LGM Split Sampler: An Efficient MCMC Sampling Scheme for Latent Gaussian Models

Óli Páll Geirsson, Birgir Hrafnkelsson, Daniel Simpson and Helgi Sigurdarson

Abstract. A general and flexible class of latent Gaussian models is proposed in this paper. The latent Gaussian model is adapted to the generalized additive model for location, scale and shape (GAMLSS), that is, the data density function of each data point can depend on more than a single linear predictor of the latent parameters. We refer to this framework as extended latent Gaussian models. The most commonly applied latent Gaussian models (LGMs) are such that a linear predictor is proposed only for the location parameter. Extended LGMs allow proposing linear predictors also for the scale parameter and potentially other parameters. We propose a novel computationally efficient Markov chain Monte Carlo sampling scheme for the extended LGMs which we refer to as the LGM split sampler. It is a two block Gibbs sampling scheme designed to exploit the model structure of the extended LGMs. An extended LGM is constructed for a simulated dataset and the LGM split sampler is implemented for posterior simulations. The results demonstrate the flexibility of the extended LGM framework and the efficiency of the LGM split sampler.

Key words and phrases: Bayesian hierarchical models, Gibbs sampling, latent Gaussian models, Markov chain Monte Carlo, posterior simulation.

1. INTRODUCTION

Latent Gaussian models (LGMs) form a flexible subclass of Bayesian hierarchical models and have become popular in many areas of statistics and various fields of applications, as LGMs are practical from a statistical modeling point of view and readily interpretable. For example, LGMs play an important role in spatial statistics; see Cressie (1993), Diggle, Tawn and Moyeed (1998), Chilès and Delfiner (2012); statistical climatology (Cooley, Nychka and Naveau, 2007, Guttorp and Gneiting, 2006); disease mapping (Pettitt, Weir and Hart, 2002, Lawson,

2013); stochastic volatility models (Martino et al., 2011); and hydrology (Schaefli, Talamba and Musy, 2007), to name a few. The concept of LGMs has proven to be very successful for doing Bayesian inference, and a wealth of models may be presented as LGMs, which in turn are amenable to the Bayesian inference approaches developed specifically for LGMs; see Martins et al. (2013a) for various applications and further references.

However, many LGMs in the literature focus only on a subclass of LGMs in which the response data density is only allowed to depend on a single linear predictor of the latent parameters. While this assumption holds in some practical cases, neglecting potential dependence on multiple parameters or features of a data density may render inferential conclusions invalid as argued in Klein et al. (2013, 2015) and Kneib (2013). For example, in Hrafnkelsson, Morris and Baladandayuthapani (2012) and Geirsson, Hrafnkelsson and Simpson (2015), latent Gaussian spatial models were imposed on the location, scale and shape parameters of the data density function in order to capture underlying spatial dependence.

This paper focuses specifically on LGMs where the data density function of each data point can depend on more than a single linear predictor of the latent parameters, as

Óli Páll Geirsson is Chief Data Officer, City of Reykjavik, Borgartún 12-14, 105 Reykjavik, Iceland (e-mail: olipalli@gmail.com). Birgir Hrafnkelsson is Professor of Statistics, Department of Mathematics, Faculty of Physical Sciences, University of Iceland, Dunhagi 5, 107 Reykjavik, Iceland (e-mail: birgirhr@hi.is). Daniel Simpson is Assistant Professor, Department of Statistical Sciences, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada (e-mail: simpson@utstat.toronto.edu). Helgi Sigurdarson is Project Manager, Air Navigation Services, Isavia, Reykjavik Airport, 102 Reykjavik, Iceland (e-mail: Helgi.Sigurdarson@isavia.is).

discussed in Martins et al. (2013b). We refer to this framework as extended LGMs when the need for such clarity is necessary. In the extended LGM framework, the data density can be conditioned on several linear predictors, say $\eta^{(1)}, \ldots, \eta^{(k)}$, one for each parameter of a given data density with k parameters. As such, extended LGMs can be viewed as a specific extension of additive regression models discussed in Fahrmeir and Tutz (1994). Imposing latent models on every parameter of a given data density is in line with the structured additive distributional regression approach presented in Rigby and Stasinopoulos (2005) and Klein et al. (2013, 2015), in which each parameter of a response data density is related to a structured additive predictor for improving modeling flexibility.

The goal of this paper is to provide a framework for extended LGMs. We discuss how to construct extended LGMs and propose a novel Bayesian inference algorithm tailored for extended LGMs. From an inferential perspective, the mathematical structure of LGMs is well suited for the development of efficient inferential algorithms. Of particular notice is the integrated nested Laplace approximation (INLA) proposed by Rue, Martino and Chopin (2009). INLA is an inferential algorithm which performs approximate Bayesian inference on a subclass of LGMs, through use of deterministic nested Laplace approximations. INLA has been shown to give fast and accurate estimates of posterior marginals, and has also been shown to be a valuable tool in practice via the R-package R-INLA; see Rue et al. (2017) and the web-site www.r-inla.org for available software. However, R-INLA only provides support for LGMs in which the data density of each data point only depends on a single linear predictor of the latent parameters, although the INLA idea has already been reimplemented in a special case of an extended LGM; see (Ferkingstad et al., 2008). The extension of INLA to extended LGMs might seem trivial but it is not (Martins et al., 2013b). Thus, in order to provide an efficient framework for extended LGMs we propose a novel efficient Markov chain Monte Carlo (MCMC) inferential algorithm which is adapted to extended LGMs. The proposed MCMC algorithm will be referred to as the LGM split sampler.

Several MCMC sampling strategies have been suggested for Bayesian hierarchical models to improve the mixing properties of MCMC algorithms. For example, methods based on approximate diffusions such as the Metropolis-adjusted Langevin algorithm (MALA), see Roberts and Rosenthal (1998); methods based on Hamiltonian mechanics (HMC), suggested by Neal (1993), which use the gradient of the target density to drive the proposal mechanism toward regions of higher posterior density; manifold methods, proposed by Girolami and Calderhead (2011), which provide a systematic way of designing proposal densities for MALA and HMC by making use of the gradient and curvature information of the target density; and various block sampling strategies such as the single block updating strategy of Knorr-Held and Rue (2002). Filippone, Zhong and Girolami (2013) conducted a detailed comparison of these methods for LGMs and found that the single block strategy of Knorr-Held and Rue (2002), in which the latent parameters and its corresponding hyperparameters are updated jointly in a single block, performed best in most situations. However, the single block updating strategy is only applicable to a subclass of LGMs, where the latent Gaussian model is imposed only on the location parameter.

The LGM split sampler is a two block Gibbs sampling scheme (Rue and Held, 2005) designed to utilize the model structure of extended LGMs for efficient MCMC sampling. The principal idea behind the LGM split sampler is to split the latent Gaussian parameters into two vectors, such that the single block sampler of Knorr-Held and Rue (2002) becomes applicable to a subset of the model parameters. The first vector, η, consists of elements that appear in the data density function and the second vector, v, consists of elements that do not appear in it. We construct the two sampling blocks by placing v and all hyperparameters into a block we refer to as the data-poor block, while η is placed in another sampling block we refer to as the data-rich block.

We show that the conditional posterior density of v conditioned on the parameter vector η and the hyperparameters becomes Gaussian. Consequently, a modified version of the single block sampler (Knorr-Held and Rue, 2002) can be constructed and implemented for the parameter vector v and the hyperparameters in the data-poor block. This proposed sampling scheme is invariant of the choice of a data density function as the data-poor block is updated conditioned on the data-rich block. Moreover, if the joint Gaussian prior density for the latent parameters η and v is a Gaussian Markov random field (GMRF) (Rue and Held, 2005) with a sparse precision matrix, we show how the proposed sampling strategy for the data-poor block conserves the sparse GMRF precision structure in the cases of both the hyperparameters conditional on η, and v conditional on η and the hyperparameters, thus reducing computational cost. We further demonstrate that the proposed sampling strategy scales well in terms of MCMC sampling efficiency as the dimension of v increases.

In practical applications of extended LGMs with non-Gaussian data density functions, especially in the field of spatial statistics, the vector v in the data-poor block has sometimes higher dimension than the vector η in the data-rich block; see, for example, Hrafnkelsson, Morris and Baladandayuthapani (2012). Additionally, the conditional posterior of η in the data-rich block is in most

cases only known up to a constant and is computationally costly to evaluate. However, in the proposed blocking scheme of the LGM split sampler the data-rich block contains a minimum number of parameters which are needed for the evaluation of the usually computationally costly conditional posterior density. In this paper, we implement a Metropolis–Hasting-type MCMC algorithm with a tailored independence proposal density (Rue and Held, 2005) in the data-rich block. We construct the independence proposal density by approximating the conditional posterior density in the data-rich block with a Gaussian approximation evaluated at the mode. We give an example to demonstrate that the proposed sampling strategy in the data-rich block is highly efficient. The LGM split sampler is modular in the sense that η in the data-rich block can be tackled with any sampling scheme; the same is true for the hyperparameters in the data-poor block while the proposed sampling scheme for v in the data-poor block conditional on η and the hyperparameters is optimal. So, other sampling schemes, more efficient than the ones proposed in this paper, can potentially be found or developed for η and the hyperparameters; however, this is beyond the scope of the paper.

The structure of the paper is such that in Section 2 the extended LGM is presented and the LGM split sampler is proposed. In Section 3, an example is given to demonstrate how the extended LGM setup and the LGM split sampler can be applied. Finally, in Section 4 a discussion is given.

2. EXTENDED LATENT GAUSSIAN MODELS

2.1 LGMs and Additive Models

A general formulation for latent Gaussian models can be presented as follows. The vectors y, x and θ contain the observations, the latent parameters and the hyperparameters, respectively.

Data-level: The observations y depend on the latent parameters x, through some choice of data distribution with a data density function $\pi(y \mid x, \theta)$.

Latent level: The prior density for the latent parameters x is Gaussian and is potentially dependent on the hyperparameters θ, that is,

$$\pi(x \mid \theta) = \mathcal{N}(x \mid \mu(\theta), Q(\theta)^{-1}).$$

Hyperparameter level: A prior distribution is assigned for the hyperparameters θ, with a density function $\pi(\theta)$.

