Direct fitting of dynamic models using integrated nested Laplace approximations — INLA

Ramiro Ruiz-Cárdenas [a,*], Elias T. Krainski [b], Håvard Rue [c]

[a] *Department of Statistics, Federal University of Minas Gerais, Belo Horizonte, Brazil*
[b] *Department of Statistics, Federal University of Paraná, Curitiba, Brazil*
[c] *Norwegian University for Science and Technology, Trondheim, Norway*

A B S T R A C T

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 dataset 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 datasets 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. R code is available online for all the examples presented.

© 2011 Elsevier B.V. All rights reserved.

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 dataset is fixed, 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 datasets and, therefore, enable us to do fully Bayesian analysis of complex state-space models more easily.

* Correspondence to: Department of Statistics, Federal University of Minas Gerais, CEP 31270901, Belo Horizonte, Brazil. Tel.: +55 31 34095905; fax: +55 31 34095924.
 E-mail addresses: ramiro@est.ufmg.br, ramiro_rc1@yahoo.com.br (R. Ruiz-Cárdenas), eliaskr@ufpr.br (E.T. Krainski), hrue@math.ntnu.no (H. Rue).

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 (see the supplementary material for a brief review on the INLA approach). 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 geoadditive 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, 2012), 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 how the INLA approach can be extended to allow for the formulation of specific latent models in a state-space form in order to perform approximate Bayesian inference on them using INLA, even in the above complex cases.

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 enables 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 organized 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. The proposed framework to fit state-space models using INLA is illustrated in Section 3 through a series of simulated examples, ranging from simple univariate models to realistically complex spatio-temporal dynamic models. In Section 4 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 5.

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

$$y_t = F_t' x_t + v_t, \quad v_t \sim N(0, V_t) \tag{1}$$

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

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 v_t and ω_t are usually 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. It uses a singular value decomposition-based algorithm to decompose the relevant variance matrices needed in the filtering, smoothing and simulation steps. Unknown parameters can be estimated by maximum likelihood or Bayesian inference, but just for the Gaussian linear case.

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., 2009), 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, 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 (one exception is Chan and Jeliazkov (2009)), 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 dataset 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.

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

3. INLA for state-space models

In this section we illustrate, through a series of simulated datasets, the proposed computational framework to formulate and fit the most common types of dynamic models using the INLA approach. Firstly, the steps of the proposed framework to perform fast full Bayesian inference on these models using the INLA library are described in detail using a simple univariate
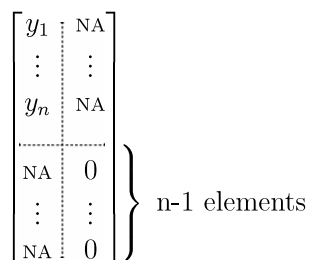
$$\left.\begin{bmatrix} y_1 & \vdots & \text{NA} \\ \vdots & & \vdots \\ y_n & \vdots & \text{NA} \\ \cdots & & \cdots \\ \text{NA} & \vdots & 0 \\ \vdots & & \vdots \\ \text{NA} & \vdots & 0 \end{bmatrix}\right\} \text{n-1 elements}$$

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

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 log-gamma (1, 0.00005). As little sensitivity to prior specification was, in general, observed in all the cases and due to space constraints, marginal posterior densities for the hyperparameters with these priors are only detailed for some of the examples. Full results can be found at the technical report by Ruiz-Cárdenas et al. (2010).

The analyses of all the examples in this paper were carried out in an Intel Core 2 Duo model laptop at 2.40 GHz and 4 GB RAM running under a Windows 7 operating system. Computer time is reported for the most representative cases. The complete R code to simulate and fit all the examples in this section is available online as supplementary material. Further coding details can also be found in Ruiz-Cárdenas et al. (2010).

3.1. 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:

$$y_t = x_t + \nu_t, \qquad \nu_t \sim N(0, V), \ t = 1, \ldots, n \tag{3}$$
$$x_t = x_{t-1} + \omega_t, \qquad \omega_t \sim N(0, W), \ t = 2, \ldots, n. \tag{4}$$

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. Therefore, the vector of hyperparameters is given by $\boldsymbol{\theta} = \{V, W\}$, while the latent field corresponds to $\boldsymbol{\xi} = \{x_1, \ldots, x_n\}$. 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 (see R code at Appendix B). 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 direct fitting of complex state-space models as will be shown later on in this section. The key feature of this generic approach begins equating to zero the system equations of the state-space model, that is, in our simple example, we re-write (4) as

$$0 = x_t - x_{t-1} - \omega_t, \quad \omega_t \sim N(0, W), \ t = 2, \ldots, n. \tag{5}$$

Then we build an augmented model, which will have dimension $n + (n-1)$ in our example, merging these "faked zero observations" from system equation (5) with the actual observations from Eq. (3) in a unique structure, as shown in Diagram 1. The first column of this structure is associated to the actual observations, $\boldsymbol{y}_t = \{y_1, \ldots, y_n\}$, 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. (5), are forced to be zero. All the other entries in the structure have no values (they are filled in with NA's). This augmented structure will have additional columns when the state-space model contains more than one evolution equation.

Inference in this augmented model using the INLA approach is performed considering different likelihoods for each column of the augmented structure (two in our simple example). 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 (5). 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, $t = \{2, \ldots, T\}$, which are the only stochastic terms in the system equation (4). Thus, we are considering that there is no information about x_t
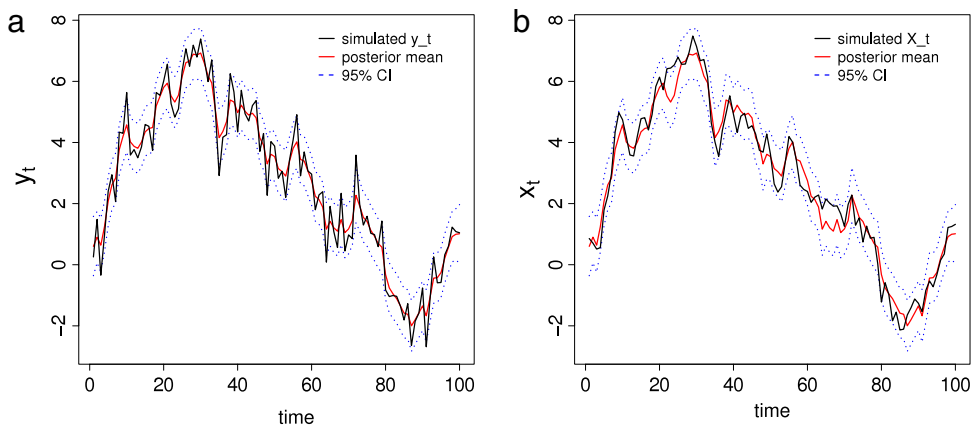
Fig. 1. Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and states (b) in the toy example.
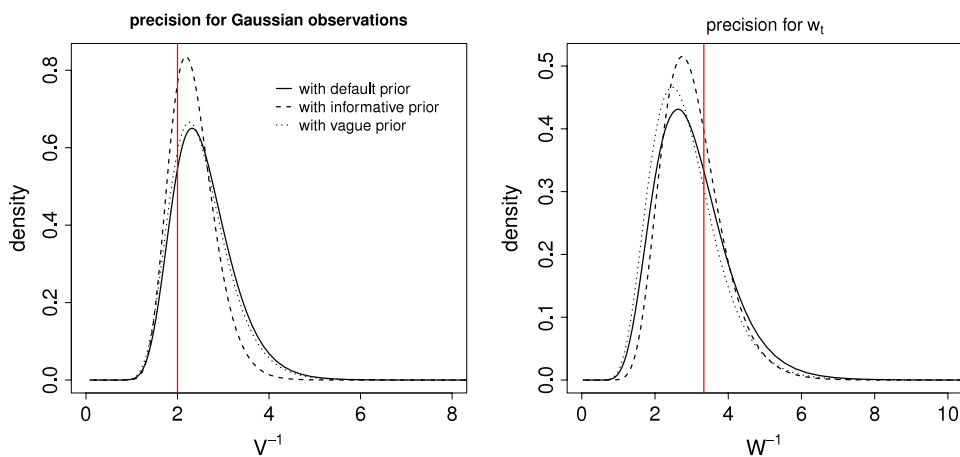

Fig. 2. Marginal posterior densities for the hyperparameters in the toy example. Red lines indicate true simulated values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 $\boldsymbol{x}$ 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 $t = 2, \ldots, n$, x_t follows Eq. (4).

