Spatial Statistics 21 (2017) 440–459

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

A comparison of estimation methods for multilevel models of spatially structured data

Roger Bivand[a,*], Zhe Sha[b], Liv Osland[c], Ingrid Sandvig Thorsen[d,e]

[a] *Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway*
[b] *University of Oxford, 24-29 St Giles', OX13LB, Oxford, UK*
[c] *Stord/Haugesund University College, Bjørnsonsgt. 45, 5528 Haugesund, Norway*
[d] *Department of Mathematics, University of Bergen, Thormøhlensgate 41A, 5006 Bergen, Norway*
[e] *Analysetjenester AS, Thormøhlensgate 41A, 5006 Bergen, Norway*

A R T I C L E I N F O

Article history:
Received 24 September 2016
Accepted 9 January 2017
Available online 17 January 2017

Keywords:
Spatially structured random effects
Spatial multilevel models
Estimation of CAR and SAR models

A B S T R A C T

Two recent contributions (Dong et al., 2015; Osland et al., 2016) point to the relevance of multilevel models for spatially structured data. In Osland et al. (2016) these models are used to examine the importance of district-level covariates for house prices in Stavanger, Norway, in Dong et al. (2015) similarly for land parcel prices in Beijing; we use these data sets in our comparison. In Osland et al. (2016), a district-level spatial random effect was fitted using an intrinsic CAR model estimated using WinBUGS. Dong et al. (2015) used R code provided in supplementary materials to their article, and subsequently improved in an R package (Dong et al., 2016a); computation there used custom MCMC C++ code to fit a SAR district-level spatial random effect. This article compares approaches to estimating models of this kind, using the R packages R2WinBUGS, HSAR, INLA, R2BayesX, hglm and the new package mclcar for Monte Carlo maximum likelihood estimation (Sha, 2016b). We show that multilevel models of spatially structured data may be estimated readily using a variety of approaches, not only the intrinsic CAR model more typically found in the existing literature. We also point to a range of issues for further research in

* Corresponding author.
E-mail addresses: roger.bivand@nhh.no (R. Bivand), zhesha1006@gmail.com (Z. Sha), liv.osland@hsh.no (L. Osland), ingrid.s.thorsen@analysetjenester.no (I.S. Thorsen).

http://dx.doi.org/10.1016/j.spasta.2017.01.002
2211-6753/

situations in which data acquired at different levels of spatial resolution are combined.

1. Introduction

Spatial data are often organized in multilevel structures, and can be characterized by a mixture of two important features: spatial autocorrelation and spatial heterogeneity (Anselin, 1988). For example, when analysing housing markets, individual houses are located in neighbourhoods which are located within census tracts, which are situated in even higher or more aggregated levels such as municipalities, counties and regions. Typical feature of multilevel data are correlations between observations located within one of the spatial zones or defined neighbourhoods. Houses and, hence, housing prices within a neighbourhood may share similar features, inter alia because they are built at the same time, or because there may be a relatively large proportion of wealthy people living there, or because they are located close to environmental amenities or disamenities. The result could be within-zonal correlations at various levels. So when using geographical data, it could be unrealistic to assume that observations within delimited spatial zones are independent. Following, Corrado and Fingleton (2012), analysts should consider using multilevel models in these situations.

Similarly to spatial regression models, but in contrast with the ordinary least squares estimator, multilevel estimators are not based on the assumption that observations are independent. The multilevel models account for unexplained local correlations, by introducing for instance a random intercept, one for each local zone. Frequently, these random intercepts are not the main focus of the research, given that they capture the impact of excluded factors found in the zones. However, by studying the significance, sign and magnitude of the random effects, they may provide important or useful information in the modelling process. This is explained in for instance Rabe-Hesketh and Skrondal (2008) or Osland et al. (2016). We choose not to examine an alternative random slope approach to these models here (Rabe-Hesketh and Skrondal, 2008), but note that a very recent paper by Dong et al. (2016b) takes up the equivalent spatial random slope approach.

In addition to inducing correlations within zones, variants of multilevel models as described in e.g. Rabe-Hesketh and Skrondal (2008), capture unobserved spatial heterogeneity which is a common feature in spatial data. By way of example, coefficients related to housing market structures may vary between the most urbanized area and semi-urban areas. In this way, classical multilevel models are useful in that they via estimation of the variance of random effects also account for the interzonal variations in the data.

One important assumption in the conventional multilevel models is that there is independence between the zonal random effects (Rabe-Hesketh and Skrondal, 2008, p. 61). For many types of data and analyses, this may be a reasonable assumption. For instance, when studying what determines birthweight of children, one may find that birthweights of children of the same mother are correlated. Birthweight of children of different mothers which represents the group level can probably be assumed to be uncorrelated. Another frequently used example is related to academic results in schools: it could be reasonable to assume that there are correlations or more similar results for pupils within the same school. It may also be a reasonable assumption that correlations of educational results between different schools could be ignored. So in these situations the classical multilevel models could be a useful approach, and may provide improved efficiency of estimates at the lower observational level (Dong et al., 2015).

The conventional random effect estimator may not be efficient when analysing spatial data, however. For this type of data it could be highly relevant to account for correlations between random effects located in close geographic proximity (Osland et al., 2016). When studying regional housing prices, for instance, urban residential neighbourhoods may have many similar features. As we move further away from these neighbourhoods, the zones may more or less gradually change character, from being highly urban to being more semi-urban and even rural. Zones located close to each other, will thus be more similar compared to zones located further away. In this way, there may exist correlations between the higher level zones located nearby, similar to what we may find at the individual observational data level in the traditional spatial regression models.

The new class of models termed Hierarchical Spatial Autoregressive (HSAR) is introduced and developed in contributions by Dong and Harris (2015) and Dong et al. (2015). Their proposed model adds an observation level spatial autoregressive component to the model, the spatially lagged response Wy, where y is a vector of n observations on the response variable, and W is a spatial weights matrix. The HSAR model may also be estimated without this component. In the following, Wy will be omitted to maintain comparability with Osland et al. (2016), although HSAR fits a simultaneous autoregressive (SAR) spatial random effect rather than a conditional autoregressive (CAR) spatial random effect. Both Dong et al. (2015) and Osland et al. (2016) analyse housing data in a very similar format, with some covariates observed for each individual-level observation, and some others observed only at the aggregated, district level. Both employ district level spatial random effects.

It is worth noting that similar approaches to spatial multilevel models have been used by Savitz and Raudenbush (2009) in studying what they term as collective efficacy – the fusion of social cohesion and informal social control – in Chicago neighbourhoods, and by Pierewan and Tampubolon (2014) in examining well-being across regions in Europe. Spatial multilevel approaches in regional and housing economics are inviting, but do not seem to have been applied commonly before Dong et al. (2015) and Osland et al. (2016), apart from Gelfand et al. (2007) and Liu and Roberts (2012); other exceptions may exist in a very dispersed literature. In other research domains, work is documented in, for example, the Munich (Fahrmeir and Lang, 2001; Rue and Held, 2005) and Zambia (Kandala et al., 2001) data sets included in several R packages (Umlauf et al., 2015b; Rue et al., 2015). Both of these examples are for non-Gaussian responses, and many of the estimation methods discussed below extend to discrete responses (see Table 1). In an associated but separate discussion, Bivand (2016) uses weighted aggregation to the district level in the case of a data set in which the estimate of the coefficient of a district level covariate is of central concern.

In this contribution, we examine the software implementations available for estimating district-level multilevel spatially structured random effects models of the kind described by Dong et al. (2015) and Osland et al. (2016); we use the data sets and variables from these articles below. The intrinsic CAR model, used by Osland et al. (2016) and common in disease mapping, may be considered as a standard CAR model with a constrained process coefficient set to its maximum. We will not only look at implementations of the intrinsic CAR model, but also at alternatives such as HSAR and multilevel CAR models in which the autoregressive coefficient is estimated. Comparisons will be carried out using scatterplot matrices of the estimated spatial random effects, tables of the estimated fixed effects and "caterpillar" plots (Spiegelhalter et al., 2003, p. 28). These plots of random effects sorted by mean value and shown with 95% error bars are useful for presenting condensed displays of estimates of the posterior marginal variances; Blangiardo and Cameletti (2015) recommend that these model outputs be shown in detail.

Our approach is first to apply the chosen estimation methods and implementations to the subset of data from the Beijing land parcel data set used in Dong et al. (2015) and published in the R HSAR package (Dong et al., 2016a), as this data set is available for readers to replicate using a script provided as supplementary material (see Appendix A). Next, we apply the same methods to the Stavanger data set used in Osland et al. (2016), but where the house price data have confidential status. Most of the methods used will be described briefly here, as they are well-documented in the literature, but special attention will be given to the Monte Carlo likelihood approach to estimating multilevel CAR models (Sha, 2016a,b).

2. Estimation methods

Our approach in this comparison is to advance the methodological proposals made by Osland et al. (2016) and Dong et al. (2015). If we have a multilevel data structure, and the upper level entities (districts, zones) contain at least one lower level observation, we can attempt to apply a multilevel model in some specification. We will first describe the general class of models that we will

Table 1
Software implementations of multilevel models; iid are aspatial only, the remainder provide various spatially structured random effects (Gelman et al., 2015; Umlauf et al., 2015b; Rue et al., 2015; Dong et al., 2016a; Sha, 2016b; Alam et al., 2015b; Bates et al., 2016; Dorie, 2015). The family column in the table shows the type of response accommodated by the functions; mclcar accepts discrete responses for single level spatial models, but currently not for multilevel spatial models.