LGMs are related to regression models (Fahrmeir and Tutz, 1994) and generalized linear regression models. To see how, assume, for example, a Gaussian data density where each data point y_i only depends on a single linear predictor η_i, for example, through the mean of the Gaussian density, of the latent parameters, that is,

$$(2.1) \qquad \eta_i = \sum_j \beta_j z_{ij} + \sum_k u_k a_{ik} + \varepsilon_i.$$

Here, $\{z_{ij}\}$ are known covariates and $\{\beta_j\}$ are the corresponding coefficients, sometimes referred to as fixed effects. The terms $\{u_k\}$ are structured random effects, where each u_k serves as a model component which contributes to the ith component of the linear predictor if $a_{ik} \neq 0$, where $\{a_{ik}\}$ are known weights. Examples of such structured random effects are time-series models, stochastic spline models, spatial models, etc. The term ε_i is an unstructured random effect and serves as a model error term.

Now, the linear predictor equation in (2.1) in its vectorized form becomes

$$(2.2) \qquad \eta = X\beta + Au + \varepsilon,$$

where X is a design matrix based on the covariates $\{z_{ij}\}$, β is the corresponding vector of the covariate coefficients $\{\beta_j\}$, u is a vector containing the structured random effects, A is a fixed matrix based on $\{a_{ik}\}$ mapping the structured random effects $\{u_k\}$ to their related component of the linear predictor and ε is a vector of the unstructured random effects. When Gaussian priors are assigned to β, u and ε, the structured additive regression formulation in (2.2) fits into the LGMs setup since the joint distribution of the vector

$$x = (\beta^{\mathsf{T}}, u^{\mathsf{T}}, \varepsilon^{\mathsf{T}})^{\mathsf{T}}$$

becomes Gaussian, and thus, this vector is the latent parameter vector in the general LGM setup presented above.

2.2 The Structure of Extended LGMs

The generalized additive model for location, scale and shape (GAMLSS) of Rigby and Stasinopoulos (2005), also presented in Klein et al. (2013), is such that each natural parameter of the data density function is modeled by a structural additive predictor. In this section, we highlight how models in the extended LGM framework can be expressed in order to impose latent models on any of the natural parameter of a given data density function in a fashion similar to that of Rigby and Stasinopoulos (2005) and Klein et al. (2013). To that end, consider without loss of generality, a data density function $\pi(y \mid \mu, \tau)$ where μ and τ are vectors of location and log-scale parameters, respectively. Next, impose latent Gaussian models on both μ and τ, for example, the following additive model structure:

$$
\begin{aligned}
(2.3) \qquad \mu &= X_\mu \beta_\mu + A_\mu u_\mu + \varepsilon_\mu, \\
\tau &= X_\tau \beta_\tau + A_\tau u_\tau + \varepsilon_\tau,
\end{aligned}
$$

where X_μ and X_τ are fixed design matrices; β_μ and β_τ are the corresponding coefficients; A_μ and A_τ are fixed matrices; u_μ and u_τ are structured random effects; and ε_μ and ε_τ are unstructured random effects which serve as model error terms. Adding the model error terms ε_μ and ε_τ to the latent models is in line with structured additive regression discussed in Fahrmeir and Tutz (1994) and

the latent Gaussian model presented in Rue, Martino and Chopin (2009). Furthermore, small variances can be imposed a priori on the model errors terms even if they are not desired in the model. However, adding the unstructured random effects is reasonable in many cases from a statistical modeling point of view as they serve as error terms for the latent models. Moreover, adding the unstructured random effects makes the posterior inference computationally feasible. On the other hand, if the variance of an unstructured random effect is set equal to a very small value or the data suggest that this variance is very small then the posterior inference may not be as computationally feasible since the dependence between the parameters found at the data level, for example, μ and τ in the above example, and the parameters that are only found at the latent level, for example, β_μ, u_μ, β_τ and u_τ in the above example, increases when this variance decreases.

The model structure in (2.3) requires informative prior densities for the hyperparameters. In particular, in a spatial model in which μ and τ contain the latent parameters at individual sites and $A_\mu u_\mu$ and $A_\tau u_\tau$ represent the spatial model components at the individual sites, the hyperparameters are hard to identify. This is supported by results from Zhang (2004), namely, the log-likelihood function for the range parameter is very flat for any value of the marginal variance.

Assign the following Gaussian prior density functions to the latent model parameters

$$\pi(\beta_\mu) = \mathcal{N}(\beta_\mu \mid \mu_{\beta\mu}, Q_{\beta\mu}^{-1}),$$
$$\pi(\beta_\tau) = \mathcal{N}(\beta_\tau \mid \mu_{\beta\tau}, Q_{\beta\tau}^{-1}),$$
(2.4)
$$\pi(u_\mu) = \mathcal{N}(u_\mu \mid \mu_{u\mu}, Q_{u\mu}^{-1}),$$
$$\pi(u_\tau) = \mathcal{N}(u_\tau \mid \mu_{u\tau}, Q_{u\tau}^{-1}),$$
$$\pi(\varepsilon_\mu) = \mathcal{N}(\varepsilon_\mu \mid 0, Q_{\varepsilon\mu}^{-1}),$$
$$\pi(\varepsilon_\tau) = \mathcal{N}(\varepsilon_\tau \mid 0, Q_{\varepsilon\tau}^{-1}),$$

where the parameters of the prior density functions can potentially depend on the hyperparameters θ, and $Q_{\varepsilon\mu}^{-1}$ and $Q_{\varepsilon\tau}^{-1}$ are diagonal matrices.

Since the vector μ in equation (2.3) is a linear combination of β_μ, u_μ and ε_μ, it is equivalent to obtain MCMC samples from the posterior distribution of (μ, β_μ, u_μ) and from the posterior distribution of $(\beta_\mu, u_\mu, \varepsilon_\mu)$. Analogous argument holds for the parameters corresponding to τ. The LGM split sampler is designed to obtain MCMC samples from the posterior distribution of $(\mu, \tau, \beta_\mu, \beta_\tau, u_\mu, u_\tau)$ as opposed to $(\beta_\mu, \beta_\tau, u_\mu, u_\tau, \varepsilon_\mu, \varepsilon_\tau)$ as in the former parameterization only the vector (μ, τ) enters the data density function, while all the elements of the latter vector enter the data density function in the latter parameterization. This parameterization for posterior inference is along the lines

of the posterior inference scheme proposed in Rue, Martino and Chopin (2009). Thus, define

$$\eta = (\mu^\mathsf{T}, \tau^\mathsf{T})^\mathsf{T}, \qquad v = (\beta_\mu^\mathsf{T}, u_\mu^\mathsf{T}, \beta_\tau^\mathsf{T}, u_\tau^\mathsf{T})^\mathsf{T}$$

which will act as the splitting of the latent parameters. The latent model structure in (2.3) and the prior distributions in (2.4) can be written in a joint matrix form, which forms the basis for the derivation of the LGM split sampler. Define the following matrices and vectors:

$$Z = \begin{pmatrix} X_\mu & A_\mu & \cdot & \cdot \\ \cdot & \cdot & X_\tau & A_\tau \end{pmatrix},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_\mu \\ \varepsilon_\tau \end{pmatrix}, \qquad Q_\varepsilon = \begin{pmatrix} Q_{\varepsilon\mu} & \cdot \\ \cdot & Q_{\varepsilon\tau} \end{pmatrix}$$

and group the following parameters and matrices together:

$$\mu_v = \begin{pmatrix} \mu_{\beta\mu} \\ \mu_{u\mu} \\ \mu_{\beta\tau} \\ \mu_{u\tau} \end{pmatrix},$$

$$Q_v = \begin{pmatrix} Q_{\beta\mu} & \cdot & \cdot & \cdot \\ \cdot & Q_{\mu u} & \cdot & \cdot \\ \cdot & \cdot & Q_{\beta\tau} & \cdot \\ \cdot & \cdot & \cdot & Q_{\tau u} \end{pmatrix},$$

where the dotted entries denote zero entries. The additive model structure implied by (2.3) for both latent parameters is thus equivalent to the matrix form

(2.5) $$\eta = Zv + \varepsilon$$

and the Gaussian prior assumptions in (2.4) are equivalent to

(2.6)
$$\pi(\eta \mid v, \theta) = \mathcal{N}(\eta \mid Zv, Q_\varepsilon^{-1}),$$
$$\pi(v \mid \theta) = \mathcal{N}(v \mid \mu_v, Q_v^{-1}).$$

By using the Gaussian prior assumptions in (2.6), it follows that the joint density of η, $v|\theta$ is a Gaussian density of the form

$$\pi\left(\begin{pmatrix} \eta \\ v \end{pmatrix} \Big| \theta \right)$$

(2.7)
$$= \mathcal{N}\left(\begin{pmatrix} \eta \\ v \end{pmatrix} \Big| \begin{pmatrix} Z\mu_v \\ \mu_v \end{pmatrix}, \right.$$

$$\left. \begin{pmatrix} Q_\varepsilon & -Q_\varepsilon Z \\ -Z^\mathsf{T} Q_\varepsilon & Q_v + Z^\mathsf{T} Q_\varepsilon Z \end{pmatrix}^{-1} \right).$$

As the data density function and the corresponding parameters were arbitrarily chosen above, analogous derivations can be carried out for any parametric data density function and all of its parameters jointly. For example, in addition to imposing latent Gaussian models on the location and log-scale parameters of the generalized extreme

value distribution, equations (2.3) and (2.4) can be extended by imposing a latent Gaussian model on the shape parameter; see Section 3 for details. Therefore, equations (2.5), (2.6) and (2.7) are general for all LGMs in the extended LGM framework. We will thus adapt equations (2.5) and (2.6) as the setup for LGMs henceforth in this paper. Note that, as the vector $x = (\eta^\mathsf{T}, v^\mathsf{T})^\mathsf{T}$ is Gaussian it can be viewed as the latent parameter x in the general LGM setup in Section 2.1.

