_{t-1} + \omega_{3t}, \quad \omega_{3t} \sim N(0, W_3), \quad t = 2, \dots, n \quad (37)$$

$$b_t = b_{t-1} + \omega_{4t}, \quad \omega_{4t} \sim N(0, W_4), \quad t = 2, \dots, n \quad (38)$$



Figure 27: Number of vandriviers killed and estimated trend + intervention. Solid and dotted lines in blue correspond to the posterior mean and 95% credibility intervals, respectively, obtained with the INLA library. The line in red corresponds to the estimated trend + intervention with the `sspir` package.

As in example 7, a mixed approach is suitable here, where the polynomial trend in system equation (34) is merged with the observational equation, yielding an augmented model with two different likelihoods (Poisson for the n actual observations and Gaussian for the $n - 1$ “pseudo” observations). The slope and seasonal terms in equations (35) to (38) follow a random walk evolution and were just modeled with a first order random walk model option from the INLA library. The relevant parts of the code to formulate and fit this model with INLA are shown next.

```
# building the augmented model
# -----
m <- n-1
Y <- matrix(NA, n+m, 2)
Y[1:n, 1] <- mumps
Y[1:m + n, 2] <- 0

## indices for the INLA library
# -----
i <- c(1:n, 2:n) # indices for T_t
j <- c(rep(NA,n), 1:m) # indices for T_{t-1}
```

```

weight1 <- c(rep(NA,n), rep(-1,m))      # weights for T_{t-1}
l       <- c(rep(NA,n), 1:m)           # indices for \beta_{t-1}
weight2 <- c(rep(NA,n), rep(-1,m))    # weights for \beta_{t-1}
w1      <- c(rep(NA,n), 2:n)          # indices for w_{1,t}
q       <- c(1:n, rep(NA,m))           # indices for a_t
cosine  <- c(cosw,rep(NA,m))           # weights for a_t
rr      <- c(1:n, rep(NA,m))           # indices for b_t
sine    <- c(sinw,rep(NA,m))           # weights for b_t

# formulating the model
# -----
formula = Y ~ f(q, cosine, model="rw1",param=c(1,0.01),initial=4, constr=F) +
              f(rr, sine, model="rw1",param=c(1,0.01),initial=4, constr=F) +
              f(l, weight2, model="rw1",param=c(1,0.2),initial=4, constr=F) +
              f(i, model="iid", initial=-10, fixed=TRUE) +
              f(j, weight1, copy="i") + f(w1, model="iid") -1

# call to fit the model
# -----
r <- inla(formula, data = data.frame(cosine,sine,i,j,weight1,l,weight2,q,rr,w1),
          family = c("poisson","gaussian"),
          control.data = list(list(),list(initial=10, fixed=TRUE)))

```

It is important to note here that in the formulation of this model, the seasonal terms, following an RW1 process, must be declared first in the formula to be passed to the `inla` function, followed by the terms in the equations that forms the augmented model. Otherwise the INLA library can made a wrong interpretation of the indices of these terms, which can lead to misleading results.

The comparison of results obtained with the INLA library and with the `sspir` package (Dethlefsen and Lundbye-Christensen, 2006) for the variation of mumps incidence are shown in Figure 28. They were very similar for the two approaches. According to Figure 28, seasonal pattern of incidence changes slowly, as can be seen in the decreasing behavior of the peak-to-trough ratio and peak location series. The location of the incidence's peak also changes from middle/late April in the beginning of the study period to late May in the last four years.