Package	Function	Description	Family
R2WinBUGS	bugs	iid, intrinsic CAR	Gaussian, discrete
R2BayesX	bayes	iid, intrinsic CAR	Gaussian, discrete
INLA	inla	iid, intrinsic CAR	Gaussian, discrete
HSAR	hsar	parametric SAR	Gaussian
mclcar	OptimMCL.HCAR	parametric CAR	Gaussian
hglm	hglm	iid, parametric CAR & SAR	Gaussian, discrete
lme4	lmer	iid	Gaussian, discrete
blme	blmer	iid	Gaussian, discrete

be considering, before going on to examine implementation issues which will mostly be covered by references to the literature. Note that while the development in Dong et al. (2015) and more generally in spatial econometrics has been largely restricted to Gaussian responses, many of the models and implementations extend readily to discrete responses, typical for disease mapping.

It is also sensible to explain here how we propose to use the terms hierarchical and multilevel. Unless others have used hierarchical themselves for multilevel data structures, we will call data structures and models with multiple nested levels "multilevel", and try to reserve "hierarchical" for nested model structures in which the model components are linked in sequence, typified in for example a WinBUGS model description file (see Banerjee et al., 2004). The terms are to some extent interchangeable, and models for multilevel data are often hierarchical, but calling these "hierarchical multilevel" models to distinguish them from "hierarchical single-level" models seems unnecessarily complex.

The multilevel conditional autoregressive (CAR) model with a Gaussian response can be written in its most general form as follows (see for example Gelfand, 2010, p. 533):

$$\begin{aligned} y &= X\beta + Z\gamma + \Delta\eta + \epsilon \\ \gamma &\sim N(0, I_K\sigma_\gamma^2) \\ \eta &\sim N(0, (I_K - \lambda M)^{-1}\sigma_\eta^2) \\ \epsilon &\sim N(0, (I_n - \rho W)^{-1}\sigma_\epsilon^2) \end{aligned} \tag{1}$$

where K is the number of districts or zones, n is the number of individual spatial observations, X, Z, Δ are given design matrices, and W, M are given spatial weight matrices, η, γ are the district multilevel spatial and non-spatial (independent and identically distributed—iid) random effects, ϵ are the individual observation spatial random effects that collapse to residuals when $\rho = 0$, and $\rho, \lambda, \beta, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_\gamma^2$ are parameters to be estimated. When the higher level variables include the intercept only, then $Z = \Delta$ and the term $Z\gamma$ is absorbed into the fixed effect intercept.

In both the considered data sets below, the higher level variables values are repeated for each individual observation belonging to each zone. If we remove both spatial random effects by setting $\rho = \lambda = 0$, we obtain the conventional iid random effects model, with random effects at the district level. Note that the model described by Eq. (1) differs from the HSAR model developed by Dong et al. (2015), as they do not consider an individual, lower level spatial random effect expressed as ϵ, but rather choose to model a spatially autoregressive response, replacing y by $(I - \rho W)y$ and $\epsilon \sim N(0, I_n\sigma_\epsilon^2)$. This introduces further important issues of interpretation, as discussed in detail by LeSage and Pace (2009). In addition, they consider a simultaneous autoregressive version of the spatial random effects η:

$$\eta \sim N(0, \sigma_\eta^2((I_K - \lambda M)^{-1}(I_K - \lambda M^T)^{-1}).$$

If there are only the spatial random effects, the above model in Eq. (1) simplifies to:

$$\begin{aligned} y &= X\beta + \Delta\eta + \epsilon \\ \eta &\sim N(0, (I_K - \lambda M)^{-1}\sigma_\eta^2) \\ \epsilon &\sim N(0, (I_n - \rho W)^{-1}\sigma_\epsilon^2). \end{aligned} \tag{2}$$

If the errors at the individual level are iid instead of spatially correlated, the model can be simplified to:

$$\begin{aligned} y &= X\beta + \Delta\eta + \epsilon \\ \eta &\sim N(0, (I_K - \lambda M)^{-1}\sigma_\eta^2) \\ \epsilon &\sim N(0, I_n\sigma_\epsilon^2). \end{aligned} \tag{3}$$

In the following we focus on the multilevel CAR model defined in Eq. (3) that considers spatial correlation at the district level only. The response follows a multivariate normal distribution:

$$y \sim N(X\beta, \qquad \Delta Q_\eta^{-1}\Delta^T + Q_e^{-1}) \tag{4}$$

where $Q_\eta = (I_K - \lambda M)/\sigma_\eta^2)$ and $Q_e = I_n/\sigma_\epsilon^2$ are the precision matrices of η and ϵ.

The likelihood of this model can be computed directly but the covariance and precision matrices are usually dense by construction and thus evaluation can be computationally demanding especially when the number of spatial units is large. Meanwhile, we can also write the conditional distribution of y given the random effects η:

$$y \mid \eta \sim N(X\beta + \Delta\eta,\ Q_e^{-1}). \tag{5}$$

So the likelihood function can be viewed as the integral of the random effects:

$$L(\theta, y) = \int f_\theta(y \mid \eta) f_\theta(\eta) d\eta \tag{6}$$

where $f_\theta(y \mid \eta)$ and $f_\theta(\eta)$ are the probability density functions for $y \mid \eta$ and η indexed by parameter θ. Both $y \mid \eta$ and η have sparse precision matrices, which can be useful in simulation and computing.

2.1. Multilevel models

As Dong et al. (2015) point out, it is entirely possible to use a standard linear mixed effects model with an iid random effects term η at the upper, district level, rather than taking η as a spatially structured random effect at the district level. Indeed, if the specified model of η includes a parameter λ taking a value close to zero, the spatially structured random effect will collapse to an iid random effect. Because the typically applied intrinsic CAR models for η assume that λ is at its upper bound, that is that strong autocorrelation is present, it is often informative to compare the random effects output of an iid random effect and a spatially structured random effect. It may very well be the case that the iid representation, with implied strong intra-district dependence between the lower-level observations, but little inter-district dependence, does represent the data generation process quite adequately.

The chief advantage of iid random effects is that they are very well studied in the literature. Very recent examples include Bates et al. (2015), which includes extensive references to the field. It also shows how the R lme4 package (Bates et al., 2016) may be used to estimate the presented models, where the upper level iid random error linear mixed-effects model is among the simplest. Its use will be compared with a Bayesian version described in Chung et al. (2013) and provided by the R blme package (Dorie, 2015). An alternative implementation for hierarchical generalized linear models is given by Rönnegård et al. (2010) in R package hglm (Alam et al., 2015b). Two further implementations are furnished by INLA (Rue et al., 2015) and BayesX (Umlauf et al., 2015b) through their R interfaces; we could have added many others.

2.2. Spatial multilevel models

In fitting the multilevel model described by Eq. (3), we can choose to follow Osland et al. (2016) and use the "car.normal()" specification in WinBUGS (Thomas et al., 2004; Spiegelhalter et al., 2003). Because WinBUGS can be interfaced from R using the R2WinBUGS package (Gelman et al., 2015) described by Sturtz et al. (2005), automating the writing of the data, initial values and script files which may conveniently be merged into the workflow. Osland et al. (2016) chose to use the "car.normal()"

specification, which is for an intrinsic CAR model (Besag, 1974; Besag and Kooperberg, 1995). An alternative Markov chain Monte Carlo (MCMC) implementation of the intrinsic CAR model, termed the "mrf" Markov Random Field smoother and implemented in the mgcv package (Wood, 2006, 2016) is provided in the R R2BayesX package (Umlauf et al., 2015b), which interfaces BayesX as described in Umlauf et al. (2015a). Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation (INLA) is a viable alternative to MCMC, and may be used for a widening range of spatial models (Gómez-Rubio et al., 2014; Bivand et al., 2014; Lindgren and Rue, 2015; Bivand et al., 2015; Blangiardo and Cameletti, 2015). Using the INLA package for R (Rue et al., 2015) gives a choice of spatial models, where the "besag" intrinsic CAR specification matches that chosen by Osland et al. (2016). It is also possible to estimate a Besag–York–Mollie (BYM) specification including both an iid random effect and an intrinsic CAR spatial random effect (Besag et al., 1991), and this will be presented briefly in the comparison below. Care in setting up the estimation of the BYM model is required, as shown convincingly by Gerber and Furrer (2015), who point to potential serious practical failures in the implementation of the model. It would be attractive to use the R CARBayes package offering a number of alternative specifications, but it does not at present support the indexing of the spatial process to a district level (Lee, 2013, 2016).

Before going on to describe the Monte Carlo maximum likelihood estimation approach, we will mention our use of the R package hglm (Alam et al., 2015b) for fitting multilevel spatial model specifications. Alam et al. (2015a), based on Rönnegård et al. (2010), show that the hierarchical generalized linear model approach can also be used. Finally, the HSAR approach (Dong et al., 2015; Dong and Harris, 2015) will be compared with the other model fitting implementations, especially to establish a baseline using the Beijing land parcel data set distributed with the package (Dong et al., 2016a).