2.3 Inference for Extended LGMs

The vector η in the LGM setup in (2.5) and (2.6) consists of the latent parameters that explicitly enter the data density function while the vector v consists of latent parameters which do not enter it. Therefore, the data vector y is conditionally independent of v conditioned on η, that is, $\pi(y \mid \eta, v) = \pi(y \mid \eta)$. The parameters η and v are referred to as the *data-rich latent parameters* and the *data-poor latent parameters*, respectively, in this paper. The corresponding posterior distribution, where the data-poor latent parameters are potentially dependent on a vector of hyperparameters, denoted by θ, is thus proportional to

$$(2.8) \qquad \pi(\eta, v, \theta \mid y) \propto \pi(y \mid \eta)\pi(\eta, v \mid \theta)\pi(\theta).$$

The block sampling scheme of the LGM split sampler consists of grouping all the model parameters into two separate blocks. That is, both v and the hyperparameters θ are placed in one block referred to as the *data-poor* block, while η is placed in the other block referred to as the *data-rich* block in this paper. A Gibbs sampling strategy is then implemented for each block, conditioned on the other block. That is, the $(k + 1)$th MCMC sample from the posterior density $\pi(\eta, v, \theta \mid y)$ is obtained by using the following two block Gibbs sampling scheme:

Data-poor block: sample (v^{k+1}, θ^{k+1}) jointly from $\pi(v, \theta \mid y, \eta^k)$.
Data-rich block: sample η^{k+1} from $\pi(\eta \mid y, v^{k+1}, \theta^{k+1})$.

To elaborate on the potential efficiency gain offered by this blocking scheme, consider the following. Due to the splitting of the latent field expressed in equation (2.8), the parameters v and θ in the data-poor block become conditionally independent of the data y conditioned on the vector η from the data-rich block, that is,

$$\pi(v, \theta \mid y, \eta) = \pi(v, \theta \mid \eta).$$

In other words, the form of the conditional posterior $\pi(v, \theta \mid y, \eta)$ becomes invariant of the data density function. The resulting conditional posterior density function in the data-poor block is thus proportional to

$$(2.9) \qquad \pi(v, \theta \mid y, \eta) \propto \pi(\theta)\pi(v \mid \eta, \theta).$$

Furthermore, the conditional posterior density of the data-poor latent parameters v conditioned on the data-rich latent parameters η and the hyperparameters θ becomes Gaussian and is independent of the data y, that is,

$$(2.10) \qquad \begin{aligned} &\pi(v \mid y, \eta, \theta) \\ &= \pi(v \mid \eta, \theta) \\ &= \mathcal{N}(v \mid Q_{v\mid\eta}^{-1}(Q_v\mu_v + Z^\mathsf{T} Q_\varepsilon\eta), Q_{v\mid\eta}^{-1}), \end{aligned}$$

where $Q_{v\mid\eta} = Q_v + Z^\mathsf{T} Q_\varepsilon Z$; see Appendix A.1 for derivation. Due to the Gaussianity offered by equation (2.10), it is possible to obtain exact samples from $\pi(v\mid\eta, \theta)$ in every MCMC iteration.

When the data density function is non-Gaussian or Gaussian with a latent model on the log-variance, the conditional posterior density $\pi(\eta \mid y, v, \theta)$ in the data-rich block becomes computationally challenging in posterior simulations. However, a key part of the design is such that the data-rich block contains a minimum number of parameters needed for the evaluation of the computationally demanding conditional posterior density $\pi(\eta \mid y, v, \theta)$. In other words, the dimension of the computationally demanding data-rich block is designed to be as low as possible. This is particularly helpful when the dimension of the data-rich component of the latent field, η, is low. However, when the dimension of η is high the conditional posterior density $\pi(\eta \mid y, v, \theta)$ will be computationally demanding.

The LGM split sampler is designed such that, in principle, any efficient MCMC sampler can be implemented for each block. In this paper, computationally efficient sampling strategies tailored to the particular conditional model structure of each block are proposed. These strategies are described in the subsequent sections.

2.3.1 *Sampler for the data-poor block.* The relation in (2.9) and the Gaussianity of the conditional posterior density $\pi(v\mid\eta, \theta)$ in (2.10) demonstrate that the single-block updating scheme of Knorr-Held and Rue (2002) is applicable within the data-poor block with some modifications. As discussed in the Introduction, the single-block updating scheme is known to outperform most modern sampling algorithms in terms of MCMC sampling efficiency when applicable. Therefore, the following modified version of the single-block updating scheme, adapted for the data-poor block of the MCMC split sampler, is proposed.

For some proposal density $q(\theta^* \mid \theta^k)$ for the hyperparameters θ, a new proposed value (v^*, θ^*) is generated jointly as follows:

$$(2.11) \qquad \begin{aligned} \theta^* &\sim q(\theta^* \mid \theta^k), \\ v^* &\sim \pi(v^* \mid \eta^k, \theta^*). \end{aligned}$$

Denote the proposal density implied by (2.11) with $q(\boldsymbol{v}^*, \boldsymbol{\theta}^* | \boldsymbol{v}^k, \boldsymbol{\theta}^k)$. The proposed value $(\boldsymbol{v}^*, \boldsymbol{\theta}^*)$ is then accepted jointly with acceptance probability

$$(2.12) \quad \alpha = \min\left\{1, \frac{\pi(\boldsymbol{v}^*, \boldsymbol{\theta}^* | \boldsymbol{y}, \boldsymbol{\eta}^k)}{\pi(\boldsymbol{v}^k, \boldsymbol{\theta}^k | \boldsymbol{y}, \boldsymbol{\eta}^k)} \frac{q(\boldsymbol{v}^k, \boldsymbol{\theta}^k | \boldsymbol{v}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{v}^*, \boldsymbol{\theta}^* | \boldsymbol{v}^k, \boldsymbol{\theta}^k)}\right\}.$$

It can be shown, see Appendix A.2, that the acceptance ratio in (2.12) can be simplified to

$$
\begin{aligned}
(2.13) \quad & \frac{\pi(\boldsymbol{v}^*, \boldsymbol{\theta}^* | \boldsymbol{y}, \boldsymbol{\eta}^k)}{\pi(\boldsymbol{v}^k, \boldsymbol{\theta}^k | \boldsymbol{y}, \boldsymbol{\eta}^k)} \frac{q(\boldsymbol{v}^k, \boldsymbol{\theta}^k | \boldsymbol{v}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{v}^*, \boldsymbol{\theta}^* | \boldsymbol{v}^k, \boldsymbol{\theta}^k)} \\
& = \frac{\pi(\boldsymbol{\theta}^* | \boldsymbol{\eta}^k)}{\pi(\boldsymbol{\theta}^k | \boldsymbol{\eta}^k)} \frac{q(\boldsymbol{\theta}^k | \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^k)}.
\end{aligned}
$$

The relation in (2.13) shows that the acceptance ratio is only dependent on the acceptance ratio for $\boldsymbol{\theta}$. Thus, this steps, which is conditional on $\boldsymbol{\eta}$, amounts to sampling first from the marginal distribution of $\boldsymbol{\theta}$ and then from the conditional distribution of $\boldsymbol{v}$. As a result, the dependence between $\boldsymbol{\theta}$ and $\boldsymbol{v}$ is broken down. To save computation time, sampling of $\boldsymbol{v}^*$ is delayed until acceptance is verified. Furthermore, since the conditional posterior $\pi(\boldsymbol{v} | \boldsymbol{\eta}, \boldsymbol{\theta})$ is a known Gaussian (see equation (2.10)), and the elements of $\boldsymbol{v}$ can be sampled jointly, the proposed sampling strategy scales well in terms of MCMC sampling efficiency as the dimension of the data-poor latent parameters, $\boldsymbol{v}$, increases.

If the Gaussian models in the prior assumptions (2.6) are GMRF density functions (Rue and Held, 2005), with a sparse precision structure, the ratio in (2.13) is computationally costly to calculate directly, since $\pi(\boldsymbol{\theta} | \boldsymbol{\eta}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\eta} | \boldsymbol{\theta})$ and the density $\pi(\boldsymbol{\eta} | \boldsymbol{\theta})$ does not necessarily preserve the sparse GMRF structure. However, from the law of conditional independence it follows that

$$(2.14) \quad \pi(\boldsymbol{\eta} | \boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\eta} | \boldsymbol{v}, \boldsymbol{\theta})\pi(\boldsymbol{v} | \boldsymbol{\theta})}{\pi(\boldsymbol{v} | \boldsymbol{\eta}, \boldsymbol{\theta})},$$

where the density functions on the right-hand side in (2.14) preserve the sparse GMRF precision structure. Furthermore, as the density $\pi(\boldsymbol{\eta} | \boldsymbol{\theta})$ is invariant of $\boldsymbol{v}$, the value of $\boldsymbol{v}$ can be set to zero for computational gain on the right-hand side of (2.14). Main result 1 summarizes how the ratio in (2.12) can be calculated with low computational cost by using the results in (2.13), (2.15) and (2.16) in the case of GMRFs with sparse precision structures. This is a key result for the implementation of the proposed sampling scheme in the data-poor block for Gaussian prior densities with sparse GMRFs precision structures. The proof can be seen in Appendix A.3. The algorithm for the sampling scheme in the data-poor block is summarized in Algorithm 1.