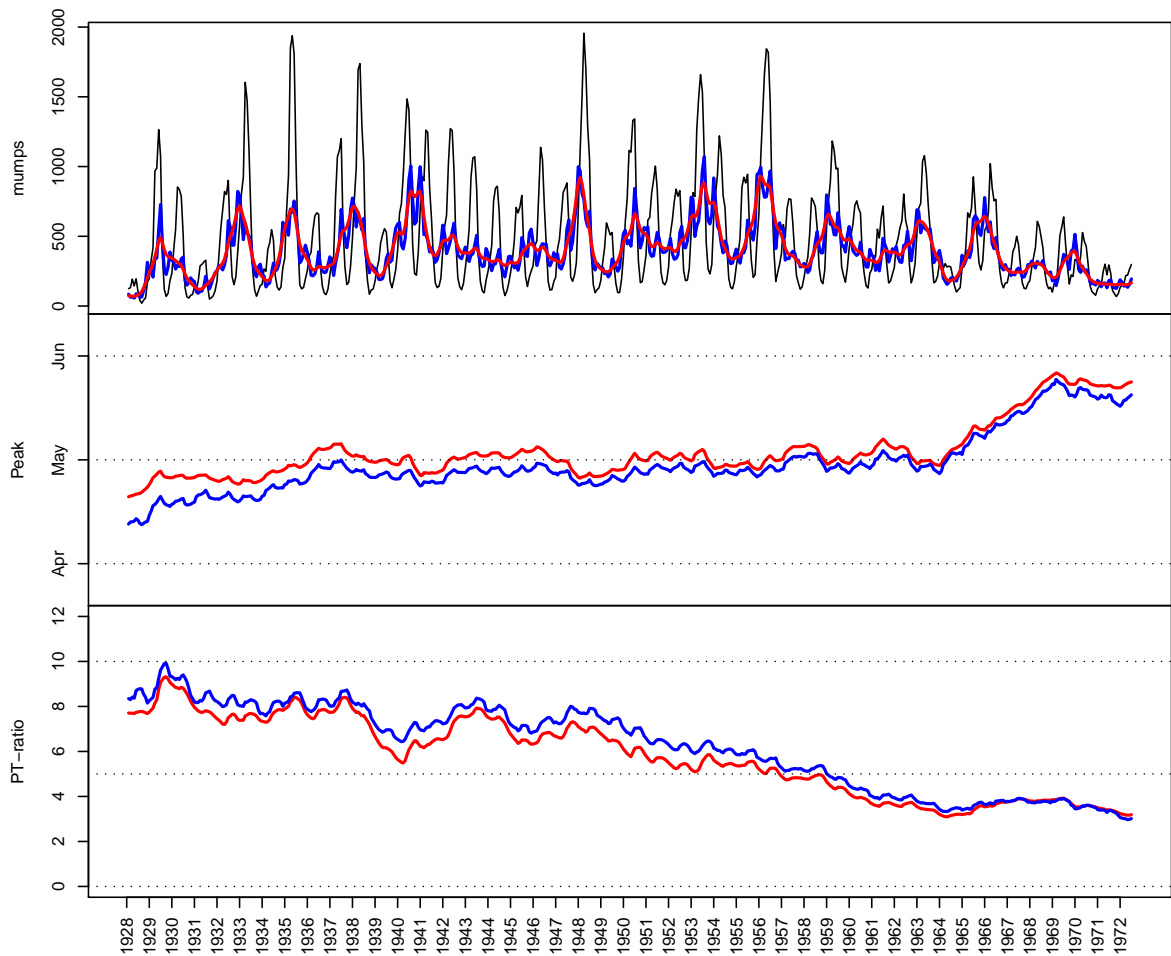


Figure 28: Comparison between INLA (blue lines) and `sspir` (red lines) results for the variation in the incidence of mumps in New York city from 1927 to 1972. The upper frame shows the observed number of cases jointly with the de-seasonalized trend. The location of the peak of the seasonal pattern is shown in the middle frame, while the lower frame is for the variation in the peak-to-trough ratio over the study period.

Example 10: Market share

In our last worked example we analyze percent market share for a consumer product. This example was fully analyzed in Pole et al. (1994, Chapter 5). The model for market share utilize weekly available information for 1990 and 1991 on product price and measures of promotional activity. The objectives are to determine a model with good predictive power and to assess the importance of suggested explanatory variables. The response y_t is assumed to be Gaussian distributed. For this data set we use a dynamic regression model with three covariates, `price`, `prom` and `cprom`, where `price` is the measured price relative to a number competitor's average prices; `prom` and `cprom` are producer and competitor promotion indices. Following Pole et al. (1994), the level is modeled as fixed and the regression coefficients have a random walk evolution. One point identified as outlier on week 34 of 1990 and excluded from the analysis in Pole et al. (1994), was also excluded in our analysis for comparison purposes. The observational and

system equations for this model are as follows

$$\begin{aligned}
 y_t &= \alpha_t + \beta_{1t}\text{price}_t + \beta_{2t}\text{prom}_t + \beta_{3t}\text{cprom}_t + \nu_t, & \nu_t &\sim N(0, V), & t &= 1, \dots, n \\
 \beta_{1t} &= \beta_{1,t-1} + \omega_{1t}, & \omega_{1t} &\sim N(0, W_1), & t &= 2, \dots, n \quad (39) \\
 \beta_{2t} &= \beta_{2,t-1} + \omega_{2t}, & \omega_{2t} &\sim N(0, W_2), & t &= 2, \dots, n \quad (40) \\
 \beta_{3t} &= \beta_{3,t-1} + \omega_{3t}, & \omega_{3t} &\sim N(0, W_3), & t &= 2, \dots, n \quad (41)
 \end{aligned}$$