The proposed augmented model framework slightly modifies the standard structure of the state-space model introducing artificial observations following additional likelihoods (just one in this simple 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. The impact of this new variables on the inference of the model when compared with that obtained under a standard modeling approach is formalized in Appendix A.

The series of simulated and estimated values for the vectors of observations and states, respectively, are illustrated in Fig. 1. Precisions of the two perturbation terms were also well estimated, as can be seen in Fig. 2, with credibility intervals always including the true simulated values. The R code that implements the above augmented approach for our toy example using the INLA library is given in Appendix B.

3.1.1. 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 datasets of size $n = 100$ were simulated from the model in (3) and (4), 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.
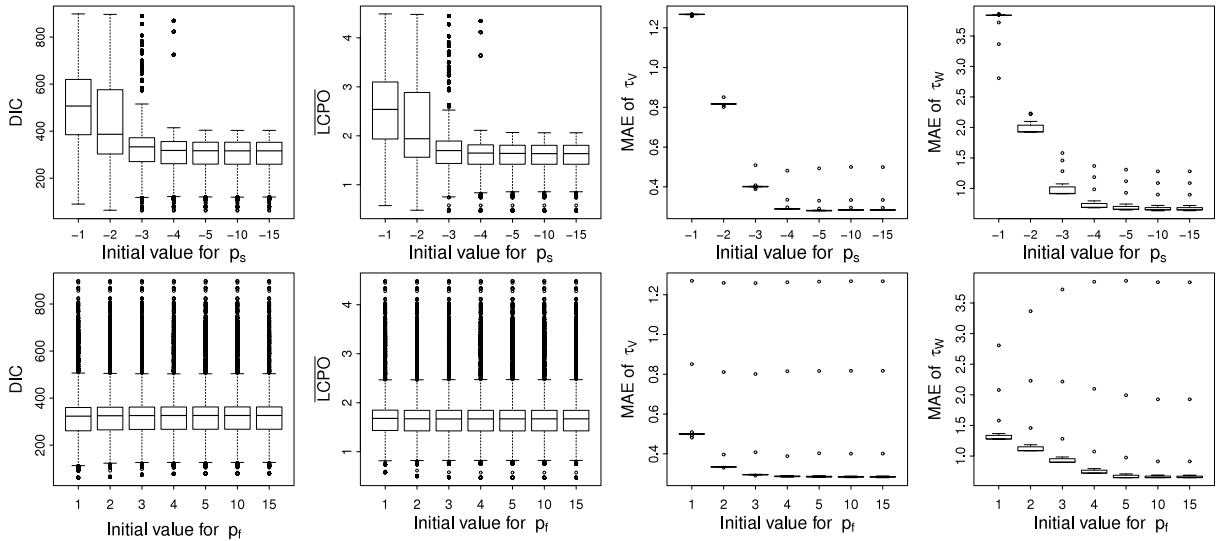
Fig. 3. Boxplots of the Deviance Information Criterion (DIC), the mean logarithmic CPO ($\overline{\text{LCPO}}$) and the mean absolute error (MAE) of the posterior mean of hyperparameters τ_V and τ_W for different initial values of the fixed log-precision of the states (upper frame) and "faked" observations (lower frame) in the toy example.

Each dataset was fitted 225 times with different combinations of initial values of p_f and p_s from a 15×15 grid, where $p_f \in \{1, 2, \ldots, 14, 15\}$ and $p_s \in \{-1, -2, \ldots, -14, -15\}$. 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 $\overline{\text{LCPO}} = -\frac{1}{n} \sum_{i=1}^{n} \log(\text{CPO}_i)$, where $\text{CPO}_i = \pi(y_i \mid \boldsymbol{y}_{-i})$ and the subscript $-i$ indicates that element i of vector of observations $\boldsymbol{y}$ was removed. Lower values of MAE, DIC and $\overline{\text{LCPO}}$ 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 artifact to allow the modeling of the states but do not constitute real observations itself. The above 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 were used to 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 Fig. 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 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.

3.1.2. Comparison with MCMC

The same 1000 simulated datasets generated in Section 3.1.1 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 (3) and (4) 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, $\boldsymbol{x}$, were sampled through FFBS (forward filtering backward sampling – Frühwirth-Schnatter, 1994; Carter and Kohn, 1994). From a unique MCMC chain 55 000 posterior samples were drawn by using a thinning of 20 and a burn-in of 5000. Hence, 2500

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

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 datasets $\times$ 3 types of priors) was vastly favorable to the INLA approach, with 2.1 s against 260.3 s spent by the MCMC algorithm.

3.2. 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. Some of these functionalities are illustrated in this subsection through an example where a generalized dynamic regression model is fitted using existing model options from the INLA library. The main features of this model are 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 generalized dynamic regression model
We simulated data from a multiple Poisson regression model with two regressors, Z_{1t} and Z_{2t}. Thus, the linear predictor is given by $\lambda_t = F_t x_t$, where $F_t = (1, Z_{1t}, Z_{2t})$ and the regression coefficients $x_t = (\beta_{0t}, \beta_{1t}, \beta_{2t})$ follow a simple random walk evolution. The model has the following observational and system equations:

$$(y_t \mid \mu_t) \sim \text{Poisson}(\mu_t)$$
$$\log(\mu_t) = \lambda_t = \beta_{0t} + \beta_{1t} Z_1 + \beta_{2t} Z_2 \quad t = 1, \ldots, n$$
$$\beta_{0t} = \beta_{0,t-1} + \omega_{0t}, \qquad \omega_{0t} \sim N(0, W_0), \quad t = 2, \ldots, n$$
$$\beta_{1t} = \beta_{1,t-1} + \omega_{1t}, \qquad \omega_{1t} \sim N(0, W_1), \quad t = 2, \ldots, n$$
$$\beta_{2t} = \beta_{2,t-1} + \omega_{2t}, \qquad \omega_{2t} \sim N(0, W_2), \quad t = 2, \ldots, n$$

The interest is to obtain full Bayes estimates of the dynamic regression coefficients, $\beta_{it} = \{\beta_{i1}, \ldots, \beta_{in}\}$, 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 s 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 Fig. 4. The marginal posterior densities of the hyperparameters are also available at the online supplementary material.