Algorithm 1 The proposed algorithm for obtaining the $(k+1)$th sample from $\pi(\boldsymbol{v}, \boldsymbol{\theta} | \boldsymbol{y}, \boldsymbol{\eta})$ in the data-poor block

Input: $(\boldsymbol{v}^k, \boldsymbol{\theta}^k, \boldsymbol{\eta}^k)$

1: Sample each element of $\boldsymbol{\theta}^*$ from a proposal density $q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^k)$

2: Calculate

$$r = \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k)} \cdot \frac{\pi(\boldsymbol{\eta}^k | \mathbf{0}, \boldsymbol{\theta}^*)\pi(\mathbf{0} | \boldsymbol{\theta}^*)}{\pi(\mathbf{0} | \boldsymbol{\eta}^k, \boldsymbol{\theta}^*)}$$

$$\cdot \frac{\pi(\mathbf{0} | \boldsymbol{\eta}^k, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{\eta}^k | \mathbf{0}, \boldsymbol{\theta}^k)\pi(\mathbf{0} | \boldsymbol{\theta}^k)}$$

$$\cdot \frac{q(\boldsymbol{\theta}^k | \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^k)}$$

on a logarithmic scale, using the equations in (2.16) for the conditional posterior densities functions

3: Calculate $\alpha = \min\{1, r\}$

4: Sample $u \sim \mathcal{U}(0, 1)$

5: **if** $\alpha > u$

6:     Calculate $\boldsymbol{Q}_{v|\eta} = \boldsymbol{Q}_v + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\varepsilon\boldsymbol{Z}$

7:     Sample $\boldsymbol{v}^*$ from

$$\boldsymbol{v}^* | \boldsymbol{\eta}^k, \boldsymbol{\theta}^*$$

$$\sim \mathcal{N}(\boldsymbol{v} | \boldsymbol{Q}_{v|\eta}^{-1}(\boldsymbol{Q}_v\boldsymbol{\mu}_v + \boldsymbol{Z}^\mathsf{T}\boldsymbol{Q}_\varepsilon\boldsymbol{\eta}^{k+1}), \boldsymbol{Q}_{v|\eta}^{-1})$$

8:     $(\boldsymbol{v}^{k+1}, \boldsymbol{\theta}^{k+1}) = (\boldsymbol{v}^*, \boldsymbol{\theta}^*)$

9: **else if** $\alpha < u$

10:    $(\boldsymbol{v}^{k+1}, \boldsymbol{\theta}^{k+1}) = (\boldsymbol{v}^k, \boldsymbol{\theta}^k)$

11: **end if**

Output: $(\boldsymbol{v}^{k+1}, \boldsymbol{\theta}^{k+1})$

MAIN RESULT 1. *The term $\pi(\boldsymbol{\theta}^* | \boldsymbol{\eta}^k)/\pi(\boldsymbol{\theta}^k | \boldsymbol{\eta}^k)$ in* (2.13) *can be rewritten as*

$$
\begin{aligned}
& \frac{\pi(\boldsymbol{\theta}^* | \boldsymbol{\eta}^k)}{\pi(\boldsymbol{\theta}^k | \boldsymbol{\eta}^k)} \\
(2.15) \quad & = \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k)} \cdot \frac{\pi(\boldsymbol{\eta}^k | \mathbf{0}, \boldsymbol{\theta}^*)\pi(\mathbf{0} | \boldsymbol{\theta}^*)}{\pi(\mathbf{0} | \boldsymbol{\eta}^k, \boldsymbol{\theta}^*)} \\
& \quad \cdot \frac{\pi(\mathbf{0} | \boldsymbol{\eta}^k, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{\eta}^k | \mathbf{0}, \boldsymbol{\theta}^k)\pi(\mathbf{0} | \boldsymbol{\theta}^k)}.
\end{aligned}
$$

Additionally, the conditional density functions on the right-hand side in (2.15) *on a logarithmic scale are*

$$\log\pi(\boldsymbol{\eta} | \mathbf{0}, \boldsymbol{\theta})$$

$$= \frac{1}{2}\log\det\boldsymbol{Q}_\varepsilon - \frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{Q}_\varepsilon\boldsymbol{\eta} + \text{const},$$

$$\log\pi(\mathbf{0} | \boldsymbol{\theta})$$

$$(2.16) \quad = \frac{1}{2}\log\det\boldsymbol{Q}_v - \frac{1}{2}\boldsymbol{\mu}_v^\mathsf{T}\boldsymbol{Q}_v\boldsymbol{\mu}_v + \text{const},$$

$$\log \pi(\mathbf{0} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$$

$$= \frac{1}{2} \log \det \boldsymbol{Q}_{\nu \mid \eta}$$

$$- \frac{1}{2} \big(\boldsymbol{Q}_{\nu \mid \eta}^{-1} (\boldsymbol{Q}_\nu \mu_\nu + \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\varepsilon \boldsymbol{\eta}) \big)^\mathsf{T}$$

$$\cdot (\boldsymbol{Q}_\nu \mu_\nu + \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\varepsilon \boldsymbol{\eta}) + \text{const},$$

where $\boldsymbol{Q}_{\nu \mid \eta} = \boldsymbol{Q}_\nu + \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\varepsilon \boldsymbol{Z}$. Moreover, if the Gaussian prior density functions in (2.6) *are GMRFs with sparse precision structures, then all of the conditional density functions on the right-hand side of* (2.16) *are GMRFs with sparse precision structures.*

2.3.2 *Sampler for the data-rich block.* The conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ in the data-rich block is

(2.17)
$$\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) = f(\boldsymbol{\eta}) - \frac{1}{2} \boldsymbol{\eta}^\mathsf{T} \boldsymbol{Q}_\varepsilon \boldsymbol{\eta}$$
$$+ (\boldsymbol{Q}_\varepsilon \boldsymbol{Z} \boldsymbol{v})^\mathsf{T} \boldsymbol{\eta} + \text{const},$$

where $f(\boldsymbol{\eta}) = \log \pi(\boldsymbol{y} \mid \boldsymbol{\eta})$ for notational convenience. As the conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ is often only known up to a constant, a Metropolis–Hastings algorithm with an independence proposal density is proposed for the data-rich block. The proposal density is constructed by approximating the conditional posterior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ with a Gaussian approximation evaluated at the mode, $\boldsymbol{\eta}^0$, of $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$. It can be shown that the Gaussian approximation of the conditional posterior becomes

(2.18)
$$\tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\eta} \mid \boldsymbol{\eta}^0, (\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1}),$$

where $\boldsymbol{H} = \nabla^2 \log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$; see Appendix A.4 for derivation.

Once a new $\boldsymbol{\eta}^*$ is proposed in the $(k+1)$th iteration of the MCMC chain with the independence proposal density $q(\boldsymbol{\eta}^*) = \tilde{\pi}(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{v}^k, \boldsymbol{\theta}^k)$, it is accepted with the following acceptance probability:

(2.19) $\alpha = \min \left\{ 1, \dfrac{\pi(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{v}^{k+1}, \boldsymbol{\theta}^{k+1})}{\pi(\boldsymbol{\eta}^k \mid \boldsymbol{y}, \boldsymbol{v}^{k+1}, \boldsymbol{\theta}^{k+1})} \cdot \dfrac{q(\boldsymbol{\eta}^k)}{q(\boldsymbol{\eta}^*)} \right\}.$

The logarithm of the acceptance ratio given in (2.19) can be simplified to

(2.20)
$$r = f(\boldsymbol{\eta}^*) - \left(\frac{1}{2} (\boldsymbol{\eta}^*)^\mathsf{T} \boldsymbol{H} + \boldsymbol{b}^\mathsf{T} \right) \boldsymbol{\eta}^* - f(\boldsymbol{\eta}^k)$$
$$+ \left(\frac{1}{2} (\boldsymbol{\eta}^k)^\mathsf{T} \boldsymbol{H} + \boldsymbol{b}^\mathsf{T} \right) \boldsymbol{\eta}^k,$$

where $\boldsymbol{b} = \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H} \boldsymbol{\eta}^0$ for notational convenience; see Appendix A.5. As the gradient $\nabla f(\boldsymbol{\eta}^0)$ and the Hessian $\boldsymbol{H}$ have already been calculated to obtain (2.18), the expression in (2.19) can be computed with low computational cost in every iteration of the MCMC chain.

In many applications, conditional independence assumption is imposed on the data density function. That is, there exists a partition of $\boldsymbol{\eta}$ into subvectors $\boldsymbol{\eta}_i$, such that

$$\pi(\boldsymbol{y} \mid \boldsymbol{\eta}) = \prod_i \pi_i(\boldsymbol{y}_i \mid \boldsymbol{\eta}_i).$$

In some cases, a proposal density based on the Gaussian approximation in (2.18) can be a poor approximation of the conditional posterior density in some partition of $\boldsymbol{\eta}$. Updating the whole vector $\boldsymbol{\eta}$ in one block may then result in the MCMC chain getting stuck, thus leading to lower computational efficiency. In order to circumvent this issue and to retain the computational speed gained by using the Gaussian approximation in (2.18) as a proposal density, a modification can be made to the sampling scheme which utilizes the conditional independence of the partitions within the data-rich block. The details of the modification can be seen in the Supplementary Material (Geirsson et al., 2020).

The proposed sampling scheme for the data-poor block is outlined in Algorithm 2. Note that by choosing $I = 1$ in Algorithm 2, the above sampling scheme without the conditional independence assumptions on the likelihood is obtained, while selecting $I \geq 2$ in Algorithm 2 assumes the aforementioned partitioning of $\boldsymbol{\eta}$ and that each $\boldsymbol{\eta}_i$ is accepted or rejected separately. Note that when $I \geq 2$ steps 5–14 can be performed in parallel across I clusters to speed up the computation.

3. EXAMPLE

One example of a latent Gaussian model is presented in this section to demonstrate how the LGM split sampler can be applied to obtain posterior samples. In this example, we present a simulation study on extreme events. The dataset consists of simulations of monthly maximum instantaneous flow based on characteristics of ten river catchments across Iceland. The simulated time series were chosen to represent 150 years. Boxplots of the distribution of the simulated data can be seen in Figure 1.

Two characterizing features were chosen to simulate the time series from each river, namely, river catchment area and maximum daily precipitation index, along with the month index, as these are known to be positively correlated with maximum instantaneous flow; see Davíðsson (2015) and Crochet (2012). The maximum daily precipitation index for a given catchment is based on the maximum monthly precipitation over the catchment. Then an average of these monthly maximum values is taken over a reference period. So, the maximum daily precipitation index varies with month while the river catchment area stays the same over months as expected. The precipitation is computed with a linear orographic precipitation model (Crochet et al., 2007) which provides 24-h precipitation on a grid with 1 km resolution.
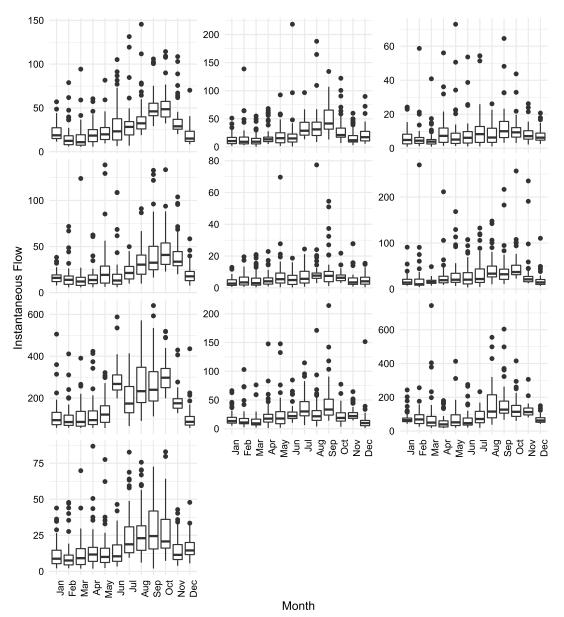
FIG. 1. *Boxplots of the simulated annual maximum instantaneous flow as a function of months for all of the ten river catchments.*

The values of the parameters used to simulate the time series were based on data from the Icelandic Meteorological Office on monthly maximum instantaneous flow from the ten aforementioned river catchments. In Davíðsson (2015), these data were analyzed, and estimates of parameters from that analysis were used as basis for generating the simulated time series analyzed in this section. The reasons for using simulated time series as opposed to the real time series are (i) the capability of verifying that the LGM sampler captures the underlying values, and (ii) a large fraction of the real time series were short, only between 20 to 40 observation per month, which is limiting for an extreme value analysis.