The model was formulated in INLA considering a simple random walk evolution form for each regression coefficient. Figure 29 shows the predicted market share values and the estimated level obtained with the INLA library and that reported in Pole et al. (1994) using the Bats software.

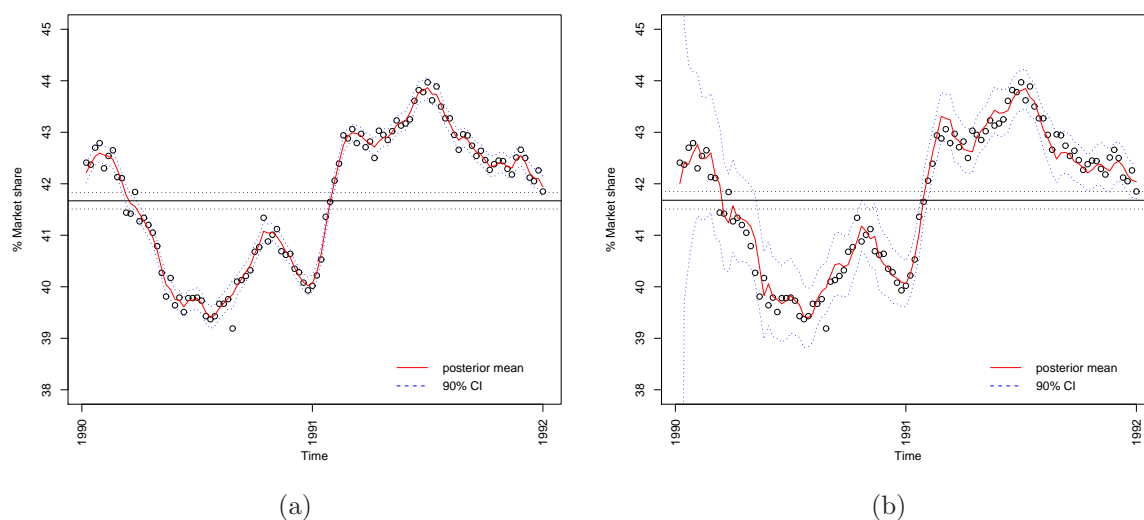


Figure 29: Observed and predicted values (posterior mean and 90% credibility interval) for the market share example using the INLA library (a) and the Bats software (b). Horizontal black lines in both plots indicate the estimated level with its 90% credibility interval.

Results of the week-by-week forecasts for the first five weeks of 1992 under four different scenarios, as considered in Pole et al. (1994), are shown in Table 2, using the INLA library and the Bats software. The four scenarios were:

1. `prom` and `cprom` indices set to 0,
2. `prom` set to its first five values of 1990 and `cprom` set to zero,
3. `prom` set to zero and `cprom` set to its first five values of 1990,
4. `prom` and `cprom` set to their first five values of 1990.

Relative price was maintained fixed in all cases at 0.206, the final value for 1991.

Table 2: Week-by-week forecasts for percent market share for the first five weeks of 1992 with INLA and with Bats software for four different promotion scenarios (see text for details). 0.05 and 0.95 quantiles are indicated by $0.05q$ and $0.95q$, respectively.