3.3. 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
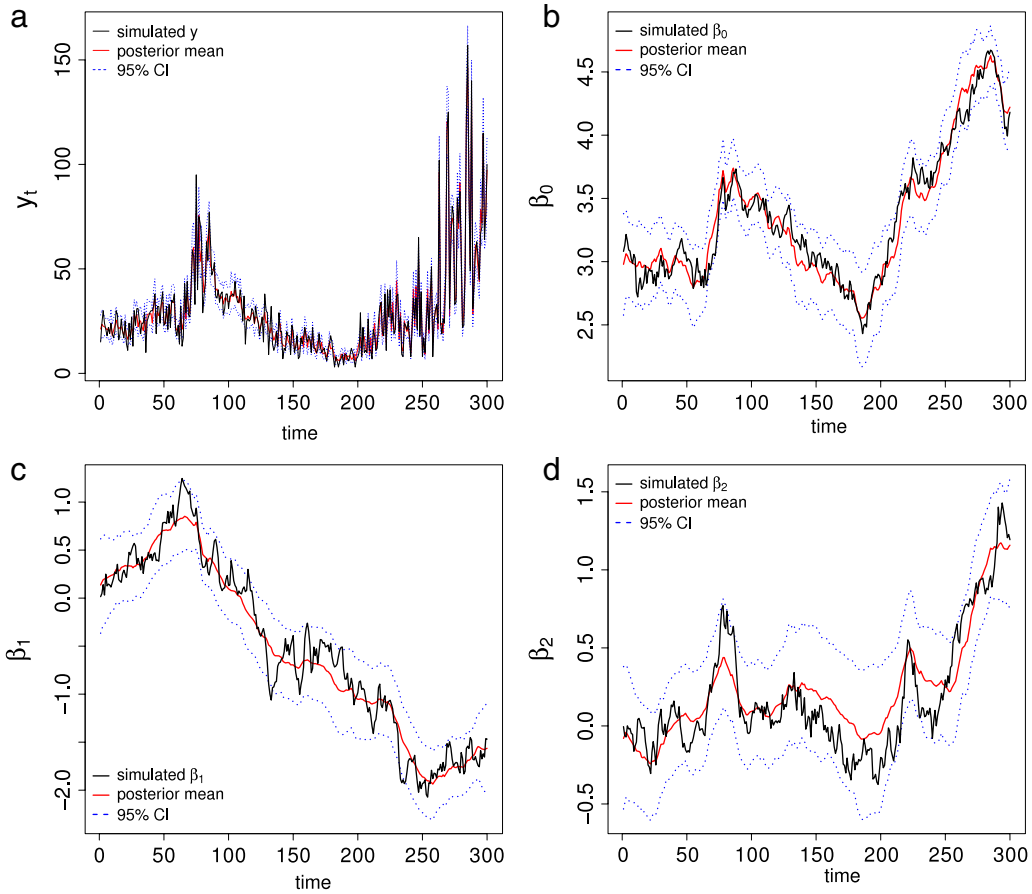
Fig. 4. Simulated and predicted values (posterior mean and 95% credibility interval) for the observations (a) and regression coefficients, β_0, β_1 and β_2 (b–d) in the generalized dynamic regression example.

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 Section 3.1 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.

3.3.1. Models with different evolution structures for the states

Here a dynamic seasonal model with two harmonics is 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 2: 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:

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

with observational and system equations given by

$$y_t = a_t + v_t, \quad v_t \sim N(0, V), \ t = 1, \ldots, n \tag{6}$$

$$a_t = \cos(\phi)a_{t-1} + \sin(\phi)b_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \ t = 2, \ldots, n \tag{7}$$

$$b_t = -\sin(\phi)a_{t-1} + \cos(\phi)b_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), \ t = 2, \ldots, n. \tag{8}$$

Note that the temporal evolution for a_t in (7) depends on both, a_{t-1} and b_{t-1}. This evolution structure cannot be directly accounted for the INLA library, therefore, we made use here of the augmented model structure described in Section 3.1, equating (7) and (8) to zero and then merging these "pseudo" observations with the actual observations from Eq. (6). The augmented structure has dimension $n + 2(n - 1)$ and three different likelihoods, being the elements in the first column
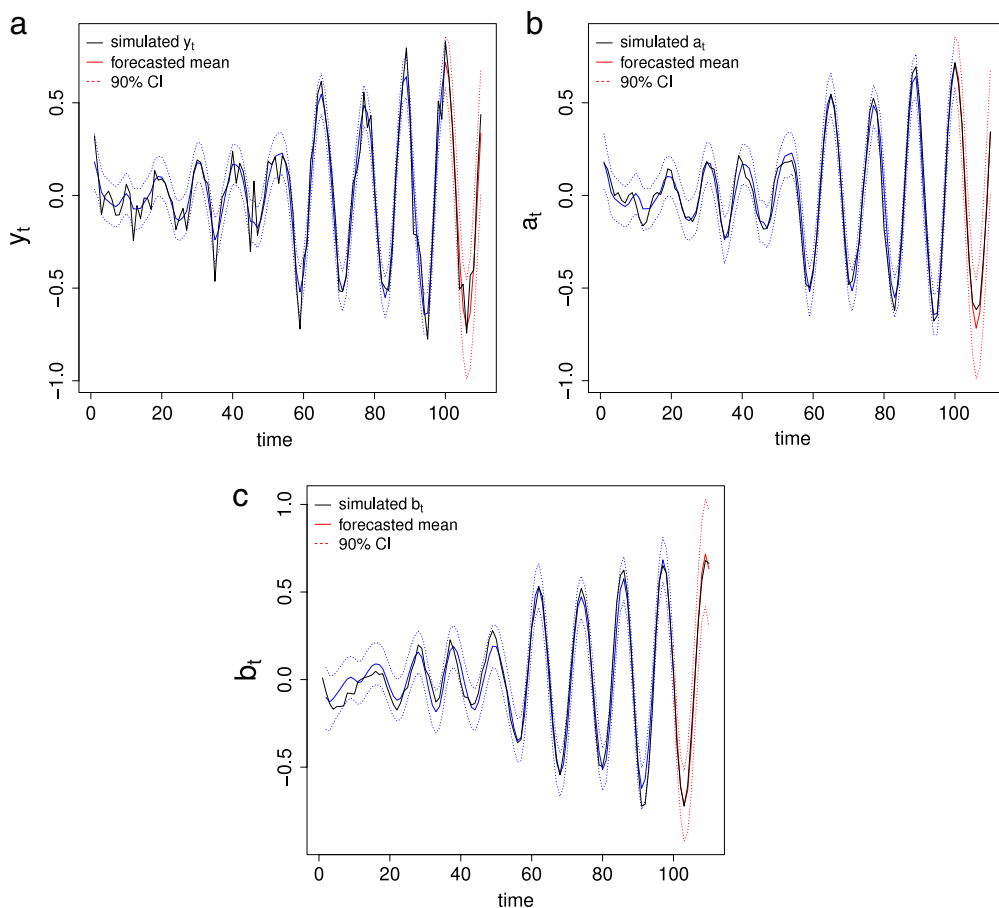
Fig. 5. Simulated and predicted values (posterior mean and 90% credibility interval) for the observations (a) and for a_t (b) and b_t (c) state vectors in the harmonic seasonal model. Red lines are for the 10 steps ahead forecast. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 s. Posterior means for the observations and states, including the forecast period, and its 90% credibility interval are shown in Fig. 5.

3.3.2. 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 3: 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, $t = \{1, \ldots, T\}$; $s = \{1, \ldots, S\}$, the response y_{ts} is specified as:

$$\boldsymbol{y}_t = \boldsymbol{F}_t'\boldsymbol{x}_t + \boldsymbol{\Upsilon}_t\boldsymbol{z}_t + \boldsymbol{\omega}_{1t}, \quad \boldsymbol{\omega}_{1t} \sim \text{PGMRF}\left(\boldsymbol{0s}, \boldsymbol{W}_1^{-1}\right) \tag{9}$$

$$\boldsymbol{x}_t = \boldsymbol{G}_t\boldsymbol{x}_{t-1} + \boldsymbol{\omega}_{2t}, \quad \boldsymbol{\omega}_{2t} \sim \text{PGMRF}\left(\boldsymbol{0s}, \boldsymbol{W}_2^{-1}\right) \tag{10}$$

$$\boldsymbol{z}_t = \boldsymbol{z}_{t-1} + \boldsymbol{\omega}_{3t}, \quad \boldsymbol{\omega}_{3t} \sim \text{PGMRF}\left(\boldsymbol{0s}, \boldsymbol{W}_3^{-1}\right) \tag{11}$$