Model setup

The data level. The data were modeled with an LGM assuming the generalized extreme value distribution (GEV)

for the observations. To that extend, let $y_{mj,t}$ denote the value from river j at month m and year t, with a density function of the form

$$f(y_{mj,t})$$

$$= \frac{1}{\sigma_{mj}} \left(1 + \xi_{mj} \left(\frac{y_{mj,t} - \mu_{mj}}{\sigma_{mj}} \right) \right)^{-1 - \frac{1}{\xi_{mj}}}$$

$$\cdot \exp\left\{ -\left(1 + \xi_{mj} \left(\frac{y_{mj,t} - \mu_{mj}}{\sigma_{mj}} \right) \right)^{-\frac{1}{\xi_{mj}}} \right\}$$

if $1 + \xi_{mj}(y_{mj,t} - \mu_{mj})/\sigma_{mj} > 0$, and $f(y_{mj,t}) = 0$; otherwise, for $j = 1, \ldots, J$, $t = 1, \ldots, T$, $m = 1, \ldots, 12$. The parameters μ_{mj}, σ_{mj} and ξ_{mj} are the location, scale and shape parameters, respectively, of the GEV distribution for river j in month m. Additionally, J is the number of rivers and T is the number of years. Furthermore, the

Algorithm 2 The proposed algorithm for obtaining the $(k+1)$th sample from $\pi(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ in the data-rich block. By choosing $I = 1$, the sampling scheme introduced in Section 2.3.2 is obtained. For $I \geq 2$, the modified sampling scheme, which is derived in the Supplementary Material (Geirsson et al., 2020), is obtained for the partitions

Input: $(\boldsymbol{v}^{k+1}, \boldsymbol{\theta}^{k+1}, \boldsymbol{\eta}^k)$
1: Find the mode $\boldsymbol{\eta}^0 = \arg\max \log \pi(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{v}^{k+1}, \boldsymbol{\theta}^{k+1})$
2: Calculate $\boldsymbol{H} = \nabla^2 f(\boldsymbol{\eta}^0)^{\boldsymbol{\eta}}$ and $\boldsymbol{b} = \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0$
3: Sample $\boldsymbol{\eta}^* \sim \mathcal{N}(\boldsymbol{\eta}^0, (\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1})$
4: Calculate $\boldsymbol{\rho}(\boldsymbol{\eta}^k)$ and $\boldsymbol{\rho}(\boldsymbol{\eta}^*)$, where

$$\boldsymbol{\rho}(\boldsymbol{\eta}) = \left(-\frac{1}{2}\boldsymbol{\eta}^\mathsf{T}\boldsymbol{H} - \boldsymbol{b}^\mathsf{T}\right) \circ \boldsymbol{\eta}$$

    and $\circ$ denotes an entrywise product
5: **for** $i = 1, \ldots, I$
6:      Calculate $r_i = f_i(\boldsymbol{\eta}_i^*) + \boldsymbol{\rho}(\boldsymbol{\eta}^*)_i^\mathsf{T}\boldsymbol{1} - (f_i(\boldsymbol{\eta}_i^k) + \boldsymbol{\rho}(\boldsymbol{\eta}^k)_i^\mathsf{T}\boldsymbol{1})$
7:      Calculate $\alpha_i = \min\{1, \exp r_i\}$
8:      Sample $u_i \sim \mathcal{U}(0, 1)$
9:      **if** $\alpha_i > u_i$
10:         $\boldsymbol{\eta}_i^{k+1} = \boldsymbol{\eta}_i^*$
11:      **else if** $\alpha_i < u_i$
12:         $\boldsymbol{\eta}_i^{k+1} = \boldsymbol{\eta}_i^k$
13:      **end if**
14: **end for**
Output: $\boldsymbol{\eta}^{k+1}$

data are assumed independent between rivers and between months.

The latent level. The location and scale parameters are modeled on a logarithmic scale at the latent level, which is modeling setup along the lines presented in Cunnane and Nash (1971) and GREHYS (1996). Thus, define $\lambda_{mj} = \log \mu_{mj}$ and $\tau_{mj} = \log \sigma_{mj}$. The shape parameter is modeled on its native scale.

As discussed in Davíðsson (2015), the underlying processes of monthly maximum instantaneous flow exhibit a seasonal behavior. Therefore, the following seasonal model is proposed for the location parameter on a logarithmic scale. That is,

$$(3.1) \quad \begin{aligned} \lambda_{mj} = {}& \beta_{0,\lambda} + u_{0,m,\lambda} + x_{1,mj}(\beta_{1,\lambda} + u_{1,m,\lambda}) \\ & + \cdots + x_{p,mj}(\beta_{p,\lambda} + u_{p,m,\lambda}) + \varepsilon_{mj,\lambda}, \end{aligned}$$

where $\beta_{0,\lambda}$ denotes an overall intercept term; $x_{i,mj}$ denotes the ith covariate in month m at the jth river; $\beta_{i,\lambda}$ denotes the weight of the ith covariate for $i = 1, \ldots, p$; $u_{0,m,\lambda}$ denotes the seasonal random effect of the mth month; $u_{i,m,\lambda}$ denotes the seasonal additional weight of the ith covariate within month m; and $\varepsilon_{mj,\lambda}$ denotes an unstructured random effect.

In order to write the model in a matrix form for the implementation of the LGM split sampler, combine the location parameters for river j over months. That is,

$$\boldsymbol{\lambda}_j = (\lambda_{1j}, \ldots, \lambda_{12j})^\mathsf{T}, \quad j = 1, \ldots, J$$

and define the following:

$$\boldsymbol{u}_{i,\lambda} = (u_{i,1,\lambda}, \ldots, u_{i,12,\lambda})^\mathsf{T},$$
$$\boldsymbol{A}_{i,j} = \mathrm{diag}(x_{i,1j}, \ldots, x_{i,12j}),$$

where $i = 0, \ldots, p$ and $x_{0,mj} = 1$ denotes the intercept term for river j and month m. Additionally, define

$$\boldsymbol{X}_j = \begin{pmatrix} 1 & x_{1,1j} & \cdots & x_{p,1j} \\ 1 & x_{1,2j} & \cdots & x_{p,2j} \\ \vdots & & \vdots & \\ 1 & x_{1,12j} & \cdots & x_{p,12j} \end{pmatrix},$$
$$\boldsymbol{A}_j = (\boldsymbol{A}_{0,j}, \ldots, \boldsymbol{A}_{p,j}).$$

The seasonal model presented in (3.1) for the log-location parameter for river j can be written in matrix form as

$$\boldsymbol{\lambda}_j = \boldsymbol{X}_j\boldsymbol{\beta}_\lambda + \boldsymbol{A}_j\boldsymbol{u}_\lambda + \boldsymbol{\varepsilon}_{j,\lambda},$$

where $\boldsymbol{\beta}_\lambda = (\beta_{0,\lambda}, \ldots, \beta_{p,\lambda})^\mathsf{T}$, $\boldsymbol{u}_\lambda = (\boldsymbol{u}_{0,\lambda}, \ldots, \boldsymbol{u}_{p,\lambda})^\mathsf{T}$, and $\boldsymbol{\varepsilon}_{j,\lambda} = (\varepsilon_{1j,\lambda}, \ldots, \varepsilon_{12j,\lambda})^\mathsf{T}$. By combing the seasonal model over rivers, the following holds:

$$\boldsymbol{\lambda} = \boldsymbol{X}\boldsymbol{\beta}_\lambda + \boldsymbol{A}\boldsymbol{u}_\lambda + \boldsymbol{\varepsilon}_\lambda,$$

where

$$\boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \vdots \\ \boldsymbol{\lambda}_J \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_J \end{pmatrix},$$
$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_J \end{pmatrix}, \qquad \boldsymbol{\varepsilon}_\lambda = \begin{pmatrix} \boldsymbol{\varepsilon}_{1,\lambda} \\ \vdots \\ \boldsymbol{\varepsilon}_{J,\lambda} \end{pmatrix}.$$

Analogous model structure was also implemented for the log-scale parameter. That is,

$$\boldsymbol{\tau} = \boldsymbol{X}\boldsymbol{\beta}_\tau + \boldsymbol{A}\boldsymbol{u}_\tau + \boldsymbol{\varepsilon}_\tau.$$

A reduced model with a similar structure was implemented for the shape parameter ξ. That is,

$$(3.2) \quad \xi_{mj} = \beta_{0,\xi} + u_{0,m,\xi} + \varepsilon_{mj,\xi},$$

where $\beta_{0,\xi}$ denotes an overall intercept term; $u_{0,m,\xi}$ denotes the seasonal random effect of the mth month; and $\varepsilon_{mj,\xi}$ denotes an unstructured random effect. The full matrix model for ξ becomes

$$\boldsymbol{\xi} = \boldsymbol{1}_{12J}\beta_{0,\xi} + (\boldsymbol{1}_J \otimes \boldsymbol{I}_{12})\boldsymbol{u}_\xi + \boldsymbol{\varepsilon}_\xi,$$

where $\boldsymbol{1}_n$ denotes an n-dimensional vector of ones, and $\otimes$ is the Kronecker product.