2.3. Monte Carlo maximum likelihood estimation

The mclcar approach (Sha, 2016a) approximates the log-likelihood ratio for a given parameter value ψ and any parameter value θ by importance sampling (see Evans and Swartz, 2000). Data from the given CAR model can be simulated with less cost by using the sparsity of the precision matrix.

Denote all the parameters of interest by $\theta = (\rho, \lambda, \sigma^2_\epsilon, \sigma^2_u, \beta)^T$. For a given ψ value, the likelihood ratio can be written as

$$L = \frac{L(\theta)}{L(\psi)} = \int \frac{f_\theta(y, \eta)}{f_\psi(y, \eta)} f_\psi(\eta \mid y) d\eta \tag{7}$$

where f_θ denotes in general the probability density function indexed by parameter θ.

The Monte Carlo approximation to the above integral is

$$\hat{L} = \frac{1}{s} \sum_{i=1}^{s} \frac{f_\theta(y, \eta_i)}{f_\psi(y, \eta_i)}, \quad \eta_i \sim f_\psi(\eta \mid y) \tag{8}$$

where s is the Monte Carlo sample size.

The log-likelihood ratio for two different parameter values θ and ψ can be approximated by $\hat{\ell} = \log \hat{L}$. This is called the Monte Carlo likelihood (MCL) (Geyer and Thompson, 1992). By maximizing the MCL $\hat{\ell}$ we can get the Monte Carlo Maximum Likelihood Estimator (MC-MLE) (Geyer and Thompson, 1992). The evaluation of the Monte Carlo likelihood requires simulating samples from $f_\psi(\eta \mid y)$ and calculating the ratio of $\frac{f_\theta(y,\eta_i)}{f_\psi(y,\eta_i)}$. Convergence of the MCL, MC-MLE and corresponding level sets to the exact log-likelihood values is discussed in Geyer (1994).

The MCL-MLE can be obtained by directly maximizing the MCL but the variance of MCL increases as θ moves away from ψ. Therefore, it is recommended to monitor the Monte Carlo variance in the maximization procedure. The variance of the MCL can be approximated by

$$\hat{V}(\hat{\ell}) = \frac{\sum_{i}^{s} \mathrm{Var}(l_i/\hat{L})}{s\hat{L}} \tag{9}$$

where $l_i = f_\theta(y, \eta_i)/f_\psi(y, \eta_i)$ and s is the Monte Carlo sample size (Sha, 2016a). Note that Cappé et al. (2002) have shown that for a model with latent variables with an arbitrary ψ the convergence rate of MCL will depend on the data size exponentially and for a ψ carefully chosen to be a consistent estimator, the convergence rate of MCL depends on the data size linearly. So in short, the convergence rate of MCL for model with latent variables is always affected by the data size.

Similarly, when obtaining the MC-MLE, the asymptotic variance of MC-MLE can be approximated by

$$\hat{V}(\hat{\theta}_{MC-MLE}) = H^{-1}AH^{-1} \tag{10}$$

where H is the approximated Hessian at the MC-MLE and $A = \hat{V}(\nabla\hat{\ell})$ (Geyer, 1994; Sha, 2016a), which can also be approximated by the Monte Carlo samples.

Instead of using the maximum of the MCL from a single importance sampler parameter value ψ_0, an iterative procedure (Sha, 2016a) can be used to obtain sequences of ψ and $\hat{\theta}_{MC-MLE}$ values, and thus reducing the Monte Carlo variance of the approximation.

The procedure can start from an arbitrary value for ψ_0, then at the kth iteration, it maximizes $\hat{\ell}_{\psi_k}$ to get $\hat{\theta}^{(k)}_{MC-MLE}$ and then in the $k+1$th iteration, it uses $\hat{\theta}^{(k)}_{MC-MLE}$ as the value for ψ_{k+1}.

To prevent the maximization at each step from overshooting and missing the target maximum, the procedure combines the estimated Monte Carlo variance in updating ψ. At the end of each maximization, if the estimated variance of $\hat{\theta}^{(k)}_{MC-MLE}$ is larger than some given bound, say $b_1 = 1/s$, then we update ψ_{k+1} to be

$$\psi_{k+1} = (1 - b_2)\psi_k + b_2\hat{\theta}^{(k)}_{MC-MLE} \tag{11}$$

where b_2 is a parameter that can be tuned by the user to control the step size in the convex optimization; otherwise, we can simply proceed as described previously. The procedure iterates until the change between two consecutive values of the MC-MLEs become smaller than some given tolerance.

3. Model estimation results

We move now to consider the use of software implementations listed above and in Table 1 to obtain estimators for multilevel spatial models. The implementations we have used include iid random effects at the district level, and SAR or CAR spatially structured random effects at the district level. Finally two implementations of the Besag–York–Mollie model with both an iid random effect and an intrinsic CAR spatially structured random effect will also be used. Results from a selection of these implementations will be presented in this section. The comparisons will be shown in figures presenting scatterplot matrices of the estimated random effects, and collections of "caterpillar" plots of random effects sorted by size and shown with 95% error bars. Presenting these comparisons using scalar summary measures has turned out to be challenging; for example, within one software implementation, the values of the Deviance Information Criterion (DIC) are comparable, but between the implementations it may vary in representation. Tables of coefficient estimates across implementations will be included, omitting the coefficients of year dummies. We have chosen to use standard or default vague priors and initial values for estimation methods using these; we believe that these should not affect the posterior results too much, and in any case this is not the focus of this comparison.

3.1. Beijing land parcel data set

The subset of the Beijing land parcel data set used here is contained in the R package HSAR (Dong et al., 2016a), using data objects `landprice`, `landSPDF` and `Beijingdistricts`. It differs slightly from the data distributed as supplementary material and analysed in Dong et al. (2015), because a small number of covariates are omitted, and the spatial objects are provided with coordinate reference systems. Consequently, results shown here do not correspond completely with those in the published article, but may be reproduced easily using the package in R; the variables used here are described in Table 2. The data set is for 1117 residential land parcels in Beijing for which prices were available at a

Table 2
Beijing land parcel data set: included variables (see Dong et al., 2015, p. 12, Table 1).

Variable	Level	Description
log (price)	Land parcel	Leasing price per m^2
log (area)	Land parcel	Area (m^2))
log (distance to CBD)	Land parcel	Distance to city centre (m)
log (distance to school)	Land parcel	Distance to nearest school (m)
log (distance to park)	Land parcel	Distance to nearest park (m)
log (distance to subway)	Land parcel	Distance to nearest subway stop (m)
year dummies	Land parcel	Observation year
Crime rate	District	Reported serious crimes per 1000 inhabitants
Pop. density	District	Population density (1000/km^2)

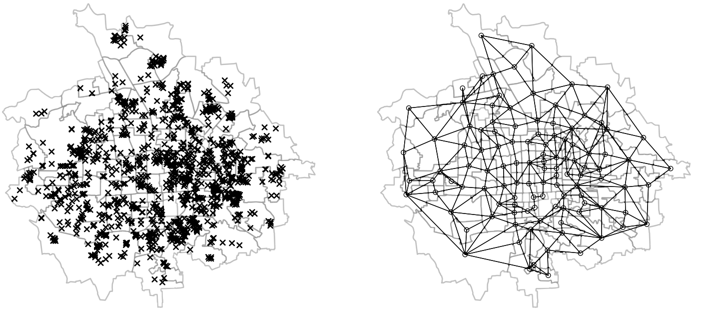

Fig. 1. Beijing districts: left panel: location of land parcel centroids in districts; right panel: contiguity neighbour (rook) relationships between districts.

year in between 2003 and 2009 (year dummies are included in the model formula; most observations are for 2003). The response is the logarithm of the leasing price per square metre; covariates at the land parcel level include the logarithms of distances to the nearest subway, elementary school, green park, to central Beijing, and the logarithm of the area of the parcel. There are 111 districts into which the 1117 parcels are nested, as shown in Fig. 1. Two covariates are provided at the district level because they are not known at the land parcel level; the population density in thousands of persons per square kilometre, and the crime rate (number of serious crimes reported per thousand persons). Some central districts are omitted from the study because they do not contain residential land parcels for which price data were available.

Table 3 shows the estimates of the coefficients, omitting the intercept and the coefficients related to the year dummies; in comparison with Table 2 in Dong et al. (2015)[p. 13], their variables log (distance to river) and buildings from before 1949 are omitted in the HSAR package data set. In addition, we have used binary, not row-standardized, spatial weights for better comparison with the CAR models for which row-standardized weights are unsuitable, and have not fitted the HSAR model including the spatial lag of the land parcel level prices. These steps leave us with three multilevel intrinsic CAR models (BayesX, INLA and WinBUGS), two multilevel SAR models (HGLM and HSAR), and one multilevel CAR model. As we can see from Table 3, the coefficient and standard error estimates are very consistent within the group of intrinsic CAR models, but differ from the other model fitting procedures to some extent. The two multilevel SAR models, HGLM SAR and HSAR, are very similar, with few differences in coefficient values from the intrinsic CAR estimators. The implementation of the Monte Carlo likelihood estimator differs a little more in a few variables, and rather more in the coefficient standard errors. Using the Monte Carlo likelihood estimator would change inferences on the covariates, but possibly more in a borderline rather than a radical way. For this data set, the MCMC estimators had the following output sample counts: HSAR 4500, BayesX 5000, WinBUGS 3000; with the MCL CAR sample count set to 500. Increasing the MCL CAR sample count would reduce the observed divergences. Convergence of the BayesX and WinBUGS MCMC chains was checked and found unproblematic (Gelman, 1996; Banerjee et al., 2004).