with $\boldsymbol{y}_t = (y_{t1}, \ldots, y_{tS})'$ and $\boldsymbol{\Upsilon}_t = (\Upsilon_{t1}, \ldots, \Upsilon_{tS})'$ denoting the observed field and the vector of covariates at time t, respectively. $\boldsymbol{F}_t = \boldsymbol{Is}$, $\boldsymbol{G}_t = \rho\boldsymbol{Is}$, $\boldsymbol{0s}$ is an $S \times S$ null matrix, and $\boldsymbol{Is}$ is the $S \times S$ identity matrix. The vectors of errors $\boldsymbol{\omega}_{i1}, \ldots, \boldsymbol{\omega}_{iT}$, $i = \{1, 2, 3\}$, are assumed to be independent and modeled as proper Gaussian Markov random fields (PGMRF).
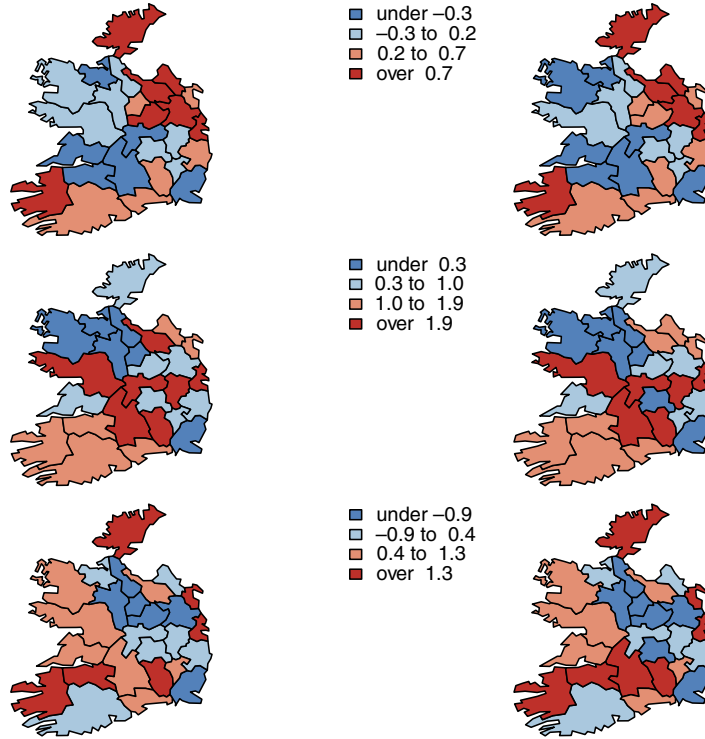
Fig. 6. Maps of simulated (left) and predicted values (right) for observations, $\boldsymbol{y}_t$, at times 18, 51 and 77 in example 3.

Matrices $\boldsymbol{W}_i$ describe the spatial covariance structure of $\boldsymbol{\omega}_{it} = (\omega_{it1}, \ldots, \omega_{itS})'$. Precision matrices $\boldsymbol{W}_i^{-1}$ are modeled as $\boldsymbol{W}_i^{-1} = \tau_i \left(\boldsymbol{Is} - \frac{\phi_i}{\lambda_{\max}} \boldsymbol{C} \right)$, with $\boldsymbol{C}$ 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 $h_{k,l} = 1$). $\lambda_{\max}$ is the maximum eigenvalue of matrix $\boldsymbol{C}$; 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.

Predicted observations and states, in general, closely followed their corresponding simulated values for all areas, as shown in the maps of Figs. 6–8 for some instant times. See also the complete time series of simulated and predicted values for some areas at the online supplementary material. Precision and correlation parameters were also well estimated even when default INLA values for hyperprior parameters and initial values were specified, as can be seen in the figure of marginal posterior densities for these hyperparameters at the supplementary material.

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 that five minutes on a modest laptop. Currently inference in models like this involves computationally expensive MCMC algorithms, which are also hard to tune.

Example 4: 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:

$$\boldsymbol{y}_t = \boldsymbol{F}_t' \boldsymbol{x}_t + \boldsymbol{\omega}_{1t}, \quad \boldsymbol{\omega}_{1t} \sim \text{PGMRF} \left(\boldsymbol{0s}, \boldsymbol{W}_1^{-1} \right) \tag{12}$$

$$\boldsymbol{x}_t = \boldsymbol{G}_t \boldsymbol{x}_{t-1} + \boldsymbol{\omega}_{23t}, \quad \boldsymbol{\omega}_{23t} \sim \text{PGMRF} \left(\boldsymbol{0s}, \boldsymbol{W}_{23}^{-1} \right) \tag{13}$$

where,

$$\boldsymbol{x}_t = \begin{pmatrix} \boldsymbol{x}_{1t} \\ \boldsymbol{x}_{2t} \end{pmatrix}, \qquad \boldsymbol{F}_t = \begin{pmatrix} \boldsymbol{Is} \\ \boldsymbol{0s} \end{pmatrix}, \qquad \boldsymbol{G}_t = \begin{pmatrix} \rho_1 \boldsymbol{Is} & \rho_1 \boldsymbol{Is} \\ \boldsymbol{0s} & \rho_2 \boldsymbol{Is} \end{pmatrix}, \qquad \boldsymbol{\omega}_{23t} = \begin{pmatrix} \boldsymbol{\omega}_{2t} \\ \boldsymbol{\omega}_{3t} \end{pmatrix} \quad \text{and} \quad \boldsymbol{W}_{23}^{-1} = \begin{pmatrix} \boldsymbol{W}_2^{-1} & \boldsymbol{0s} \\ \boldsymbol{0s} & \boldsymbol{W}_3^{-1} \end{pmatrix},$$
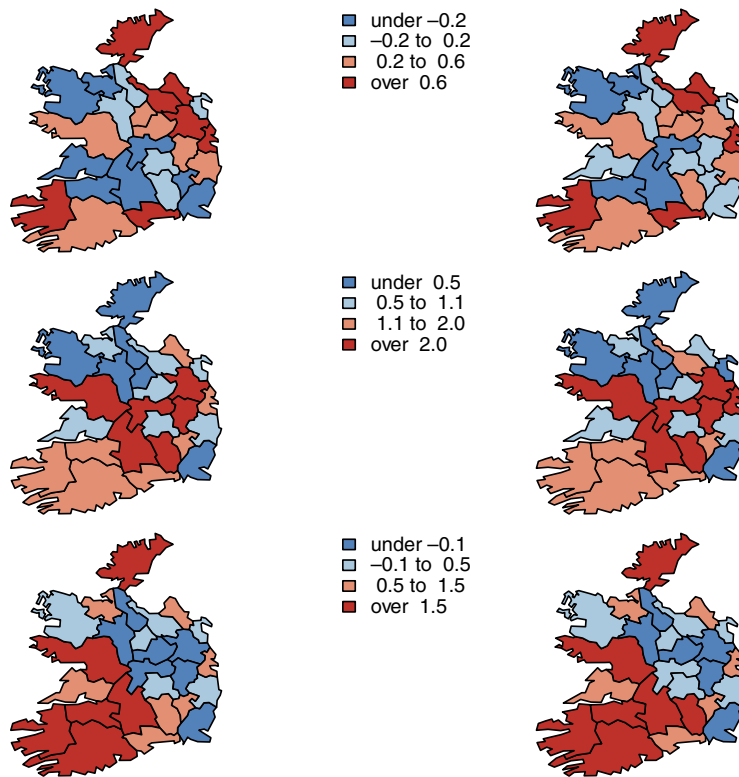
Fig. 7. Maps of simulated (left) and predicted values (right) for state vector, $\boldsymbol{x}_t$, at times 18, 51 and 77 in example 3.
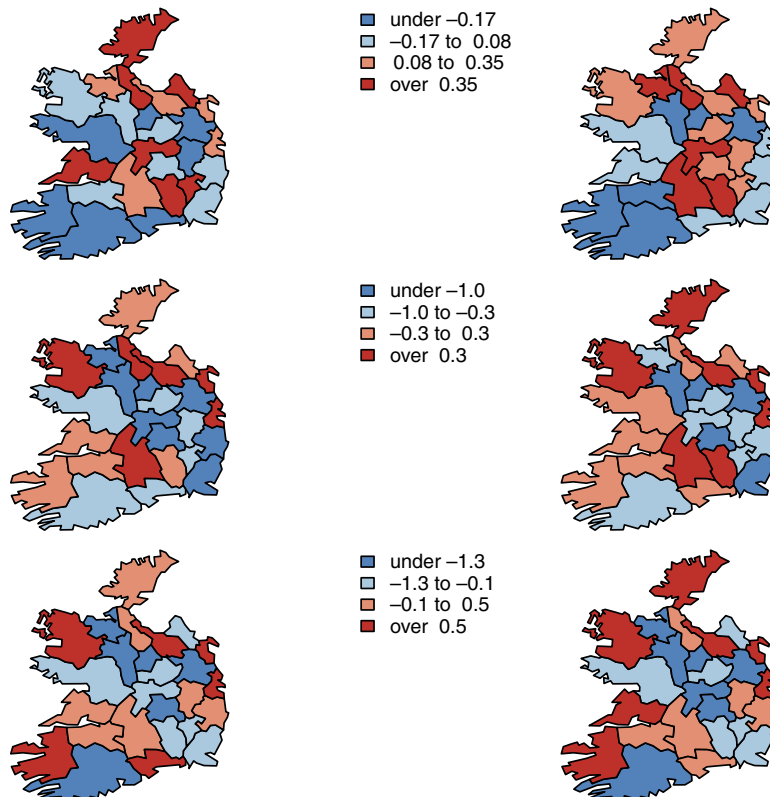