Working within the LGM framework, the following prior density functions were assigned to the latent parameters. First assign

$$\pi(\boldsymbol{\beta}_\lambda) = \mathcal{N}(\boldsymbol{\beta}_\lambda \mid \mathbf{0}, \sigma_{\beta\lambda}^2 \boldsymbol{I}),$$

$$\pi(\boldsymbol{\beta}_\tau) = \mathcal{N}(\boldsymbol{\beta}_\tau \mid \mathbf{0}, \sigma_{\beta\tau}^2 \boldsymbol{I}),$$

$$\pi(\beta_\xi) = \mathcal{N}(\beta_\xi \mid 0, \sigma_{\beta\xi}^2).$$

The parameters $\boldsymbol{\beta}_\lambda$, $\boldsymbol{\beta}_\tau$ and β_ξ are assumed a priori to have a low precision on their native scales in order to let the data play the dominate role in their inference. Thus, the parameter values $\sigma_{\beta\lambda} = 4$, $\sigma_{\beta\tau} = 4$ and $\sigma_{\beta\xi} = 2$ were chosen for the prior density functions.

Secondly, the selection of prior density functions for the seasonal random effects needs to incorporate a correlation structure that induces a strong correlation between neighboring months. This is achieved by assigning the following prior density functions:

$$\pi(\boldsymbol{u}_\lambda) = \mathcal{N}(\boldsymbol{u}_\lambda \mid \mathbf{0}, \operatorname{diag}(\boldsymbol{\psi}_\lambda) \otimes \boldsymbol{Q}_u^{-1}),$$

$$\pi(\boldsymbol{u}_\tau) = \mathcal{N}(\boldsymbol{u}_\tau \mid \mathbf{0}, \operatorname{diag}(\boldsymbol{\psi}_\tau) \otimes \boldsymbol{Q}_u^{-1}),$$

$$\pi(\boldsymbol{u}_\xi) = \mathcal{N}(\boldsymbol{u}_\xi \mid \mathbf{0}, \psi_\xi \boldsymbol{Q}_u^{-1}),$$

where $\boldsymbol{\psi}_\lambda = (\psi_{0,\lambda}, \ldots, \psi_{p,\lambda})^\mathsf{T}$, $\boldsymbol{\psi}_\tau = (\psi_{0,\tau}, \ldots, \psi_{p,\tau})^\mathsf{T}$ and ψ_ξ serve as scaling parameters for the monthly random effects corresponding to the three intercepts and the covariates; and $\boldsymbol{Q}_u(\kappa)$ is a 12×12 circular band precision matrix that has the vector

$$\begin{bmatrix} 1 & -2(\kappa^2 + 2) & \kappa^4 + 4\kappa^2 + 6 & -2(\kappa^2 + 2) & 1 \end{bmatrix}$$

on the diagonal band, as discussed in Lindgren, Rue and Lindström (2011), which captures the autocorrelation between months. In this example, the decay parameters was fixed to simplify the inference and set equal to $\kappa = 1$. Furthermore, this value of κ induces an autocorrelation a priori between consecutive months. Third, for the unstructured random effects, the following priors were chosen:

$$\pi(\boldsymbol{\varepsilon}_\lambda) = \mathcal{N}(\boldsymbol{\varepsilon}_\lambda \mid \mathbf{0}, \sigma_{\varepsilon\lambda}^2 \boldsymbol{I}),$$

$$\pi(\boldsymbol{\varepsilon}_\tau) = \mathcal{N}(\boldsymbol{\varepsilon}_\tau \mid \mathbf{0}, \sigma_{\varepsilon\tau}^2 \boldsymbol{I}),$$

$$\pi(\boldsymbol{\varepsilon}_\xi) = \mathcal{N}(\boldsymbol{\varepsilon}_\xi \mid \mathbf{0}, \sigma_{\varepsilon\xi}^2 \boldsymbol{I}).$$

The hyperparameter level. Let $\boldsymbol{\theta}$ denote all the hyperparameters of the model that are not fixed. They are on a logarithmic scale and the parameters are parameterized as precision parameters for computational purposes. That is,

$$\boldsymbol{\theta} = (\log \psi_{0,\lambda}^{-1}, \ldots, \log \psi_{p,\lambda}^{-1}, \log \psi_{0,\tau}^{-1}, \ldots, \log \psi_{p,\tau}^{-1},$$

$$\log \psi_\xi^{-1}, \log \sigma_{\varepsilon\lambda}^{-2}, \log \sigma_{\varepsilon\tau}^{-2}, \log \sigma_{\varepsilon\xi}^{-2})$$

Gaussian prior distributions with fixed parameters were assigned to the hyperparmeters in $\boldsymbol{\theta}$.

Posterior inference. The data-rich block includes $\boldsymbol{\eta} = (\boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\xi})$ and the data-poor block consists of $\boldsymbol{v} = (\boldsymbol{\beta}_\lambda, \boldsymbol{u}_\lambda, \boldsymbol{\beta}_\tau, \boldsymbol{u}_\tau, \beta_\xi, \boldsymbol{u}_\xi)$ and the hyperparameters $\boldsymbol{\theta}$. For the implementations of the LGM split sampler, define $\boldsymbol{\mu}_v = \mathbf{0}$ and the following sparse matrices:

$$(3.3) \quad \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{A} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{X} & \boldsymbol{A} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \boldsymbol{1}_{12J} & \boldsymbol{I}_{12J} \end{pmatrix},$$

$$\boldsymbol{Q}_\varepsilon = \begin{pmatrix} \sigma_{\varepsilon\lambda}^{-2} \boldsymbol{I} & \cdot & \cdot \\ \cdot & \sigma_{\varepsilon\tau}^{-2} \boldsymbol{I} & \cdot \\ \cdot & \cdot & \sigma_{\varepsilon\xi}^{-2} \boldsymbol{I} \end{pmatrix}$$

and

$$(3.4) \quad \boldsymbol{Q}_v^{-1} = \operatorname{bdiag}(\sigma_{\beta\lambda}^2 \boldsymbol{I}, \operatorname{diag}(\boldsymbol{\psi}_\lambda) \otimes \boldsymbol{Q}_u^{-1}, \sigma_{\beta\tau}^2 \boldsymbol{I},$$

$$\operatorname{diag}(\boldsymbol{\psi}_\tau) \otimes \boldsymbol{Q}_u^{-1}, \sigma_{\beta\xi}^2, \psi_\xi \boldsymbol{Q}_u^{-1}),$$

where bdiag denotes a block diagonal matrix.

Data-poor block. The sampling scheme outlined in Section 2.3.1 was used to obtain MCMC samples from the conditional posterior $\pi(\boldsymbol{v}, \boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\eta})$ in the data-poor block. A proposal density based on the Gaussian distribution centered on the last draw of $\boldsymbol{\theta}$, as discussed in Roberts, Gelman and Gilks (1997), was selected for Algorithm 1, with a precision matrix $c^{-1}(-\boldsymbol{H})$ where $\boldsymbol{H}$ is a finite difference estimate of the Hessian matrix of $\log \pi(\boldsymbol{\theta}|\hat{\boldsymbol{\eta}})$ evaluated at the mode. That is,

$$(3.5) \quad \boldsymbol{H} \approx \nabla^2 \log \pi(\boldsymbol{\theta}|\hat{\boldsymbol{\eta}})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

where $\hat{\boldsymbol{\eta}}$ is the maximum likelihood estimate of $\boldsymbol{\eta}$ for each river and month; $\boldsymbol{\theta}_0$ is the mode of $\log \pi(\log \boldsymbol{\theta}|\hat{\boldsymbol{\eta}})$; and c is a scaling constant. Conditioning on $\hat{\boldsymbol{\eta}}$, as opposed to $\boldsymbol{\eta}^{k+1}$ for example, removes the necessity to estimate $\boldsymbol{H}$ in every iteration. Moreover, setting a specific scaling constant c removes the need for tuning. The scaling constant $c = 2.382^2/\dim(\boldsymbol{\theta})$ was implemented, as it is optimal in a particular scenario; see Roberts, Gelman and Gilks (1997). The resulting proposal density therefore becomes

$$(3.6) \quad q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k) = \mathcal{N}(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k, c(-\boldsymbol{H})^{-1}).$$

Algorithm 1 was thus implemented to obtain MCMC samples from the conditional posterior within the data-poor block, with $\boldsymbol{\mu}_v = \mathbf{0}$; $\boldsymbol{Z}$, $\boldsymbol{Q}_\varepsilon$ as in equation (3.3); $\boldsymbol{Q}_v$ as in equation (3.4); and the proposal density in (3.6).

Data-rich block. The modified version of the sampling scheme in Section 2.3.2, outlined in the Supplementary Material (Geirsson et al., 2020), was used to obtain MCMC samples from the conditional posterior $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$. The logarithm of the conditional posterior

is of the same form as in equation (2.17), with $\mathbf{Z}$ and $\mathbf{Q}_\varepsilon$ defined in equation (3.3) and

$$f(\boldsymbol{\eta}) = \sum_{m=1}^{12} \sum_{j=1}^{J} f_{mj}(\boldsymbol{\eta}_{mj})$$

(3.7)
$$= \sum_{m=1}^{12} \sum_{j=1}^{J} \sum_{t=1}^{T} \log \pi_{\text{gev}}(y_{mj,t} \mid \exp \lambda_{mj},$$

$$\exp \tau_{mj}, \xi_{mj}),$$

where π_{gev} denotes the density function of the generalized extreme value distribution and $\boldsymbol{\eta}_{mj} = (\lambda_{mj}, \tau_{mj}, \xi_{mj})$. Therefore, Algorithm 2 was used to obtain MCMC samples from the conditional posterior from the data-rich block, with $I = J \cdot 12 = 120$ and $f(\boldsymbol{\eta})$ as in (3.7).

Convergence diagnostics. The following convergence diagnostics are based on four MCMC chains sampled in parallel with the LGM split sampler from the proposed model. Each chain was calculated with 50,000 iterations where 10,000 iterations were burned in. Runtime, on a modern desktop (Ivy Bridge Intel Core i7-3770K, 16 GB RAM and a solid state hard drive), was approximately 7 hours. All calculations were carried out using R.