Table 3
Beijing districts: estimates of coefficients for six implementations of multilevel spatial regression models (standard errors in parentheses).

	BayesX	INLA	WinBUGS	HSAR	HGLM SAR	MCL CAR
log (area)	−0.0249	−0.0251	−0.0259	−0.0270	−0.0266	−0.0292
	(0.0184)	(0.0186)	(0.0182)	(0.0189)	(0.0187)	(0.0204)
log (distance to CBD)	−0.4113	−0.4098	−0.4103	−0.3935	−0.3947	−0.3047
	(0.1064)	(0.1061)	(0.0973)	(0.0929)	(0.0933)	(0.1624)
log (distance to school)	−0.0185	−0.0188	−0.0193	−0.0310	−0.0301	−0.0434
	(0.0390)	(0.0391)	(0.0387)	(0.0388)	(0.0385)	(0.0436)
log (distance to park)	−0.1787	−0.1799	−0.1806	−0.1927	−0.1862	−0.2505
	(0.0610)	(0.0613)	(0.0601)	(0.0614)	(0.0608)	(0.0811)
log (distance to subway)	−0.2155	−0.2160	−0.2152	−0.2062	−0.2063	−0.2123
	(0.0419)	(0.0422)	(0.0411)	(0.0422)	(0.0418)	(0.0519)
Crime rate	0.0036	0.0037	0.0032	0.0045	0.0029	0.0109
	(0.0087)	(0.0082)	(0.0084)	(0.0093)	(0.0094)	(0.0107)
Pop. density	0.0248	0.0249	0.0260	0.0267	0.0258	0.0291
	(0.0143)	(0.0142)	(0.0147)	(0.0143)	(0.0141)	(0.0141)

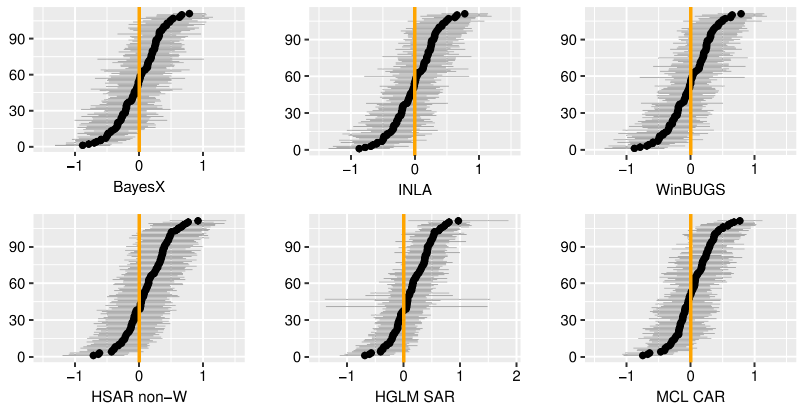

Fig. 2. Beijing districts: "caterpillar" plots of spatially structured random effects for six estimators.

While the tabular results do not show any issues with the data, the HGLM CAR model would not converge, and the equivalent HGLM SAR model warned that the land parcel observation 888, which has a very high crime rate, and the highest population density in the whole data set, deserved attention. The values of the response of log price and the other covariates in the chosen model specification are not unusual. There are only 4 of the 1117 land parcel observations with a population density over 10 (thousand persons per square km), 888 has a value of 28. In addition, this land parcel observation is alone in its district (one of 7 districts only containing a single land parcel observation each among 111 districts in total), thus potentially confusing the random effects estimate as there are no replicate land parcels in these districts.

Fig. 2 shows "caterpillar" plots (Spiegelhalter et al., 2003, p. 28) of the sorted spatially structured random effects, with error bars spanning ± 1.96 standard errors. On the top row, the three intrinsic CAR estimators are very similar in their shape and extent. The HSAR estimator is similar in shape but with longer error bars, while the HGLM SAR and MCL CAR have slightly shorter error bars. The HGLM SAR estimator also finds much larger standard errors for some of the district random errors, but the standard error estimate for district 109, to which land parcel observation 888 belongs, is in the first quartile of estimated values.

Fig. 3 permits us to see how closely correlated are the spatially structured random effects for the three intrinsic CAR estimators. Also note that the close correlation between the HSAR and HGLM SAR estimators is matched in the density shape of the two sets of random effects. For the MCL CAR

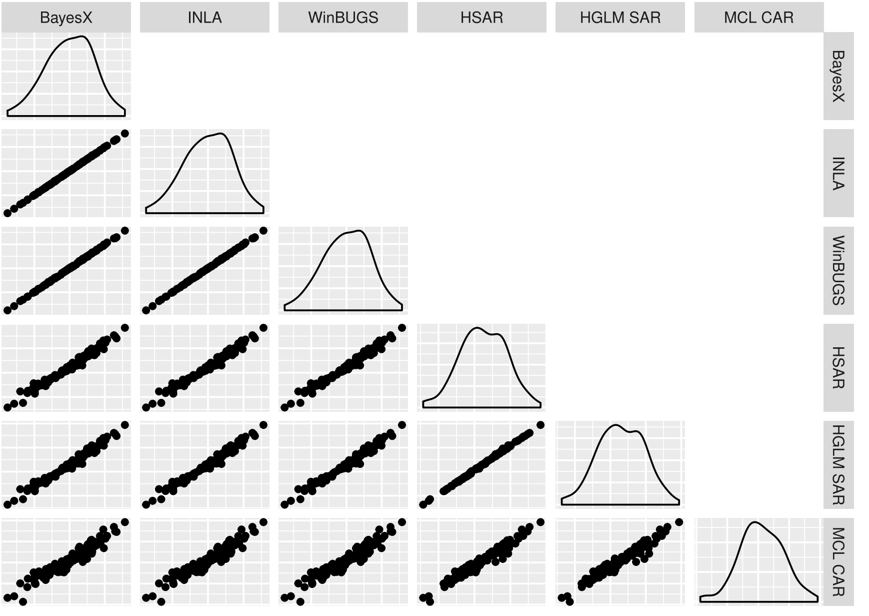

Fig. 3. Beijing districts: scatter plot matrix of spatially structured random effects for six estimators; the display shows a scatterplot for each pair of estimators being compared, with density plots of the spatially structured random effects in the diagonal panels.

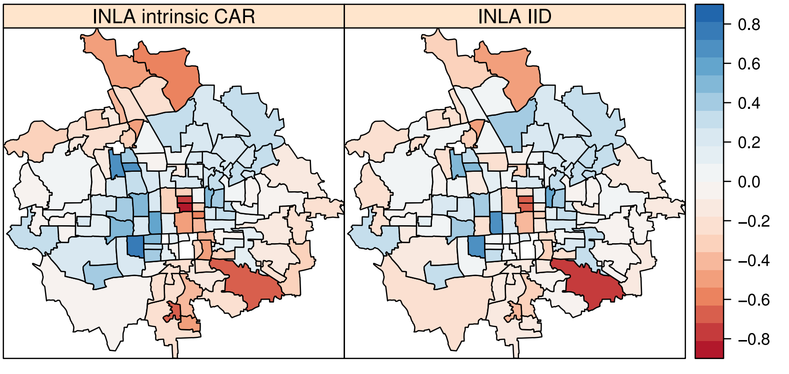

Fig. 4. Beijing districts: random effects estimated using INLA. Left panel: spatial intrinsic CAR random effects. Right panel: non-spatial iid random effects.

estimator, there are somewhat larger differences from the other estimators, but the largest differences between sets of scatterplots appear to be between the intrinsic CAR and the others. Fig. 4 displays the very similar spatial footprints of the spatially structured random effect estimated using the intrinsic CAR model in INLA, and the spatially unstructured iid random effects. In further results not presented here but included in the supplementary materials code, it appears that most of the combined random effects (spatial and non-spatial) estimated in the BYM model (Besag et al., 1991) are captured by the intrinsic CAR model, with the iid component of the BYM model only finding random effects close to zero. It looks as though the intrinsic CAR technique absorbs most of the intra-district iid variability that might more sensibly have been assigned to the non-spatial random effect.

Finally, Table 4 reports the estimates of the variance σ^2_η associated with the random effect, the residual variance σ^2_ϵ and λ, the spatial coefficient for the HGLM SAR, HSAR and MCL CAR estimators.

Table 4
Beijing districts: summary measures of random effects estimates.