Fig. 8. Maps of simulated (left) and predicted values (right) for state vector, $\boldsymbol{z}_t$, at times 18, 51 and 77 in example 3.
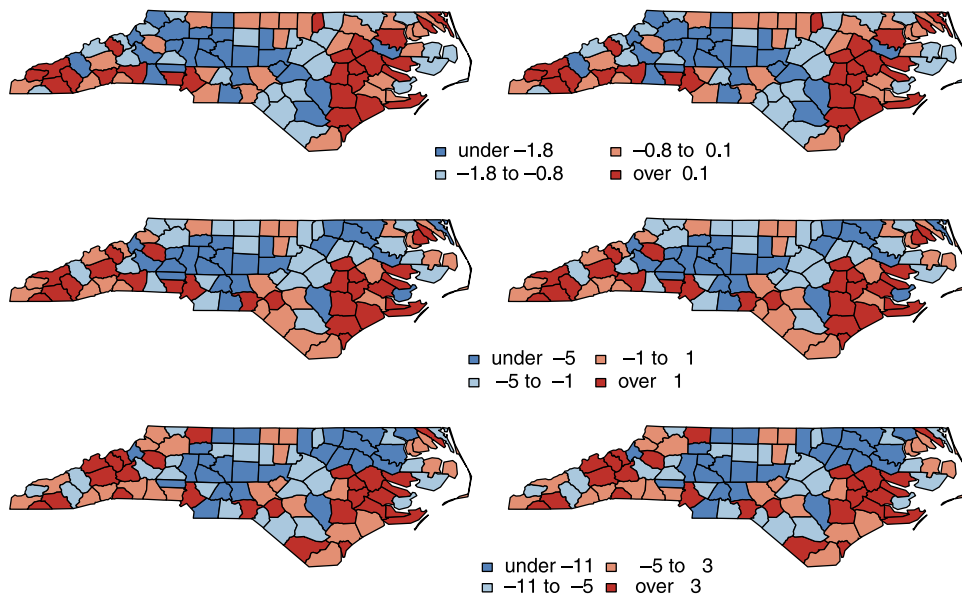
Fig. 9. Maps of simulated (left) and predicted values (right) for observations, y_t, at times 7, 15 and 29 in example 4.
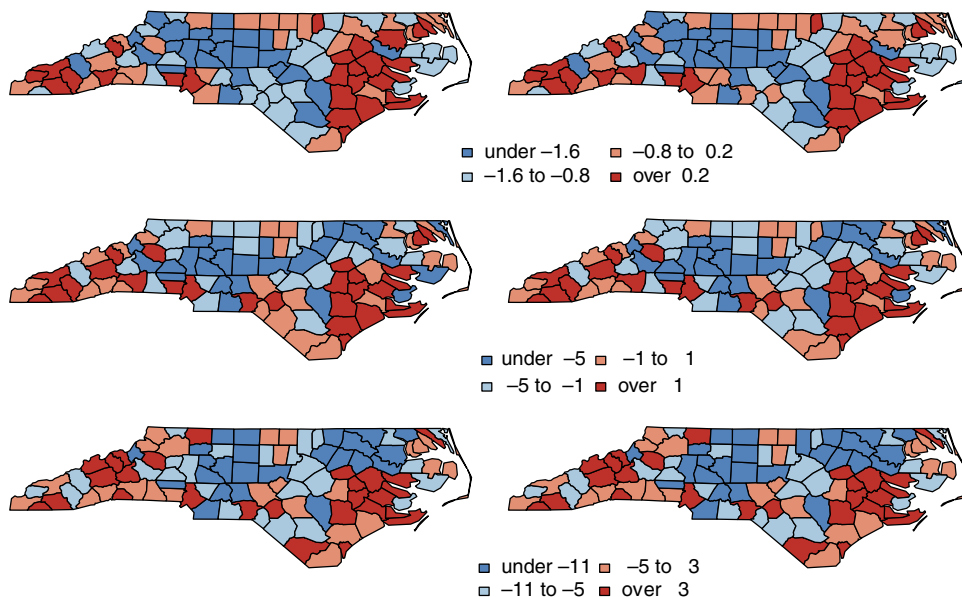

Fig. 10. Maps of simulated (left) and predicted values (right) for X_1 state vector at times 7, 15 and 29 in example 4.

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 x_1 and 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 $\rho_1 = \rho_2 = 1$.

Predicted values here also closely agreed with the simulated series of observations and states in all counties as illustrated in the maps of Figs. 9–11 for some instant times. Complete time series of simulated and predicted values for some areas is also available as online supplementary material. 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 $\{\log(10), \log(20), \ldots, \log(90), \log(100)\}$ keeping the default
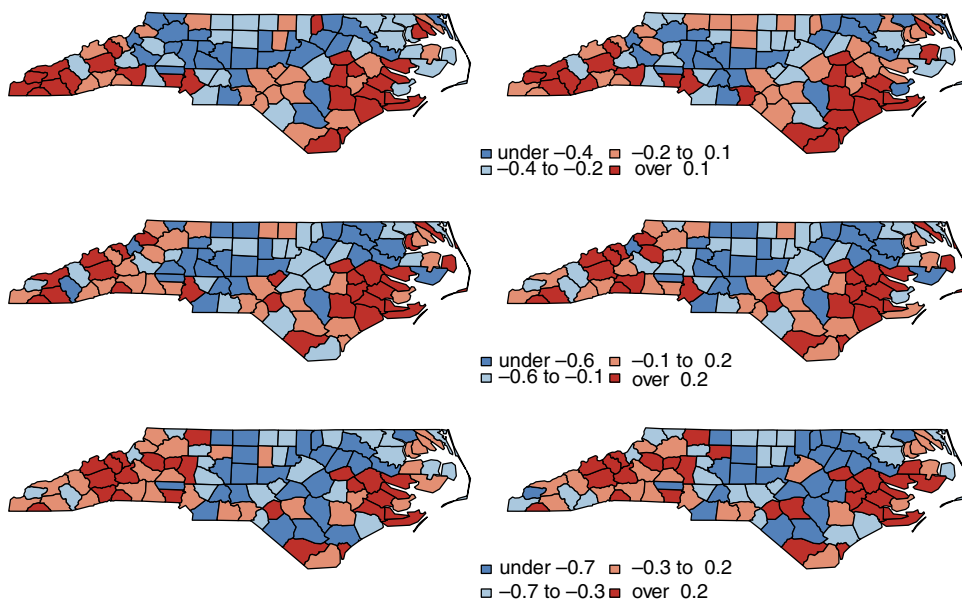
Fig. 11. Maps of simulated (left) and predicted values (right) for X_2 state vector at times 7, 15 and 29 in example 4.

initial values for the correlation parameters, and (ii) varying the initial values for ϕ_i in the set $\{0.1, 0.2, \ldots, 0.8, 0.9\}$ 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 at the online supplementary material. 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.