Gelman–Rubin plots, autocorrelation plots, trace plots and running mean plots for nine model parameters based on the MCMC run are shown in the Supplementary Material (Geirsson et al., 2020). All four plots are based on the same set of parameters and arranged identically. Three parameters were chosen from the location, scale and shape structures of the proposed model which were placed in the first, second and third rows of the figures. The first columns are based on parameters from the data-rich part of the latent field; the second columns are based on parameters from the data-poor part of the latent field; and the third column is based on hyperparameters.

The Gelman–Rubin plots in the Supplementary Material (Geirsson et al., 2020) show that the sampler has converged in the mean after roughly 10.000 iterations. Similar results hold for all the model parameters (results not shown). Furthermore, the autocorrelation plots in the Supplementary Material (Geirsson et al., 2020) demonstrate that the MCMC chains for the parameters from the data-rich and the data-poor parts of the latent field, exhibit a negligible autocorrelation after lag 10 and lag 20, respectively. The hyperparameters show a negligible autocorrelation after lag 30. Trace plots show the four MCMC have mixed well, and the running mean plots show that the four MCMC chains have converged in the mean; see the Supplementary Material (Geirsson et al., 2020).

Relying on these results, the MCMC chains exhibit all signs of having converged. Moreover, these results further indicate that the LGM split sampler, with the modified proposal density of Roberts, Gelman and Gilks (1997) implied by equation (3.6) for the hyperparameters, is highly computationally efficient in both the data-rich and data-poor blocks.

The left panel in Figure 2 compares the empirical cumulative distribution from river $j = 1$ in January with its posterior cumulative distribution functions based on the MCMC runs. The right panel shows the corresponding probability–probability plots. These results demonstrate that the LGM split sampler recaptures the known underlying values of the model parameters, which were used to generate the simulated data. Analogous results hold across all rivers and months (results not shown).

Since the data were generated from a known model setup, the results of the inference based on the MCMC runs can be compared to the known values of the model parameters. In Figure 3, the known values of the seasonal random effects are shown along with the corresponding 95% posterior intervals. The top panel in Figure 3 shows
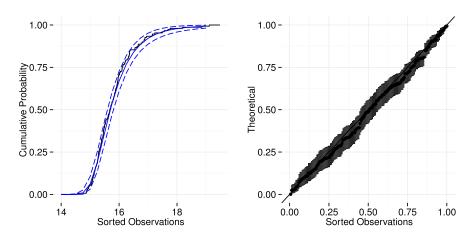

FIG. 2. *The left panel shows the empirical cumulative distribution of maximum instantaneous flow from river $j = 1$ in January (black solid curve) and the posterior mean of the corresponding cumulative distribution function (blue solid curve) and corresponding 95% posterior intervals (blue dashed curve). The right panel shows a probability–probability plot of maximum instantaneous flow from river $j = 1$ in January, along with 95% posterior intervals.*
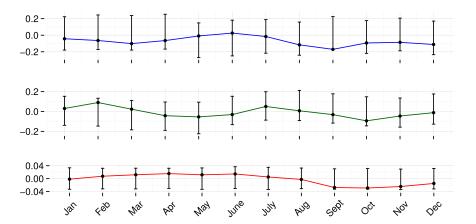
FIG. 3. *The top panel shows the known value (denoted with the blue entries) of the seasonal random effect $u_{0,m,\lambda}$ for the log-location parameter λ as function of month, m. The middle panel shows the known value (denoted with the green entries) of the seasonal random effect $u_{0,m,\tau}$ for log-scale parameter τ as function of month, m. The bottom panel shows the known value (denoted with the red entries) of the seasonal random effect $u_{0,m,\xi}$ for the shape parameter ξ as function of month, m. The errors bars in all panels represent the corresponding 95% posterior intervals based on the MCMC-runs.*

TABLE 1

The 95% posterior intervals from the MCMC run of all the covariate coefficients. The last column shows the underlying values which were used in the simulation of the data

Covariate coefficient	0.025 posterior quantile	0.975 posterior quantile	True value
$\beta_{0,\lambda}$	-5.22	-4.21	-4.30
$\beta_{1,\lambda}$	0.85	0.97	0.89
$\beta_{2,\lambda}$	0.71	0.94	0.74
$\beta_{0,\tau}$	-4.99	-4.40	-4.50
$\beta_{1,\tau}$	0.64	0.73	0.69
$\beta_{2,\tau}$	-0.05	0.09	0.00
$\beta_{0,\xi}$	0.08	0.14	0.10

TABLE 2

The 95% posterior interval from the MCMC run of all the hyperparameters in $\boldsymbol{\theta}$ on their original parameterization. The last column shows the underlying values which were used in the simulation of the data. The star denotes that the 95% posterior interval did not capture the true value

Hyper-parameter	0.025 posterior quantile	0.975 posterior quantile	True value
$\psi_{0,\lambda}$	0.00175	0.04915	0.01429
$\psi_{1,\lambda}$	0.00050	0.00476	0.00200
$\psi_{2,\lambda}$	0.00017	0.00324	0.00100
$\psi_{0,\tau}$	0.00063	0.02014	0.00500
$\psi_{1,\tau}$	0.00031	0.00243	0.00100
$\psi_{2,\tau}$	0.00013	0.00252	0.00050
$\psi_{0,\xi}$	0.00003	0.00088	0.00020
$\sigma^2_{\varepsilon\lambda}$	0.03452	0.06142	*0.06250
$\sigma^2_{\varepsilon\tau}$	0.00708	0.01536	0.01000
$\sigma^2_{\varepsilon\xi}$	0.00002	0.00674	0.000025

this comparison for the seasonal random effect $u_{0,m,\lambda}$ for the log-location parameter λ as a function of months. The middle and the bottom panels in Figure 3 show the same comparison for $u_{0,m,\tau}$ for the log-scale parameter and $u_{0,m,\xi}$ for the shape parameter ξ, respectively. The results reveal that the 95% posterior intervals for the seasonal random effects contain their known values. These results demonstrate that the LGM split sampler recaptures the known seasonal random effects. Furthermore, Tables 1 and 2 show the 95% posterior intervals of the main regression parameters and the hyperparameters and the underlying values of the parameters. These two tables reveal that the 95% posterior intervals of these parameters capture their underlying values except for $\sigma^2_{\varepsilon\lambda}$. In the case of model parameters not presented in Tables 1 and 2, their 95% posterior intervals capture their underlying values for 90% of them (results now shown).

4. DISCUSSION

In this paper, we presented extended LGMs that are such that each data point depends on more than one linear predictor at the latent level and proposed a novel sampler for the corresponding posterior density. This sampler has not been presented before and is referred to as the LGM split sampler. It is based on the sampler of Knorr-Held and Rue (2002). The LGM split sampler places the unknown parameters into two blocks in such a way that efficient sampling is achieved for each block by utilizing the particular structure of the extended LGMs and the Gaussian assumption at the latent level.

In the data-rich block, we proposed a Metropolis–Hastings algorithm with an independence proposal density which was constructed with a Gaussian approximation of the conditional posterior density evaluated at its mode. Furthermore, we proposed a modification for the sampler which is applicable if conditional independence assumptions are imposed on the data density function.

The modification can potentially increase the computational efficiency of the sampler, as discussed in the Supplementary Material (Geirsson et al., 2020).

Although the proposed sampler in the data-rich block is computationally efficient, it is only applicable in practice if the mode of conditional posterior density function can be found, and can be calculated reasonably fast. For example, in the case of models where each observed data point has more than one unique data density parameter associated with it, say of the type $y_i \sim \pi(y_i|\mu_i, \sigma_i)$ for every measurement i, finding the mode of the conditional posterior $\pi(\mu_i, \sigma_i|y_i)$ becomes computationally impractical in some cases. Models of this type include, for example, certain spatial temporal models (Hrafnkelsson, Morris and Baladandayuthapani, 2012). Similar computational issues also arise if data dependence at the data-level of an LGM is desired; see, for example, Davison, Padoan and Ribatet (2012) where t-copulas are implemented with GEV marginal density functions at the data level as a model for spatial extremes. However, in both of the aforementioned cases different sampling schemes for the data-rich block can be implemented without changing the sampling scheme of choice in the data-poor block. For example, a sampling scheme based on a MALA and HMC-type algorithm are well suited for the structure of the data-rich block in both cases.

In the data-poor block, the conditional posterior density $\pi(\nu|\eta, \theta)$ is a Gaussian of the form in (2.10) and invariant of the data density function. These results serves as one of the main computational advantage introduced by the LGM split sampler due to the following reasons. First, as the conditional posterior $\pi(\nu|\eta, \theta)$ is Gaussian, a modified version of the single block sampler of Knorr-Held and Rue (2002), which is known to be a highly efficient sampling scheme when applicable (Filippone, Zhong and Girolami, 2013), becomes applicable within the data-poor block regardless of the data-density function at the data level. As a consequence, computationally efficient sampling algorithms can be used to sample from the exact Gaussian conditional posterior density $\pi(\nu|\eta, \theta)$. Furthermore, if the prior density functions in (2.6) have a sparse GMRF precision structure, then $\pi(\nu|\eta, \theta)$ preserves the sparse structure as discussed in Section 2.3.1, which in turn allows for highly efficient sampling algorithms for the Gaussian density $\pi(\nu|\eta, \theta)$. Therefore, the proposed sampling scheme in Section 2.3.1 for the data-poor blocks scales well in terms of computational speed and efficiency with increasing dimension of the data-poor part of the latent field, which is of great importance to achieve, especially in the field of spatial statistics. It was shown that the density $\pi(\theta|\eta)$ also preserves the sparse GMRF precision structure when it is assumed a priori in (2.6). Since the densities $\pi(\nu|\eta, \theta)$ and $\pi(\theta|\eta)$ are used to sample (ν, θ) from the conditional posterior density $\pi(\nu, \theta|y, \eta)$, then

the sparse GMRF structure assumed in (2.6) is preserved, and thus supporting fast computation within the entire data-poor block.

Second, as the conditional posterior density $\pi(\nu|\eta, \theta)$ is a known Gaussian and the acceptance rate in the sampling scheme for the data-poor block in Section 2.3.1 only depends on the hyperparameters, the computational efficiency of the proposed sampling scheme for the data-poor block is only dependent on the sampling scheme used for the hyperparameters. In this sense, the sampling scheme in Section 2.3.1 is in itself modular, that is, any proposal density for the hyperparmeters is applicable. Choosing a computationally efficient sampling scheme for the hyperparameters can thus increase the computational efficiency of the overall sampling scheme within the data-poor block, as demonstrated in the example in Section 3. In this example, we proposed a modified version of the sampling scheme of Roberts, Gelman and Gilks (1997) implied by equation (3.5), which resulted in rapidly decreasing autocorrelation in the MCMC chains.