	BayesX	INLA	WinBUGS	HSAR	HGLM SAR	MCL CAR
σ_u^2	0.4164 (0.1006)	0.3742	0.3856	0.0884 (0.0241)	0.0772	0.0948 (0.0320)
σ_e^2	0.5835 (0.0256)	0.5814	0.5816	0.5853 (0.0256)	0.5829	0.5690 (0.0248)
λ				0.1365 (0.0198)	0.1499	0.1444 (0.0520)

The scales of σ_η^2 are known to vary between intrinsic CAR and fitted parameter estimators. Intrinsic CAR assumes that the K random effects are actually following a distribution with $K-1$ dimensions, so the variance estimates are always going to be larger than for the others. The estimates of σ_ϵ^2 are largely consistent. The coefficient estimates of λ are similar for the CAR and SAR estimators. The maximum value that the autoregressive parameter can take in the SAR and CAR models is 0.17379, the inverse of the maximum eigenvalue of the district level spatial weights matrix. The estimated coefficient values all lie close to this maximum. In general we would expect the autoregressive parameter in a SAR model to be smaller in absolute value than in a CAR model estimated in the same way because the SAR model effectively uses higher lags smoothing out the strength of the spatial correlation; the results here are for SAR and CAR models fitted using different estimators. There do not seem to be any formal methods for choosing between the intrinsic CAR representation, which assumes that the coefficient is at its maximum, and parametric representations in which the coefficient is estimated rather than taking an assumed value. Use of an assumed value in image restoration, in which pixels arbitrarily tessellate surface values, seems reasonable, but this does not necessarily transfer to entities defined in other ways.

3.2. Stavanger house price data set

We will use the same data and the same hedonic house price model as the one found in Osland et al. (2016). We will, consequently, provide only a short summary of relevant background information here. The study area is the Stavanger region, which is located in a coastal area in the southwestern part of Norway. In many ways the region is an ideal study area for illustration purposes. The region covers 8 municipalities. It is, therefore, a relatively large region which spans over rural, semi-urban and urban landscapes. This feature is important in this paper, given that it creates latent spatial heterogeneity and variation in the data.

Geographically the region is spatially delimited from other regions by mountains and fjords. The urban centre of the region is the city of Stavanger which is the fourth largest city in Norway. Since 1969 the region has been the capital of a booming petroleum industry; and the industry structure particularly in the most urbanized parts is dominated by oil and gas related activities. In the study period, 1997–2007, the growth in employment and population has been larger than for the country as a whole. The income level in the most populous parts of the area is among the highest in Norway. Housing prices have also grown in almost all years of the study period. So compared to an international standard, there are no parts of the studied market that could be described as deprived. In this way, the study area is considered to be highly homogeneous, owner occupancy among households is very high (80%), which leaves us with a relatively transparent housing market to be studied by way of different estimators.

A complete overview of data sources and e.g. descriptive statistics is found in Osland et al. (2016); a summary is provided in Table 5. The study period is 1997–2007, but the data are in essence cross-sectional, with a year dummy included. In all there are 7180 market sales of privately owned single family houses. A hedonic house price model is estimated, in which the dependent variable, the logarithm of housing prices, is described as a function of characteristics related to the house, the location of the house and its surrounding, in addition to accessibility factors. The following variables are characteristics related to the house itself: lot size is measured in square metres, age is the number of years since the house was built, garage is a dummy variable indicating whether the house has a garage or not, living area is the size of the house also measured in square metres, and rural lot size is

Table 5
Stavanger house price data set: variables used (see Osland et al., 2016, p. 915, Table A1); columns headed 1–3 show in which Model variables were included.

Variable	Level	1	2	3	Description
log (real price)	House	x	x	x	Deflated sale price of house
log (lot size)	House	x	x	x	Lot size (m^2)
log (rural lot size)	House	x	x	x	Lot size × rural dummy
log (age)	House	x	x	x	Age of house
log (age)^2	House	x	x	x	Square of age of house
garage	House	x	x	x	Garage dummy
log (living area)	House	x	x	x	Area of house (m^2)
year dummies	House	x	x	x	Year of sale
log (time to CBD)	Zone		x	x	Driving time to city centre (minutes)
log (time to CBD)^2	Zone		x	x	Square of driving time
log (accessibility)	Zone		x	x	Defined in Eq. (12)
log (socio-economic conditions)	Zone			x	Socio-economic conditions index
log (low education)	Zone			x	Share aged 30–39 with $\leq$ 10 years school

the variable lot size interacted with a dummy-variable indicating whether the house is located in one of the four rural municipalities of the region. The second type of variables focuses on the impact of spatial structure characteristics of the overall study area: time to CBD is the travelling time from the zone to which the house belongs and the CBD. The measure is based on driving time by car in minutes. Accessibility is the labour market accessibility of each zone. The variable is defined as a Hansen type of accessibility measure, by (Hansen, 1959):

$$\text{ACCESSIBILITY}_i = \sum_{k=1}^{w} D_k e^{\beta_e d_{ik}}. \tag{12}$$

In the expression above, D_k denotes the number of jobs in destination zone k. The measure is based on the assumption that accessibility of a destination is a decreasing function of distance to other potential destinations. Each destination is weighted by the number of opportunities available in a particular zone. See Osland and Thorsen (2008) for more information on this accessibility measure. Finally, two variables representing the socio-economic conditions in the neighbourhood where house i is located were considered. Socio-economic conditions are an aggregated index calculated on the basis of 18 different socioeconomic indicators, such as rate of unemployment, the proportion of immigrants, the median household income, etc. The lower the number of the index, the better is the socioeconomic conditions in a given location. The low education variable represents the proportion of the population at the age 30–39 years which has primary and lower secondary school as their highest level of education. Osland et al. (2016) estimate three model variants, using different sets of variables as will be explained in the sections to follow.

The data are given at various levels of aggregation. Information on housing prices and housing characteristics is given at the lowest individual observational level. Variables based on distances refer to postal zone level. In all there are 100 postal delivery zones. For the socioeconomic variables, Stavanger is divided into 68 different zones. Outside the Stavanger municipality however, we have information only at the municipality level for these variables. Hence, socioeconomic characteristics are represented by 75 different zones. The random effects are estimated based on the use of 51 different zones in the area. Inside Stavanger we have used the 39 postal delivery zones. Zones outside Stavanger are delimited at the municipality level.

The spatial weights matrix used here differs minimally from that used in the numerical results presented in Osland et al. (2016), and is based on defining neighbours as sharing at least one boundary point (Queen contiguity), edited to add a small number of inhabited islands.

Fig. 5 shows the considerable variation in the density of house sale locations and the sizes of the zones used in analysis. The right panel of the figure zooms into the most urban part of the study area, permitting more of the detail to be seen.

Table 6 shows very similar coefficient and standard error estimates for all five estimators of the first multilevel spatially structured random effects model specification for the Stavanger house price

Fig. 5. Stavanger house locations and zone boundaries. Left panel: house locations (upper) and zone contiguity neighbours (lower). Right panel: house locations (upper) and zone contiguity neighbours (lower) in central Stavanger.

Table 6
Stavanger zones: estimates of Model 1 coefficients for five implementations (including only covariates on individual houses) of multilevel spatial regression models (standard errors in parentheses).

	BayesX	INLA	WinBUGS	HSAR	MCL CAR
log (lot size)	0.1041	0.1041	0.1043	0.1070	0.1078
	(0.0055)	(0.0055)	(0.0055)	(0.0054)	(0.0124)
log (rural lot size)	−0.0697	−0.0693	−0.0682	−0.0822	−0.0991
	(0.0099)	(0.0098)	(0.0098)	(0.0087)	(0.1015)
log (age)	−0.0548	−0.0549	−0.0547	−0.0533	−0.0543
	(0.0101)	(0.0101)	(0.0100)	(0.0101)	(0.0162)
log $(\text{age})^2$	−0.0069	−0.0069	−0.0070	−0.0074	−0.0073
	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0034)
garage	0.0495	0.0495	0.0493	0.0493	0.0488
	(0.0056)	(0.0056)	(0.0056)	(0.0056)	(0.0062)
log (living area)	0.4732	0.4731	0.4732	0.4720	0.4724
	(0.0083)	(0.0082)	(0.0083)	(0.0083)	(0.0090)

data (omitting the estimates for the year dummies and the intercept). The coefficient and standard error values are very similar to those reported for the M1(CAR) specification in Table 1 of Osland et al. (2016)[p. 12 of the early access version]. The HGLM SAR and CAR estimators both failed to estimate this model, and the HGLM multilevel iid estimator reported issues associated with the greater than proportional influence of single observations. Fig. 6 confirms that the spatially structured random effects estimates are effectively identical between the three multilevel intrinsic CAR estimators, BayesX, INLA and WinBUGS. There are however differences between these and the HSAR and MCL CAR estimators. Convergence of the BayesX and WinBUGS MCMC chains for this and subsequent models was checked and found unproblematic (Gelman, 1996; Banerjee et al., 2004).