4. 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. The full code and datasets needed to reproduce all the examples in this section is available online as supplementary material.

Example 5: UK Gas consumption

The first real-life 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

$$y_t = \log_{10}(\text{UK gas})_t = T_t + S_t + v_t, \quad v_t \sim N(0, V), \ t = 1, \ldots, n \tag{14}$$

$$T_t = T_{t-1} + \beta_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \ t = 2, \ldots, n \tag{15}$$

$$\beta_t = \beta_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), \ t = 2, \ldots, n \tag{16}$$

$$S_t = -(S_{t-1} + S_{t-2} + S_{t-3}) + \omega_{3t}, \quad \omega_{3t} \sim N(0, W_3), \ t = 4, \ldots, n. \tag{17}$$

Approximate Bayesian inference in this case was performed using a mixed approach. Firstly, we utilized an augmented structure merging the observational Eq. (14) with the polynomial trend (15). The slope and seasonal terms in Eqs. (16) and (17), 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 dataset using a Gibbs sampling algorithm available in the `dlm` package. In order to obtain suitable (precise) estimates of the parameters of interest for a comparison between MCMC and INLA
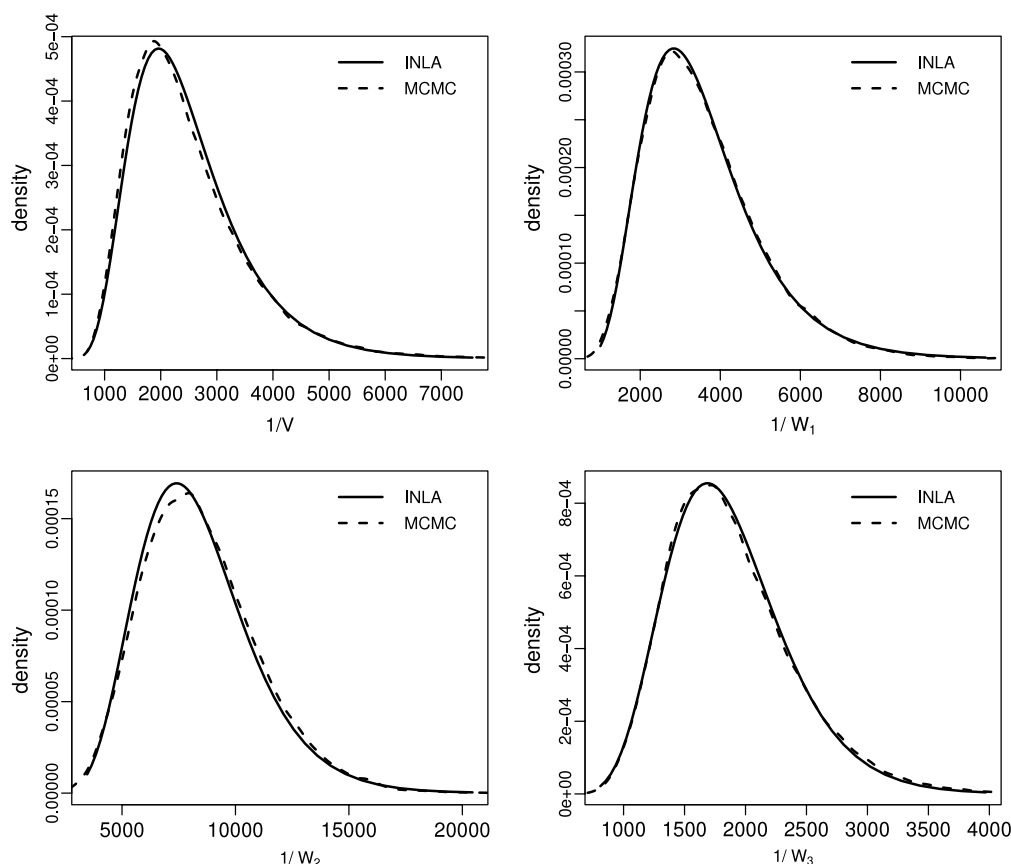
Fig. 12. Marginal posterior densities for the precision parameters obtained with INLA and MCMC in the UK gas consumption example.

approaches, 1 010 000 posterior samples were drawn from a unique MCMC chain by using a thinning of 20 and a burn-in of 10 000. Hence, 50 000 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 dataset. The CODA package (Plummer et al., 2006) was used to diagnose convergence through the Geweke and Raftery–Lewis diagnostic tests. Fig. 12 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 Fig. 13). However, we must be aware that, unlike INLA, forecasting with the dlm 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 Figs. 12 and 13, 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 dlm package used 43 629.75 s to run the MCMC algorithm in R against 3.43 s spent by the INLA library (almost 13 000 times faster). At first glance this comparison of computational time would not seem totally fear since INLA is heavily optimized whereas the dlm 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 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 s), 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 to 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.
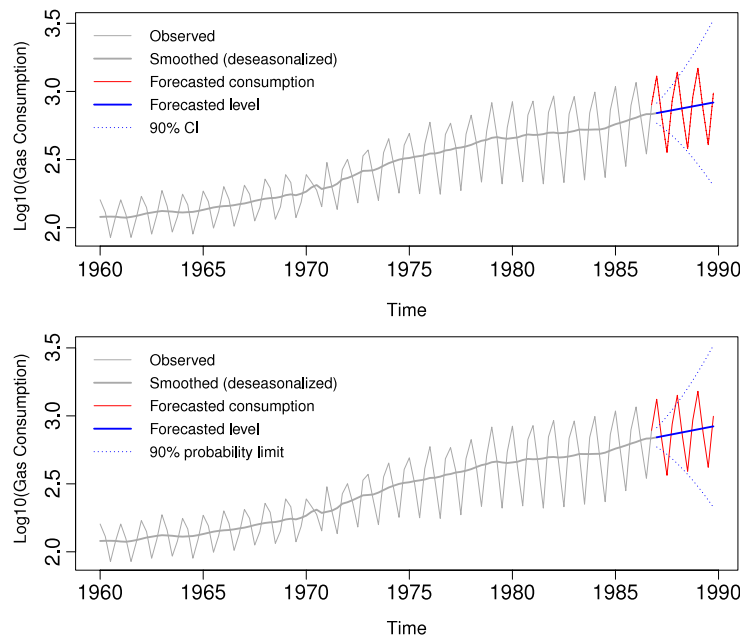
Fig. 13. UK gas consumption forecasts for the next 12 quarterly periods obtained with INLA (upper frame) and with the `dlm` package (lower frame).

Example 6: 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 dataset see Harvey and Durbin (1986) and Durbin and Koopman (2000).

This dataset has been previously analyzed 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

$$y_t \sim \text{Poisson}(\mu_t)$$
$$\log(\mu_t) = \lambda_t = T_t + \alpha * \text{seatbelt}_t + S_t, \quad t = 1, \ldots, n \tag{18}$$
$$T_t = T_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W), \ t = 2, \ldots, n \tag{19}$$
$$S_t = -(S_{t-1} + \cdots + S_{t-11}), \quad t = 12, \ldots, n. \tag{20}$$

The trend and seasonal terms in linear predictor (18) can be directly modeled using existing first order random walk and seasonal model options from the INLA library. 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 Fig. 14.