Week	INLA			Bats		
	mean	0.05q	0.95q	mean	0.05q	0.95q
	Scenario 1			Scenario 1		
1992/1	41.40	41.19	41.61	41.40	41.04	41.76
1992/2	41.40	41.18	41.61	41.40	41.04	41.76
1992/3	41.40	41.18	41.61	41.40	41.03	41.77
1992/4	41.40	41.18	41.62	41.40	41.03	41.77
1992/5	41.40	41.18	41.62	41.40	41.03	41.77
	Scenario 2			Scenario 2		
1992/1	41.36	41.14	41.59	41.36	41.00	41.73
1992/2	41.32	41.07	41.56	41.31	40.93	41.69
1992/3	41.27	41.00	41.54	41.25	40.86	41.65
1992/4	41.25	40.97	41.53	41.23	40.83	41.63
1992/5	41.22	40.92	41.52	41.19	40.77	41.61
	Scenario 3			Scenario 3		
1992/1	41.28	40.91	41.64	41.23	40.84	41.62
1992/2	41.27	40.90	41.66	41.22	40.82	41.62
1992/3	41.27	40.88	41.67	41.22	40.82	41.62
1992/4	41.27	40.87	41.68	41.22	40.81	41.63
1992/5	41.26	40.83	41.71	41.20	40.79	41.62
	Scenario 4			Scenario 4		
1992/1	41.24	40.86	41.63	41.19	40.79	41.59
1992/2	41.19	40.77	41.62	41.13	40.71	41.55
1992/3	41.14	40.69	41.62	41.07	40.63	41.51
1992/4	41.12	40.64	41.62	41.05	40.60	41.50
1992/5	41.08	40.55	41.63	41.00	40.53	41.47

Forecasting trend was similar under the two approaches for all scenarios considered.

6 Concluding remarks

In this paper we propose a computational framework to perform direct full approximate Bayesian inference in linear and generalized dynamic linear models based on the INLA approach. We illustrate our proposal through a series of simulated and real-life examples ranging from simple univariate models to realistically complex spatio-temporal dynamic models. Our approach allows an easy specification of complex dynamic models in R using a formula language, as is routinely done with the most common linear and generalized linear models. Additional examples

can be found at the technical report by ?. The proposed framework outperforms computational tools currently available in the literature of dynamic models in some important respects:

- Unknown precision parameters and its credibility intervals are straightforwardly estimated with INLA jointly with the state parameters, even for models with non-Gaussian observations, unlike other approaches in the literature which do not estimate the unknown variance parameters automatically or are restricted to the Gaussian case. The `sspir` package (Dethlefsen and Lundbye-Christensen, 2006), for example, requires the combination of numerical maximization algorithms with the output of the iterated extended Kalman smoother, while the `Bats` software (Pole et al., 1994) uses a discount factor approach to model unknown variances. The `SsfPack` package (Koopman et al., 1999) provides punctual estimates of the hyperparameters of state space models, but it requires further Monte Carlo simulation in order to get the confidence intervals through some bootstrap procedure as proposed, for example, in Franco et al. (2008). The `d1m` package (Petris, 2010) allows the Bayesian estimation of unknown hyperparameters, but it is restricted to Gaussian dynamic linear models and even within this class it does not consider important cases such as multiple dynamic regression. Computational cost with this package is also higher due to the Gibbs sampling algorithm on which it is based.
- Our approach is able to deal with complex spatio-temporal observations in an easy way, as shown in examples 3 and 4. To the best of our knowledge there are no other computational tools currently available in the literature to deal with this kind of data in a framework of dynamic models.
- The proposed framework offers large savings in computational time if compared to expensive MCMC methods as shown in example 5 and in the toy example. Although not compared in this article, it is expected that these gains will be even more evident in the case of spatial and spatio-temporal dynamic models as those in examples 3 and 4. In these cases, which are the most complex ones considered in this work, our approach spent just a few minutes to run and most of the examples were fitted in just a few seconds.

The direct approximation of the posteriors for the states and hyperparameters performed by the INLA approach, following the framework proposed in this paper, allows a fast yet easier way to inference, even for complex state-space models. It is important to recall here that when observations follow a Gaussian distribution, as most of the examples in this paper, the method is actually exact up to integration error. This is not the case with current sequential (Kalman filter-based) approaches to inference in state-space models, or even when MCMC algorithms are used. Our approach improves current recursive inference methods where estimation is performed with filtering and smoothing steps, based on the temporal structure of the observations, which become difficult to apply as the complexity of the models increases. From our view point the procedure for inference should not be confused with the dynamic nature of the model. In all the examples considered in this paper the observations were fixed, that is, all them have been already measured. Hence, disregarding (computationally) the temporal structure of the data, like INLA does, possibilitates their full Bayesian analysis, even for realistically complex cases, more easily.