In particular, the spatially structured random effects of four zones differ between the estimators: 1111 Sokndal (number of houses: 68), 1112 Lund (number of houses: 42), 1114 Bjerkreim (number of houses: 42), and 1122 Gjesdal (number of houses: 308). The numbers of individual house price observations are not unusual for these four zones, as the minimum zone house count in the whole data set is 5, the maximum 1636 (Sandnes municipality), and the median count 70. Note, however, that the variable log (rural lot size) has a structure that may correlate with these four random effect

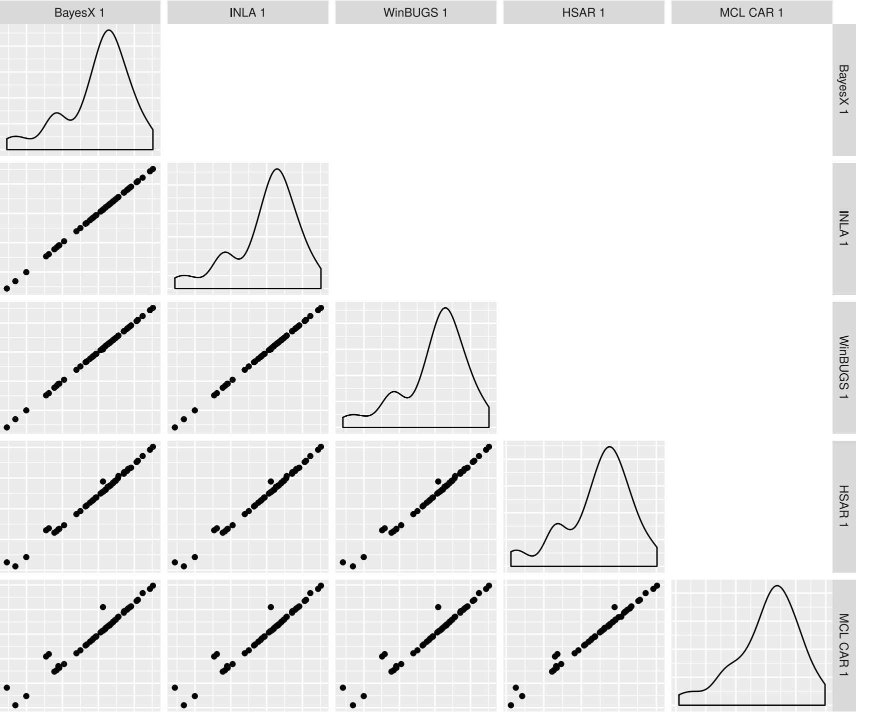

Fig. 6. Stavanger zones: scatter plot matrix of spatially structured random effects for five estimators of Model 1 including only covariates on individual houses; the display shows a scatterplot for each pair of estimators being compared, with density plots of the spatially structured random effects in the diagonal panels.

Table 7
Stavanger zones: summary measures of random effects estimates for Model 1 (including only covariates on individual houses).

	BayesX	INLA	WinBUGS	HSAR	MCL CAR
σ_η^2	0.0585	0.0539	0.0546	0.0193	0.0244
	(0.0137)			(0.0045)	(0.0052)
σ_ϵ^2	0.0419	0.0419	0.0419	0.0419	0.0418
	(0.0007)			(0.0007)	(0.0007)
λ				0.1423	0.1554
				(0.0113)	(0.0085)

estimates. If this variable is omitted, the HGLM SAR model does run to completion for the first model without zonal level covariates, but fails to estimate the two models including zonal level covariates. HSAR and MCL CAR both differ clearly from the intrinsic CAR estimators in the case of these four upper level zones, but HSAR differs less strongly than MCL CAR.

Fig. 7 presents the spatially structured random effects "caterpillar" plots for the five considered estimation methods and the three model specifications found in Osland et al. (2016). Model 2 adds accessibility measures, and Model 3 also includes socio-economic conditions at the zonal level. The point estimates for Model 1 were considered in the discussion of Fig. 6, and will be discussed for Model 3 below. The general conclusion is that the estimators all provide similar estimates, and that both the shapes of the curves and the standard error bar lengths change in the same ways as zonal level covariates are added. The HSAR spatially structured random effects standard errors are somewhat larger than those of the CAR models.

Table 7 reports the estimates of σ_η^2 and σ_ϵ^2 for all five estimators for Model 1. The pattern found for the Beijing districts data set is repeated, with σ_ϵ^2 consistent across all estimators, but where the scaling

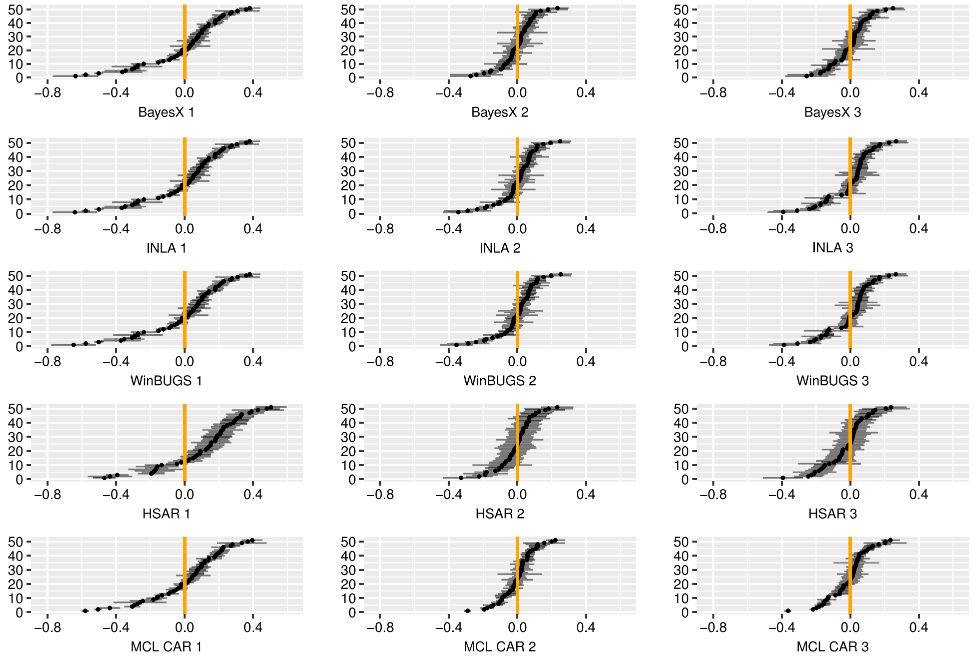

Fig. 7. Stavanger zones: "caterpillar" plots of spatially structured random effects for five estimators and three models; Model 1 includes only covariates on individual houses, Model 2 adds accessibility covariates at the zonal level, and Model 3 further adds socio-economic conditions at the zonal level.

of σ_η^2 varies between the group of intrinsic CAR estimators and the other estimators. The maximum value of the zonal level spatial autoregressive coefficient at the top of its range is 0.159895, so that we can see that both HSAR and MCL CAR lie quite close to this.

Because the results for Models 2 and 3 are qualitatively similar in character, only those for Model 3 will be discussed here. Table 8 shows more variability in coefficient and standard error estimates between the estimators than Table 6. In general, the INLA and WinBUGS estimates are closest to each other throughout, and reproduce the values presented in the M3(CAR) column of Table 1 of Osland et al. (2016, p. 906). For some covariates – the garage dummy, log age squared and log living area – there is very little difference in coefficient estimates. In other cases, coefficient estimates do differ, even between BayesX and the INLA/WinBUGS pair of estimators. In particular, MCL CAR not only fails to reach the same coefficient values, but also reaches larger standard error values for the coefficients for the zonal level covariates, changing the inferences drawn from the Model 3 results as a whole. The degree with which MCL CAR diverges from the other estimators is related to the number of Monte Carlo samples used, with increases in numbers of samples giving smaller discrepancies. For this data set, the MCMC estimators had the same output sample sizes as those used for the Beijing data set, but the MCL CAR sample count was increased to 5000 for this much larger data set. Increasing the MCL CAR sample count even further should reduce the observed divergences (Cappé et al., 2002). It appears that much more of the variability in the log house price is apportioned to the zonal level random effects than the zonal level covariates in this case. All of the estimators apart from INLA and WinBUGS struggle to handle the split between the zonal level covariates and zonal level spatially structured random effects. One reason could be the heterogeneity engendered by the four rural municipalities.

Fig. 8 shows the same picture of good agreement between the INLA and WinBUGS estimates, here for the spatially structured random effects. The divergence between the INLA and WinBUGS pair of estimators and BayesX is unexpected, as they are all intrinsic CAR implementations. The above mentioned four municipalities split HSAR and MCL CAR from INLA and WinBUGS. In addition, BayesX and INLA/WinBUGS are most different with respect to the spatially structured random effects

Table 8
Stavanger zones: estimates of Model 3 coefficients for five implementations (including covariates on individual houses and zonal covariates) of multilevel spatial regression models (standard errors in parentheses).