The posterior mean for α parameter in Eq. (18), 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 7: Mumps

In our last worked the response y_t corresponds to the monthly number of registered cases of mumps in New York City from January 1928 to June 1972. This dataset 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 dataset 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
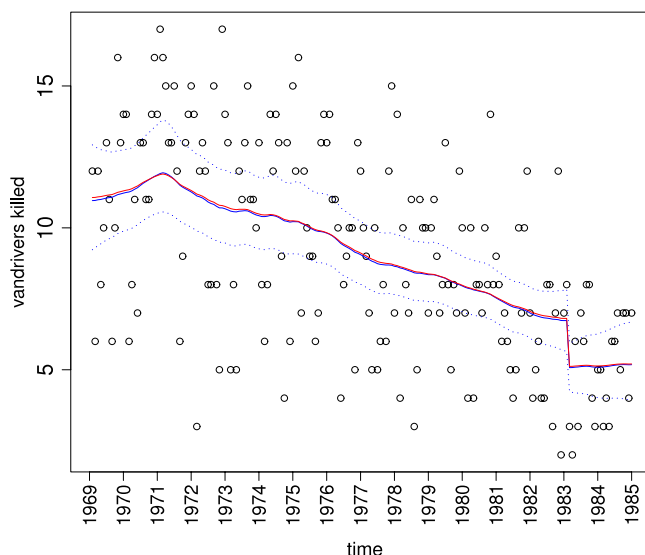
Fig. 14. Number of van drivers killed and estimated trend + intervention. Solid and dotted lines in blue correspond to the posterior mean and 95% credibility intervals, respectively, obtained with the INLA library. The line in red corresponds to the estimated trend + intervention with the `sspir` package. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

harmonic seasonal component (H_t) as suggested in Dethlefsen and Lundbye-Christensen (2006). The observational and system equations for this model are as follows

$$y_t \sim \text{Poisson}(\mu_t)$$
$$\log(\mu_t) = \lambda_t = T_t + H_t, \quad t = 1, \ldots, n$$
$$T_t = T_{t-1} + \beta_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1),\ t = 2, \ldots, n \tag{21}$$
$$\beta_t = \beta_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2),\ t = 2, \ldots, n \tag{22}$$
$$H_t = a_t \cos\left(\frac{2\pi}{12}t\right) + b_t \sin\left(\frac{2\pi}{12}t\right), \quad t = 1, \ldots, n \tag{23}$$
$$a_t = a_{t-1} + \omega_{3t}, \quad \omega_{3t} \sim N(0, W_3),\ t = 2, \ldots, n \tag{24}$$
$$b_t = b_{t-1} + \omega_{4t}, \quad \omega_{4t} \sim N(0, W_4),\ t = 2, \ldots, n. \tag{25}$$

As in example 7, a mixed approach is suitable here, where the polynomial trend in system equation (21) 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 Eqs. (22)–(25) follow a random walk evolution and were just modeled with a first order random walk model option from the INLA library.

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 Fig. 15. They were very similar for the two approaches. According to Fig. 15, 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.

5. 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 Ruiz-Cárdenas et al. (2010). 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
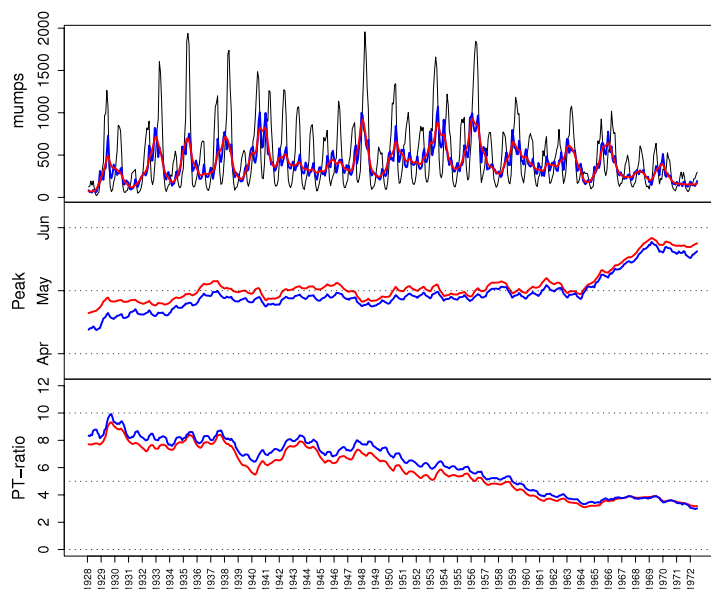
Fig. 15. Comparison between INLA (blue lines) and `sspir` (red lines) results for the variation in the incidence of mumps in New York city from 1927 to 1972. The upper frame shows the observed number of cases jointly with the de-seasonalized trend. The location of the peak of the seasonal pattern is shown in the middle frame, while the lower frame is for the variation in the peak-to-trough ratio over the study period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 `dlm` 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

Appendix A. Impact of the augmented data framework on inference

The standard state-space structure is slightly modified in the model framework proposed in this article with the introduction of new observations following additional likelihoods, as well as additional fixed precisions, associated to these new observations and to the latent states. Here we used the first order univariate dynamic linear model described in Section 3.1, 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

Firstly we will obtain the joint distribution of $\boldsymbol{y}$ given the precision parameters when a first order random walk (RW1) prior is assumed for the states $\boldsymbol{x}$. This is a default option in the INLA approach. Rewriting Eqs. (3) and (4) we can express the state-space model in the toy example in matricial form as

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v} \tag{A.1}$$
$$\boldsymbol{x}_{-1} = \boldsymbol{x}_{-n} - \boldsymbol{\omega} \tag{A.2}$$

with $\boldsymbol{y} = (y_1, \ldots, y_n)$ and $\boldsymbol{x} = (x_1, \ldots, x_n)$ denoting the observations and states vectors, respectively. $\boldsymbol{v} = (v_1, \ldots, v_n)$ and $\boldsymbol{\omega} = (\omega_2, \ldots, \omega_n)$ are the perturbation vectors for the observation and state equations, respectively, and $\boldsymbol{x}_{-i}$ denotes the state vector $\boldsymbol{x}$ without the ith entry. The likelihood for this model is

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

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

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

Then the joint distribution

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

after some calculations can be expressed as

$$\boldsymbol{y}|\tau_V, \tau_W \propto \exp\left\{-\frac{1}{2}\boldsymbol{y}'\left(\tau_V\boldsymbol{I_n} - \tau_V^2\boldsymbol{Q_X}^{-1}\right)\boldsymbol{y}\right\}, \tag{A.3}$$

with $\boldsymbol{I_n}$ being an $n \times n$ identity matrix and

$$\boldsymbol{Q_X} = (\tau_V\boldsymbol{I_n} + \tau_W\boldsymbol{R}). \tag{A.4}$$

Predictive distribution under the augmented model

Following the framework in Section 3.1 we equate to zero the system equation (A.2), which can be rewritten in matricial form as

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

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

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

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

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

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

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

$$\boldsymbol{y} \mid \tau_V, \tau_W, \tau_0, \tau_x \propto \iint \exp\left\{-\frac{\tau_V}{2}(\boldsymbol{y}-\boldsymbol{x})'(\boldsymbol{y}-\boldsymbol{x}) - \frac{\tau_0}{2}(\boldsymbol{\omega}-\boldsymbol{d})'(\boldsymbol{\omega}-\boldsymbol{d})\right\} \times$$

$$\times \exp\left\{-\frac{\tau_x}{2}\boldsymbol{x}'\boldsymbol{x} - \frac{\tau_W}{2}\boldsymbol{\omega}'\boldsymbol{\omega}\right\} d\boldsymbol{\omega} d\boldsymbol{x}$$

$$\propto \exp\left\{-\frac{\tau_V}{2}\boldsymbol{y}'\boldsymbol{y}\right\} \int \exp\left\{-\frac{\tau_V}{2}(\boldsymbol{x}'\boldsymbol{x}-2\boldsymbol{y}'\boldsymbol{x}) - \frac{\tau_0}{2}\boldsymbol{d}'\boldsymbol{d} - \frac{\tau_x}{2}\boldsymbol{x}'\boldsymbol{x}\right\}$$

$$\times \underbrace{\int \exp\left\{-\frac{\tau_0}{2}(\boldsymbol{\omega}'\boldsymbol{\omega}-2\boldsymbol{d}'\boldsymbol{\omega}) - \frac{\tau_W}{2}\boldsymbol{\omega}'\boldsymbol{\omega}\right\} d\boldsymbol{\omega}}_{(II)} d\boldsymbol{x},$$