Yet another advantage of the INLA approach is its suitability to perform model comparison. Marginal likelihoods, for example, which can be used as a basis to compare competing models through the Bayes factor, can be easily computed with INLA. Additionally, the deviance information criterion (Spiegelhalter et al., 2002) and two predictive measures, the conditional predictive ordinate and the probability integral transform, used to validate and compare models

and as a tool to detect atypical observations, are available from the INLA output (see Martino and Rue, 2010, for implementation details).

Reasonable results were found for most of the examples considered in this paper using the default INLA values for the hyperprior parameters and for the initial values of these parameters. However, for some models this choice can greatly impact the final results. Therefore, a sensitivity analysis to address the choice of those values is highly recommended.

The extension of the proposed framework to consider multivariate observations is straightforward. The approach has also potential to be applied/extended to other classes of models such as models with errors in covariates. This is subject of current research.

Acknowledgments

R. Ruiz-Cárdenas was partially founded by CAPES (Brazil).

References

- Anderson and J. B. Moore, Optimal Filtering. Prentice-Hall, Englewood Cliffs, NJ.
- Andrieu, C., Doucet, A. Singh, S.S. and Tadić V.B. (2004). Particle Methods for Change Detection, System Identification, and Control. *Proceedings of the IEEE*, 92: 423–438.
- Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B*, 72: 269–342.
- Bivand, R. (2010). spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-9. <http://CRAN.R-project.org/package=spdep>
- Cappé, O., Godsill, S.J. and Moulines, E. (2007). An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo. *Proceedings of the IEEE*, 95: 899–924.
- Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81: 541–53.
- Carter, C.K. and Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika*, 83: 589–601.
- Dethlefsen, C. and Lundbye-Christensen, S. (2006). Formulating State Space Models in R with Focus on Longitudinal Regression Models. *Journal of Statistical Software*, 16: 1–15.
- Doucet, A. and Tadić, V.B. (2003). Parameter estimation in general state-space models using particle methods. *Annals of the Institute of Statistical Mathematics*, 55: 409–422.
- Durbin J, Koopman S.J. (2000). Time Series Analysis of Non-Gaussian Observations Based on State Space Models from both Classical and Bayesian Perspectives (with discussion). *Journal of the Royal Statistical Society Series B*, 62: 3–56.
- Durbin, J. and Koopman, S.J. (2001). Time Series Analysis by State Space Methods. Oxford University Press.
- Ehlers, R.S. and Gamerman, D. (1996). Analytic approximations for dynamic non-linear models. *Brazilian Journal of Probability and Statistics*, 10: 87–101.

- Eidsvik, J., Finley, A.O., Banerjee, S. and Rue, H., 2010. Approximate Bayesian Inference for Large Spatial Datasets Using Predictive Process Models. Preprint Statistics No 9/2010, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. URL: <http://www.math.ntnu.no/preprint/statistics/2010/S9-2010.pdf>.
- Franco, G.C., Santos, T.R., Ribeiro, J.A. and Cruz, F.R.B. (2008). Confidence intervals for hyperparameters in structural models. *Communications in Statistics: Simulation and Computation*, 37: 486–497.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15: 183–202.
- Gamerman, D. (1998). Markov Chain Monte Carlo for Dynamic Generalised Linear Models. *Biometrika*, 85: 215–227.
- Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F*, 140: 107–113.
- Hipel, K.W. and McLeod, I.A. (1994). Time Series Modeling of Water Resources and Environmental Systems. Elsevier Science Publishers B.V. (North-Holland).
- Harvey, A.C. and Durbin, J. (1986). The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling (with discussion). *Journal of the Royal Statistical Society series A*, 149: 187–227.
- Held L., Schrödle B. and Rue H., 2010. Posterior and Cross-validators Predictive Checks: A Comparison of MCMC and INLA, In Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir. Editors: Tutz, G. and Kneib, T., Physica-Verlag, Heidelberg. p. 91–110.
- Helske, J. (2010). KFAS: Kalman filter and smoothers for exponential family state space models. R package version 0.6.0. <http://CRAN.R-project.org/package=KFAS>.
- Koopman, S.J., Shephard, N. and Doornik, J.A. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal*, 2: 113–166.
- Luethi, D., Erb, P. and Otziger, S. (2009). FKF: Fast Kalman Filter. R package version 0.1.0. <http://CRAN.R-project.org/package=FKF>
- Martino, S. and Rue, H. (2010). Implementing Approximate Bayesian Inference using Integrated Nested Laplace Approximation: a manual for the inla program. Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. Compiled on April 8, 2010. URL: <http://www.math.ntnu.no/~hrue/inla/manual.pdf>.
- Migon, H.S., Gamerman, D., Lopes, H.F. and Ferreira, M.A.R. (2005). Dynamic Models. In: *Handbook of Statistics*, 25, D.K. Dey and C.R. Rao (Eds), Elsevier:North-Holland, 553–588.
- Pole, A., West, M. and Harrison, J. (1994). Applied Bayesian Forecasting and Time Series Analysis, New York, Chapman & Hall.
- Pole, A. and West, M. (1990). Efficient Bayesian learning in non-linear dynamic models. *Journal of Forecasting*, 9: 119–136.
- Petris, G. (2010). dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models. R package version 1.1-1. <http://CRAN.R-project.org/package=dlm>.