	BayesX	INLA	WinBUGS	HSAR	MCL CAR
log (lot size)	0.1048	0.1045	0.1046	0.1080	0.1082
	(0.0052)	(0.0052)	(0.0053)	(0.0051)	(0.0053)
log (rural lot size)	−0.0522	−0.0478	−0.0476	−0.0633	−0.0703
	(0.0082)	(0.0088)	(0.0087)	(0.0084)	(0.0208)
log (age)	−0.0398	−0.0389	−0.0387	−0.0381	−0.0398
	(0.0097)	(0.0097)	(0.0096)	(0.0097)	(0.0163)
log (age)^2	−0.0104	−0.0105	−0.0106	−0.0109	−0.0106
	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0033)
garage	0.0468	0.0468	0.0468	0.0468	0.0470
	(0.0054)	(0.0053)	(0.0053)	(0.0054)	(0.0060)
log (living area)	0.4533	0.4536	0.4535	0.4528	0.4505
	(0.0079)	(0.0079)	(0.0080)	(0.0079)	(0.0083)
log (time to CBD)	−0.1821	−0.1896	−0.1963	−0.2358	−0.2628
	(0.0500)	(0.0505)	(0.0478)	(0.0710)	(0.8627)
log (time to CBD)^2	0.0117	0.0305	0.0312	0.0353	0.0373
	(0.0129)	(0.0121)	(0.0114)	(0.0154)	(0.1491)
log (accessibility)	0.1010	0.1604	0.1595	0.1563	0.1498
	(0.0251)	(0.0171)	(0.0165)	(0.0174)	(0.0282)
log (socio-economic conditions)	−0.0392	−0.0111	−0.0109	−0.0114	−0.0073
	(0.0179)	(0.0149)	(0.0157)	(0.0153)	(0.0317)
log (low education)	−0.1180	−0.1524	−0.1525	−0.1496	−0.1522
	(0.0218)	(0.0194)	(0.0208)	(0.0196)	(0.1208)

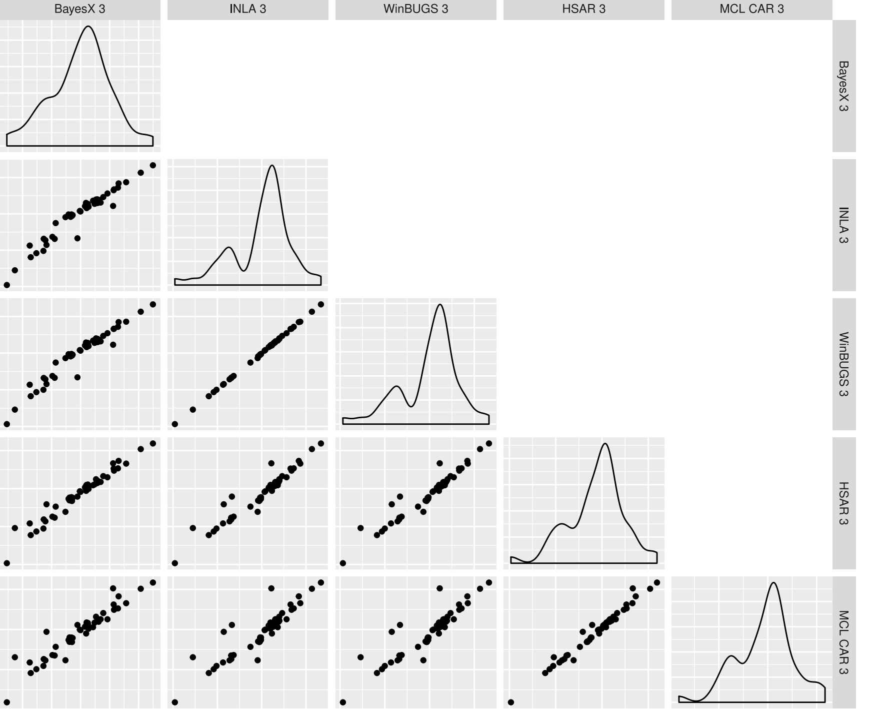

Fig. 8. Stavanger zones: scatter plot matrix of spatially structured random effects for five estimators of Model 3 (including covariates on individual houses and zonal covariates); the display shows a scatterplot for each pair of estimators being compared, with density plots of the spatially structured random effects in the diagonal panels.

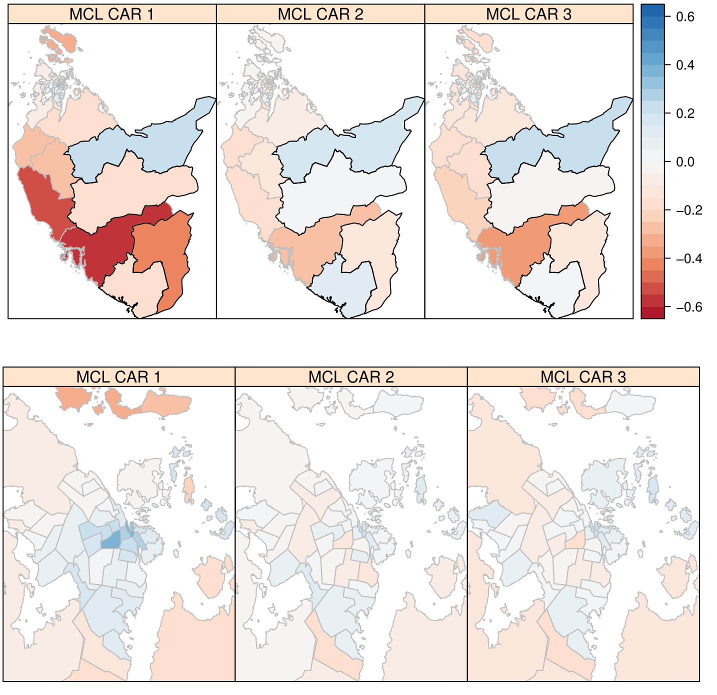

Fig. 9. Stavanger zones: Maps of spatially structured random effects for Models 1, 2 and 3 (MCL CAR estimator): upper panel: whole study area with four atypical zones bounded in black; lower panel: enlarged view of central Stavanger.

Table 9
Stavanger zones: summary measures of random effects estimates for Model 3 (including covariates on individual houses and zonal covariates).

	BayesX	INLA	WinBUGS	HSAR	MCL CAR
σ_η^2	0.0288	0.0292	0.0291	0.0111	0.0116
	(0.0075)			(0.0031)	(0.0049)
σ_ϵ^2	0.0386	0.0384	0.0384	0.0384	0.0383
	(0.0006)			(0.0007)	(0.0006)
λ				0.1182	0.1133
				(0.0242)	(0.1506)

estimates for the same four municipalities, something strengthening concern with regard to their influence on the outcome of model estimation.

Fig. 9 shows the estimates of the spatially structured random effects for the three models using MCL CAR, both for the study area as a whole, and for central parts of Stavanger shown at a greater scale. The four municipalities are shown bordered by black lines. It is easy to see that the successive introduction of zonal level covariates in Models 2 and 3 leads to the weakening of the extremes of the spatial random effects, as was also visible in Fig. 7.

Table 9 shows the values of σ_η^2 and σ_ϵ^2 for Model 3 estimated with the five different estimators. There are very few differences between the values of σ_ϵ^2, as seen in Table 7. The typical values of σ_ϵ^2 for Model 1 were 0.0419 and are 0.0384 for Model 3, so the difference is about five times the size of the standard error estimates of σ_ϵ^2. Note that while HSAR and MCL CAR report quite large values of λ, its standard error estimate for MCL CAR is large too, perhaps collapsing to the iid random effects. We can conclude that including the zonal level covariates has made a difference, but certainly MCL CAR, and possibly BayesX and HSAR are yielding results suggesting that these zonal level covariates are interacting with the multilevel construction of the model. The results for both data sets from China and Norway appear to indicate that it is the coefficients of distance related variables that are more sensitive to the choice of method of estimation.

4. Concluding remarks

Recent research has shown that multilevel models have the potential to be useful when analysing spatial data. In this paper we have shown that estimation of multilevel models, both CAR and SAR, may be carried out by using several implementations included in R packages. We have also tried out a new package mclcar for Monte Carlo maximum likelihood estimation. This method was long neglected, and has been retrospectively evaluated only recently, including its implementation in prototype form in R. So in this paper, a range of implementations have been used on two different data sets, chosen because the data sets have similar structures, and, since the Beijing land parcel price data set is fully accessible, our script comparing the implementations and using this data set can be provided as supplementary material (see Appendix A). The Stavanger data set cannot be distributed because it contains confidential information.

The findings from the reported estimations are very robust across the two different study areas. The results show that some of the implementations seem to estimate the model variants without raising concerns, while others seem to be more sensitive to peculiarities in the data. In particular, INLA and WinBUGS implementations of the intrinsic CAR model seem to run smoothly, and provide robust results across methods of estimation in both datasets. The use of MCL CAR and HGLM has been shown to be more sensitive to the inclusion of zonal level covariates, and to possible heterogeneity among the individual and district level observations. Moreover, the results from the intrinsic CAR models are very similar, whereas the results from the MCL CAR diverge the most from these models. The HSAR models produce results that are in between these extremes. In further research, it will be of interest to examine whether the apparent robustness of the intrinsic CAR approach is justified, or whether the issues flagged by the HGLM and Monte Carlo likelihood approaches are being submerged in overfitting by the intrinsic CAR assumptions. We hope that, by making the accompanying script available, we will have contributed to helping to bring more attention to this area of the application of spatial statistical methods.

Acknowledgements

The authors would like to thank Guanpeng Dong and Lars Rönnegård for helpful clarifications, and the editor and two referees for constructive comments. Any remaining confusion is the responsibility of the authors alone. This paper is based on funding from the Research Council of Norway through an Industrial Ph.D. for Ingrid Sandvig Thorsen.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.spasta.2017.01.002.