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

$$\boldsymbol{y} \mid \tau_V, \tau_W, \tau_0, \tau_x \propto \exp\left\{-\frac{\tau_V}{2}\boldsymbol{y}'\boldsymbol{y}\right\} \int \exp\left\{-\frac{1}{2}\left[\boldsymbol{x}'\left((\tau_V+\tau_x)\boldsymbol{I_n}+\tau_0\boldsymbol{R}-\frac{\tau_0^2}{\tau_0+\tau_W}\boldsymbol{R}\right)\boldsymbol{x}-2\tau_V\boldsymbol{y}'\boldsymbol{x}\right]\right\} d\boldsymbol{x}$$

$$\propto \exp\left\{-\frac{\tau_V}{2}\boldsymbol{y}'\boldsymbol{y}\right\} \int \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}'\boldsymbol{Q}_{X*}\boldsymbol{x}-2\boldsymbol{y}'\boldsymbol{x}\right)\right\} d\boldsymbol{x}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\tau_V\boldsymbol{y}'\boldsymbol{y}-\tau_V^2\boldsymbol{y}'\boldsymbol{Q}_{X*}^{-1}\boldsymbol{y}\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\boldsymbol{y}'\left(\tau_V\boldsymbol{I_n}-\tau_V^2\boldsymbol{Q}_{X*}^{-1}\right)\boldsymbol{y}\right\}, \tag{A.5}$$

with,

$$\boldsymbol{Q}_{X*} = (\tau_V+\tau_x)\boldsymbol{I_n} + \left(\frac{\tau_0}{\tau_0+\tau_W}\tau_W\right)\boldsymbol{R}. \tag{A.6}$$

Comparing Eqs. (A.3) and (A.4) with (A.5) and (A.6) we note that the impact of the augmented model framework on the joint distribution of $\boldsymbol{y}$ given the hyperparameters is due to two factors in matrix $\boldsymbol{Q}_{X*}$. 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 $\boldsymbol{Q}_{X*}$ converges to $\boldsymbol{Q}_X$ as $\tau_0 \to \infty$ and $\tau_x \to 0$, which is achieved in practice letting τ_0 and τ_x to assume a priori large and small fixed values, respectively, as shown in Section 3.1.1.

Further, we can establish criteria to set the values of these fixed precisions based on its associated factors in matrix $\boldsymbol{Q}_{X*}$. 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 Section 3.1.1.

Appendix B. R script for the toy example

Let $\boldsymbol{y}$ be a $n \times 1$ vector of observations simulated from the dynamic linear model described in Section 3.1. The following R code implements the augmented approach to this model using the INLA library:

B.1. Using current INLA modeling options

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

B.2. Using the proposed augmented approach

```
## building the augmented model
m <- n-1
Y <- matrix(NA, n+m, 2)
Y[1:n,    1] <- y               # actual observations
Y[1:m + n, 2] <- 0             # faked observations
## indices for the INLA library
i  <- c(1:n, 2:n)              # indices for 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 j
l  <- c(rep(NA,n), 2:n)        # indices for 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)))
```

Appendix C. Supplementary data

Supplementary material related to this article can be found online at doi:10.1016/j.csda.2011.10.024.

References

Anderson, , Moore, J.B., 1979. Optimal Filtering. Prentice-Hall, Englewood Cliffs, NJ.
Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society Series B 72, 269–342.
Andrieu, C., Doucet, A., Singh, S.S., Tadić, V.B., 2004. Particle methods for change detection, system identification, and control. Proceedings of the IEEE 92, 423–438.
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., 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., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.
Carter, C.K., Kohn, R., 1996. Markov chain Monte Carlo in conditionally Gaussian state space models. Biometrika 83, 589–601.
Chan, J.C.C., Jeliazkov, I., 2009. Efficient simulation and integrated likelihood estimation in state space models. International Journal of Mathematical Modelling and Numerical Optimisation 1, 101–120.
Dethlefsen, C., 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., 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., Koopman, S.J., 2001. Time Series Analysis by State Space Methods. Oxford University Press.
Ehlers, R.S., 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., 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., 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., Smith, A.F.M., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings—F 140, 107–113.
Hipel, K.W., McLeod, I.A., 1994. Time Series Modeling of Water Resources and Environmental Systems. Elsevier Science Publishers B.V., North-Holland.
Harvey, A.C., 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., Rue, H., 2010. Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In: Tutz, G., Kneib, T. (Eds.), Statistical Modelling and Regression Structures—Festschrift in Honour of Ludwig Fahrmeir. Physica-Verlag, Heidelberg, pp. 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., Doornik, J.A., 1999. Statistical algorithms for models in state space using SsfPack 2.2. Econometrics Journal 2, 113–166.
Luethi, D., Erb, P., Otziger, S., 2009. FKF: fast Kalman filter. R package version 0.1.0. http://CRAN.R-project.org/package=FKF.
Martino, S., 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., Ferreira, M.A.R., 2005. Dynamic models. In: Dey, D.K., Rao, C.R. (Eds.), Handbook of Statistics, vol. 25. Elsevier, North-Holland, pp. 553–588.
Petris, G., 2010. An R package for dynamic linear models. Journal of Statistical Software 36, 1–16.
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., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6, 7–11.

Pole, A., West, M., 1990. Efficient Bayesian learning in non-linear dynamic models. Journal of Forecasting 9, 119–136.

Pole, A., West, M., Harrison, J., 1994. Applied Bayesian Forecasting and Time Series Analysis. Chapman & Hall, New York.

Reis, E.A., Salazar, E., Gamerman, D., 2006. Comparison of sampling schemes for dynamic linear models. International Statistical Review 74, 203–214.

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.

Riebler, A., Held, L., Rue, H., 2011. Estimation and extrapolation of time trends in registry data—Borrowing strength from related populations. ArXiv e-prints, Statistics—Applications, arXiv:1108.0606v1 [stat.AP].

Riebler, A., Held, L., Rue, H., Bopp, M., 2012. 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 175, doi:10.1111/j.1467-985X.2011.01013.x.

Ripley, B.D., 2002. Time series in R 1.5.0. R News 2, 2–7.

Roos, M., Held, L., 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. Bayesian Analysis 6, 259–278.

Rue, H., 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., 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.

Ruiz-Cárdenas, R., Krainski, E.T., Rue, H., 2010. Fitting dynamic models using integrated nested Laplace approximations—INLA. Preprint Statistics No 12/2010, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. URL: http://www.math.ntnu.no/inla/r-inla.org/papers/Ruiz_Krainski_Rue_TR_20111023.pdf.

Schrödle, B., Held, L., Riebler, A., 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., 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 Transactions on Signal Processing 50, 281–289.

Vivar, J.C., Ferreira, M.A.R., 2009. Spatiotemporal models for Gaussian areal data. Journal of Computational and Graphical Statistics 18, 658–674.

West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models, second ed. Springer, New York.

West, M., Harrison, P.J., 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., Pole, A., 1988. Bats – a user guide – Bayesian analysis of time series – release 1.3 – June 1988. University of Warwick.

Whiteley, N., Andrieu, C., 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, p. 27.

Zoeter, O., Heskes, T., 2006. Deterministic approximate inference techniques for conditionally Gaussian state space models. Statistics and Computing 16, 279–292.