- Pettit, L. I., 1990. The conditional predictive ordinate for the normal distribution, *Journal of the Royal Statistical Society, Series B* 52, 175–184.
- Plummer, M., Best, N. Cowles, K. and Vines, K., 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC, *R News* 6, 7–11.
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Reis, E.A., Salazar, E. and Gamerman, D. (2006). Comparison of Sampling Schemes for Dynamic Linear Models. *International Statistical Review*, 74: 203–214.
- Riebler, A., Held, L. and Rue, H., 2011. Estimation and extrapolation of time trends in registry data - Borrowing strength from related populations. *Annals of Applied Statistics*, (to appear).
- Riebler, A., Held, L., Rue, H. and Bopp, M., 2011. Gender-specific differences and the impact of family integration on time trends in age-stratified swiss suicide rates. *Journal of the Royal Statistical Society, Series A*, (to appear).
- Ripley, B.D. (2002). Time Series in R 1.5.0. *R News*, 2: 2–7.
- Roos, M., and Held, L., 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis* 6, 259–278.
- Rue, H. and Follestad, T. (2002). GMRFLib: a C-library for fast and exact simulation of Gaussian Markov random fields. Preprint series in statistics No 1/2002, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. URL: <http://www.stat.ntnu.no/preprint/statistics/2002/S1-2002.ps>.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman and Hall/CRC Press.
- Rue H. and Martino S. (2007). Approximate Bayesian Inference for Hierarchical Gaussian Markov Random Fields Models. *Journal of Statistical Planning and Inference*, 137: 3177–3192.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society series B*, 71: 319–392.
- Schrödle, B. and Held, L. (2009). Evaluation of case reporting data from Switzerland: Spatio-temporal disease mapping using INLA. Technical Report, Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich.
- Schrödle, B., Held, L. Riebler, A. and Danuser, J., 2011. Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60, 261–279.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society series B*, 64: 583–639.
- Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE IEEE Transactions on Signal Processing*, 50: 281–289.
- Vivar, J. C. and Ferreira, M. A. R. (2009). Spatiotemporal models for gaussian areal data. *Journal of Computational and Graphical Statistics*, 18: 658–674.

West, M., Harrison, P.J. and Migon, H. (1985). Dynamic generalized linear model and Bayesian forecasting (with discussion). *Journal of the American Statistical Association*, 80: 73–97.

West, M., Harrison, P.J. and Pole, A. (1988). Bats - A user guide - Bayesian Analysis of Time Series - release 1.3 - June 1988, University of Warwick.

West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Second edition, New York:Springer.

Whiteley, N., Andrieu, C. and Doucet, A. (2010). Efficient Bayesian Inference for Switching State-Space Models using Particle Markov Chain Monte Carlo Methods. Statistics Group report 10:04, University of Bristol, Department of Mathematics, 27pp.

Zoeter, O. and Heskes, T. (2006). Deterministic approximate inference techniques for conditionally Gaussian state space models. *Statistics and Computing*, 16: 279-292.