References

Alam, M., Rönnegård, L., Shen, X., 2015a. Fitting conditional and simultaneous autoregressive spatial models in hglm. The R Journal 7, 5–18. URL: http://journal.r-project.org/archive/2015-2/alam-ronnegard-shen.pdf.
Alam, M., Rönnegård, L., Shen, X., 2015b. hglm: Hierarchical generalized linear models. URL: https://CRAN.R-project.org/package=hglm R package version 2.1-1.
Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers.
Banerjee, S., Carlin, B.P., Gelfand, A., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC, London.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1–48. http://dx.doi.org/10.18637/jss.v067.i01, URL: https://www.jstatsoft.org/index.php/jss/article/view/v067i01.
Bates, D., Maechler, M., Bolker, B., Walker, S., 2016. lme4: Linear mixed-effects models using 'Eigen' and S4. URL: https://CRAN.R-project.org/package=lme4 R package version 1.1-12.
Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B 36, 192–236.
Besag, J.E., Kooperberg, C., 1995. On conditional and intrinsic autoregression. Biometrika 82, 733–746.

Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Statist. Math. 43, 1–59.

Bivand, R., 2016. Revisiting the boston data set (harrison and rubinfeld, 1978c): changing the units of observation affects estimated willingness to pay for clean air. Region URL: http://openjournals.wu.ac.at/ojs/index.php/region/index.

Bivand, R., Gómez-Rubio, V., Rue, H., 2014. Approximate Bayesian inference for spatial econometrics models. Spatial Stat. 9, 146–165.

Bivand, R., Gómez-Rubio, V., Rue, H., 2015. Spatial data analysis with R-INLA with some extensions. Journal of Statistical Software 63, 1–31. http://dx.doi.org/10.18637/jss.v063.i20, URL: https://www.jstatsoft.org/index.php/jss/article/view/v063i20.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-temporal Bayesian Models with R-INLA. Wiley.

Cappé, O., Douc, R., Moulines, E., Robert, C., 2002. On the convergence of the Monte Carlo maximum likelihood method for latent variable models. Scand. J. Stat. 29, 615–635.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., Liu, J., 2013. A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. Psychometrika 78, 685–709. URL: http://gllamm.org/.

Corrado, L., Fingleton, B., 2012. Where is the economics in spatial econometrics? J. Reg. Sci. 52, 210–239.

Dong, G., Harris, R., 2015. Spatial autorgressive models for geographically hierarchical data structures. Geogr. Anal. 47, 173–191.

Dong, G., Harris, R., Jones, K., Yu, J., 2015. Multilevel modeling with spatial interaction effects with application to an emerging land market in Beijing, China. PLOS One 10, http://dx.doi.org/10.1371/journal.pone.0130761.

Dong, G., Harris, R., Mimis, A., 2016a. HSAR: Hierarchical Spatial Autoregressive Model (HSAR). URL: https://CRAN.R-project.org/package=HSAR R package version 0.4.0.

Dong, G., Ma, J., Harris, R., Price, G., 2016b. Spatial random slope multilevel modeling using multivariate conditional autoregressive models: A case study of subjective travel satisfaction in beijing. Ann. Am. Assoc. Geogr. 106, 19–35.

Dorie, V., 2015. blme: Bayesian linear mixed-effects models. URL: https://CRAN.R-project.org/package=blme R package version 1.0-4.

Evans, M., Swartz, T., 2000. Approximating Integrals via Monte Carlo and Deterministic Methods. OUP Oxford.

Fahrmeir, L., Lang, S., 2001. Bayesian inference for generalized additive mixed models based on Markov random field priors. J. R. Stat. Soc. Ser. C 50, 201–220.

Gelfand, A.E., 2010. Misaligned spatial data: The change of support problem. In: Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M. (Eds.), Handbook of Spatial Statistics. Chapman & Hall/CRC, Boca Raton, pp. 517–539. 23.

Gelfand, A.E., Banerjee, S., Sirmans, C.F., Y, T., E, O.S., 2007. Multilevel modeling using spatial processes: Application to the Singapore housing market. Comput. Statist. Data Anal. 51, 3567–3579.

Gelman, A., 1996. Inference and monitoring convergence. In: Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, Boca Raton.

Gelman, A., Sturtz, S., Ligges, U., 2015. R2WinBUGS: Running 'WinBUGS' and 'OpenBUGS' from 'R' / 'S-PLUS'. URL: https://CRAN.R-project.org/package=R2WinBUGS R package version 2.1-21.

Gerber, F., Furrer, R., 2015. Pitfalls in the implementation of bayesian hierarchical modeling of areal count data: An illustration using bym and leroux models. Journal of Statistical Software 63, 1–32. http://dx.doi.org/10.18637/jss.v063.c01, URL: https://www.jstatsoft.org/index.php/jss/article/view/v063c01.

Geyer, C.J., 1994. On the convergence of Monte Carlo maximum likelihood calculations. J. R. Stat. Soc. Ser. B Stat. Methodol. 56, 261–274.

Geyer, C.J., Thompson, E.A., 1992. Constrained Monte Carlo maximum likelihood for dependent data. J. R. Stat. Soc. Ser. B Stat. Methodol. 54, 657–699.

Gómez-Rubio, V., Bivand, R., Rue, H., 2014. Spatial models using Laplace approximation methods. In: Handbook of Regional Science. Springer, Berlin Heidelberg, pp. 1401–1417.

Hansen, W.G., 1959. How accessibility shapes land use. J. Am. Inst. Plan. 25, 73–76.

Kandala, N.B., Lang, S., Klasen, S., Fahrmeir, L., 2001. Semiparametric Analysis of the Socio-Demographic and Spatial Determinants of Undernutrition in Two African Countries. Technical Report 245 Discussion Paper series, Collaborative Research Center 386. Ludwig-Maximilians-Universität, München.

Lee, D., 2013. CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors. Journal of Statistical Software 55, 1–24. http://dx.doi.org/10.18637/jss.v055.i13, URL: https://www.jstatsoft.org/index.php/jss/article/view/v055i13.

Lee, D., 2016. CARBayes: Spatial generalised linear mixed models for areal unit data. URL: https://CRAN.R-project.org/package=CARBayes R package version 4.6.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman and Hall/CRC, Boca Raton FL.

Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. Journal of Statistical Software 63, 1–25. http://dx.doi.org/10.18637/jss.v063.i19, URL: https://www.jstatsoft.org/index.php/jss/article/view/v063i19.

Liu, N., Roberts, D., 2012. Do incomers pay more for rural housing? Environ. Plan. A 44, 1986–2005.

Osland, L., Thorsen, I., 2008. Effects on housing prices of urban attraction and labor market accessibility. Environ. Plan. A 40, 2490–2509.

Osland, L., Thorsen, I.S., Thorsen, I., 2016. Accounting for local spatial heterogeneities in housing market studies. J. Reg. Sci. 56, 895–920. http://dx.doi.org/10.1111/jors.12281.

Pierewan, A.C., Tampubolon, G., 2014. Spatial dependence multilevel model of well-being across regions in Europe. Appl. Geogr. 47, 168–176. http://dx.doi.org/10.1016/j.apgeog.2013.12.005, URL: http://www.sciencedirect.com/science/article/pii/S0143622813002816.

Rabe-Hesketh, S., Skrondal, A., 2008. Multilevel and Longitudinal Modeling Using Stata. Stata Press Publication.

Rönnegård, L., Shen, X., Alam, M., 2010. hglm: A package for fitting hierarchical generalized linear models. The R Journal 2, 20–28. URL: http://journal.r-project.org/archive/2010-2/RJournal_2010-2_Roennegaard~et~al.pdf.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall, London.

Rue, H., Martino, S., Lindgren, F., Simpson, D., Riebler, A., Krainski, E.T., 2015. INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximaxion. R package version 0.0-1440400394.

Savitz, N.V., Raudenbush, S.W., 2009. Exploiting spatial dependence to improve measurement of neighborhood social processes. Sociol. Methodol. 39, 151–183. http://dx.doi.org/10.1111/j.1467-9531.2009.01221.x, URL: http://dx.doi.org/10.1111/j.1467-9531.2009.01221.x.

Sha, Z., 2016a. Estimation of conditional autoregressive models (Ph.D. thesis), University of Oxford.

Sha, Z., 2016b. mclcar: Estimating Conditional Auto-Regression (CAR) models using Monte Carlo likelihood methods. URL: https://CRAN.R-project.org/package=mclcar R package version 0.1-8.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS User Manual. http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf.

Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS: A package for running WinBUGS from R. Journal of Statistical Software 12, 1–16. URL: http://www.jstatsoft.org.

Thomas, A., Best, N., Lunn, D., Arnold, R., Spiegelhalter, D., 2004. GeoBUGS User Manual. http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/geobugs12manual.pdf.

Umlauf, N., Adler, D., Kneib, T., Lang, S., Zeileis, A., 2015a. Structured additive regression models: An R interface to BayesX. Journal of Statistical Software 63, 1–46. http://dx.doi.org/10.18637/jss.v063.i21, URL: https://www.jstatsoft.org/index.php/jss/article/view/v063i21.

Umlauf, N., Kneib, T., Lang, S., Zeileis, A., 2015b. R2BayesX: Estimate Structured Additive Regression models with bayesx. URL: https://CRAN.R-project.org/package=R2BayesX R package version 1.0-0.

Wood, S., 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

Wood, S., 2016. mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. URL: https://CRAN.R-project.org/package=mgcv R package version 1.8-14.