A R script for fitting the second order dynamic spatio-temporal model in example 6

```
## simulating the data set

## Loading North Carolina's map (it has 100 areas)
require(spdep)
ncfile <- system.file("etc/shapes/sids.shp", package="spdep")[1]
nc <- readShapePoly(ncfile)

# building the structure matrix (C)
nc.nb <- poly2nb(nc)
d <- sapply(nc.nb, length)           # vector with number of neighbors
C <- nb2mat(nc.nb, style="B")       # structure matrix

n <- length(d)

## simulated values for tau_i and phi_i (i=1,2,3)
tau <- c(30, 50, 70)
phi <- c(0.8, 0.9, 0.9)

# building the precision matrix
lamb.max <- max(eigen(C, only.values=TRUE)$values) # maximum eigenvalue of C matrix
Q1 <- (diag(n)-phi[1]/lamb.max*C)
Q2 <- (diag(n)-phi[2]/lamb.max*C)
Q3 <- (diag(n)-phi[3]/lamb.max*C)

myrmvnorm <- function(n, mu, S)
  sweep(matrix(rnorm(n*nrow(S)), n)%%chol(S), 2, mu)

## defining the length of time series (number of years)
k <- 30

set.seed(1)
```

```

## simulating observational and innovation errors
w1 <- t(myrmvnorm(k, rep(0,n), solve(tau[1]*Q1)))
w2 <- t(myrmvnorm(k, rep(0,n), solve(tau[2]*Q2)))
w3 <- t(myrmvnorm(k, rep(0,n), solve(tau[3]*Q3)))

## generating the time series for observations and states
yy <- x1 <- x2 <- matrix(0, n, k)
x1[,1] <- w2[,1]
x2[,1] <- w3[,1]
for (i in 2:k) {
  x2[,i] <- x2[,i-1] + w3[,i]
  x1[,i] <- x1[,i-1] + x2[,i-1] + w2[,i]
}
yy <- x1 + w1

## defining the Cmatrix to use with model='generic1' for w1, w2 and w3
st.cmat <- kronecker(C, diag(k))
c.mat <- as(st.cmat, "dgTMatrix")

## building the augmented model
## -----
nd <- n*k
Y <- matrix(NA, nd*3-2*n, 3)
Y[1:nd, 1] <- as.vector(t(yy))
Y[1:(nd-n) + nd, 2] <- 0
Y[1:(nd-n) + 2*nd-n, 3] <- 0

## indices for the f() function
## -----
id1 <- (1:nd)[-((1:n)*k)]
id2 <- (1:nd)[-c(1,((1:(n-1))*k)+1)]
id3 <- seq(1,nd,k)
ix1 <- c(1:nd, id2, rep(NA,nd-n)) ## indices for x1_t
ix1b <- c(rep(NA,nd), id1, rep(NA,nd-n)) ## indices for x1_{t-1}
wx1b <- c(rep(NA,nd), rep(-1,nd-n), rep(NA,nd-n)) ## weights for x1_{t-1}
ix2 <- c(rep(NA,nd),rep(NA,nd-(2*n)),id3, id2) ## indices for x2_t
wx2 <- c(rep(NA,nd), rep(NA,nd-n), rep(1,nd-n)) ## weights for x2_t
ix2b <- c(rep(NA,nd), rep(id1, 2)) ## indices for x2_{t-1}
wx2b <- c(rep(NA,nd), rep(-1,2*(nd-n))) ## weights for x2_{t-1}
iw1 <- c(1:nd, rep(NA,2*(nd-n))) ## indices for w1_t
iw2 <- c(rep(NA,nd), id2, rep(NA,nd-n)) ## indices for w2_t
ww2 <- rep(c(NA,-1,NA), c(nd,length(id2),nd-n)) ## weights for w2_t
iw3 <- c(rep(NA,nd),rep(NA,nd-n), id2) ## indices for w3_t
ww3 <- rep(c(NA,-1), c(nd+nd-n,length(id2))) ## weights for w3_t

## formulating the model with default prior for log-precision parameters
formula1 = Y ~ f(iw1, model="generic1", Cmatrix=c.mat) +
  f(ix1, model="iid", initial=-10, fixed=T) +
  f(ix1b, wx1b, copy="ix1") +
  f(ix2, wx2, model="iid", initial=-10, fixed=T) +
  f(ix2b, wx2b, copy="ix2") +

```

```
f(iw2, ww2, model="generic1", Cmatrix=c.mat) +  
f(iw3, ww3, model="generic1", Cmatrix=c.mat) -1  
  
## call to fit the model  
## -----  
r1 <- inla(formula1,  
  data = data.frame(ix1,ix1b,wx1b,ix2,wx2,ix2b,wx2b,iw1,iw2,iw3,ww2,ww3),  
  family = rep("gaussian",3),  
  control.data = list(list(initial=10, fixed=T), list(initial=10, fixed=T),  
    list(initial=10, fixed=T)))
```