Due to the modularity of the LGM split sampler, sampling schemes for the data-rich block can be developed and improved independently of the sampler in the data-poor block, and vice versa. Additionally, as the conditional posterior density $\pi(\nu|\eta, \theta)$ in the data-poor block becomes invariant of the data, the computational advantages introduced by the conditional posterior structure in the data-poor block hold for all LGMs. Moreover, the LGM split sampler can be applied to various LGMs as it is designed to handle LGMs where latent models Gaussian models are imposed on more than just the mean structure of the data density function. Thus, in our view, further developing and improving sampling schemes that utilize the computational advantages introduced by the LGM split sampler presents an interesting area of future research.

APPENDIX: PROOFS FOR RESULTS IN THE MAIN BODY TEXT

A.1

PROOF.   By known results about Gaussian distributions and inverses of block matrices, the joint distribution of (η, ν) is given by

$$\pi\begin{pmatrix} \eta \\ \nu \end{pmatrix} = \mathcal{N}\left(\begin{pmatrix} \eta \\ \nu \end{pmatrix} \middle| \begin{pmatrix} Z\mu_\nu \\ \mu_\nu \end{pmatrix}, \right.$$

$$\left. \begin{pmatrix} Q_\varepsilon & -Q_\varepsilon Z \\ -Z^\mathsf{T} Q_\varepsilon & Q_\nu + Z^\mathsf{T} Q_\varepsilon Z \end{pmatrix}^{-1} \right)$$

$$= \mathcal{N}\left(\begin{pmatrix} \eta \\ \nu \end{pmatrix} \middle| \begin{pmatrix} Z\mu_\nu \\ \mu_\nu \end{pmatrix}, \right.$$

$$\left. \begin{pmatrix} Q_\varepsilon^{-1} + Z Q_\nu^{-1} Z^\mathsf{T} & Z Q_\nu^{-1} \\ Q_\nu^{-1} Z^\mathsf{T} & Q_\nu^{-1} \end{pmatrix} \right).$$

The conditional distribution of $\boldsymbol{v}$ conditioned on $\boldsymbol{\eta}$ follows directly from Lemma 2.1 in Rue and Held (2005), that is,

$$\pi(\boldsymbol{v} \mid \boldsymbol{\eta}) = \mathcal{N}(\boldsymbol{v} \mid \boldsymbol{Q}_{v|\eta}^{-1}(\boldsymbol{Q}_v \boldsymbol{\mu}_v + \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\varepsilon \boldsymbol{\eta}), \boldsymbol{Q}_{v|\eta}^{-1}),$$

where $\boldsymbol{Q}_{v|\eta} = \boldsymbol{Q}_v + \boldsymbol{Z}^\mathsf{T} \boldsymbol{Q}_\varepsilon \boldsymbol{Z}$. $\square$

A.2

PROOF. By definition of the proposal density in (2.11), the following holds:

$$\frac{q(\boldsymbol{v}^k, \boldsymbol{\theta}^k \mid \boldsymbol{v}^*, \boldsymbol{\theta}^*)}{q(\boldsymbol{v}^*, \boldsymbol{\theta}^* \mid \boldsymbol{v}^k, \boldsymbol{\theta}^k)} = \frac{\pi(\boldsymbol{v}^k \mid \boldsymbol{\eta}^k, \boldsymbol{\theta}^k) q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{\pi(\boldsymbol{v}^* \mid \boldsymbol{\eta}^k, \boldsymbol{\theta}^*) q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)},$$

where $q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^k)$ is some proposal density for $\boldsymbol{\theta}$ and $\pi(\boldsymbol{v} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$ is the conditional Gaussian density function in (2.10). Therefore, the acceptance ratio in (2.12) can be written as

$$\text{(A.1)} \quad \begin{aligned} & \frac{\pi(\boldsymbol{v}^*, \boldsymbol{\theta}^* \mid \boldsymbol{y}, \boldsymbol{\eta}^k)}{\pi(\boldsymbol{v}^k, \boldsymbol{\theta}^k \mid \boldsymbol{y}, \boldsymbol{\eta}^k)} \frac{\pi(\boldsymbol{v}^k \mid \boldsymbol{\eta}^k, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{v}^* \mid \boldsymbol{\eta}^k, \boldsymbol{\theta}^*)} \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)} \\ & = \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^k)}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^k)} \frac{q(\boldsymbol{\theta}^k \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^k)} \end{aligned}$$

since $\pi(\boldsymbol{v}, \boldsymbol{\theta} | \boldsymbol{y}, \boldsymbol{\eta}) = \pi(\boldsymbol{v}, \boldsymbol{\theta} | \boldsymbol{\eta})$, as discussed in Section 2.3.1, and $\pi(\boldsymbol{v}, \boldsymbol{\theta} \mid \boldsymbol{\eta})/\pi(\boldsymbol{v} \mid \boldsymbol{\eta}, \boldsymbol{\theta}) = \pi(\boldsymbol{\theta} \mid \boldsymbol{\eta})$ for any $\boldsymbol{\eta}$, $\boldsymbol{v}$ and $\boldsymbol{\theta}$. The result in (A.1) demonstrates that the acceptance ratio in (2.12) only depends on $\boldsymbol{\theta}$ within in the proposed setup. In other words, the acceptance ratio in (2.12) becomes independent of the value of $\boldsymbol{v}$. $\square$

A.3

PROOF. In order to rewrite $\pi(\boldsymbol{\theta} \mid \boldsymbol{\eta})$ in (2.13), we use the relation

$$\text{(A.2)} \qquad \pi(\boldsymbol{\theta} \mid \boldsymbol{\eta}) \propto \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\eta} \mid \boldsymbol{\theta}).$$

Furthermore, by the law of conditional probability, the following holds:

$$\text{(A.3)} \quad \pi(\boldsymbol{\eta} \mid \boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\eta}, \boldsymbol{v} \mid \boldsymbol{\theta})}{\pi(\boldsymbol{v} \mid \boldsymbol{\eta}, \boldsymbol{\theta})} = \frac{\pi(\boldsymbol{\eta} \mid \boldsymbol{v}, \boldsymbol{\theta}) \pi(\boldsymbol{v} \mid \boldsymbol{\theta})}{\pi(\boldsymbol{v} \mid \boldsymbol{\eta}, \boldsymbol{\theta})}.$$

As $\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})$ is independent of the value of $\boldsymbol{v}$, it follows that the two ratios in (A.3) are invariant of the choice of $\boldsymbol{v}$. In particular, the following holds:

$$\text{(A.4)} \qquad \pi(\boldsymbol{\eta} \mid \boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\eta} \mid \boldsymbol{0}, \boldsymbol{\theta}) \pi(\boldsymbol{0} \mid \boldsymbol{\theta})}{\pi(\boldsymbol{0} \mid \boldsymbol{\eta}, \boldsymbol{\theta})}$$

by choosing the value $\boldsymbol{v} = \boldsymbol{0}$. Combining (A.2) and (A.4) yield

$$\begin{aligned} \frac{\pi(\boldsymbol{\theta}^* \mid \boldsymbol{\eta}^k)}{\pi(\boldsymbol{\theta}^k \mid \boldsymbol{\eta}^k)} &= \frac{\pi(\boldsymbol{\theta}^*) \pi(\boldsymbol{\eta}^k \mid \boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k) \pi(\boldsymbol{\eta}^k \mid \boldsymbol{\theta}^k)} \\ &= \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^k)} \cdot \frac{\pi(\boldsymbol{\eta}^k \mid \boldsymbol{0}, \boldsymbol{\theta}^*) \pi(\boldsymbol{0} \mid \boldsymbol{\theta}^*)}{\pi(\boldsymbol{0} \mid \boldsymbol{\eta}^k, \boldsymbol{\theta}^*)} \\ &\quad \cdot \frac{\pi(\boldsymbol{0} \mid \boldsymbol{\eta}^k, \boldsymbol{\theta}^k)}{\pi(\boldsymbol{\eta}^k \mid \boldsymbol{0}, \boldsymbol{\theta}^k) \pi(\boldsymbol{0} \mid \boldsymbol{\theta}^k)}. \end{aligned}$$

Moreover, if the Gaussian prior density functions in (2.6) are GMRFs with sparse precision structures, then all of the conditional density functions on the rightmost side of (A.3) are GMRFs with sparse precision structures, by known results about conditioning on subvectors as demonstrated in Theorem 2.5 in Rue and Held (2005). $\square$

A.4

PROOF. The conditional posterior density function $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ is proportional to the product of the data density function and the conditional Gaussian prior density $\pi(\boldsymbol{\eta} \mid \boldsymbol{v}, \boldsymbol{\theta})$ given by (2.6), that is,

$$\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) \propto \pi(\boldsymbol{y} \mid \boldsymbol{\eta}) \pi(\boldsymbol{\eta} \mid \boldsymbol{v}, \boldsymbol{\theta}).$$

Thus, the logarithm of the conditional posterior density function is given by

$$\begin{aligned} \log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) = f(\boldsymbol{\eta}) &- \frac{1}{2} \boldsymbol{\eta}^\mathsf{T} \boldsymbol{Q}_\varepsilon \boldsymbol{\eta} \\ &+ (\boldsymbol{Q}_\varepsilon \boldsymbol{Z} \boldsymbol{v})^\mathsf{T} \boldsymbol{\eta} + \text{const}, \end{aligned}$$

where $f(\boldsymbol{\eta}) = \log \pi(\boldsymbol{y} \mid \boldsymbol{\eta})$ for notational convenience. The second-order Taylor approximation of $f(\boldsymbol{\eta})$ expanded around the mode $\boldsymbol{\eta}^0$ of the conditional posterior $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ is

$$\begin{aligned} f(\boldsymbol{\eta}) &\approx f(\boldsymbol{\eta}^0) + \nabla f(\boldsymbol{\eta}^0)^\mathsf{T} (\boldsymbol{\eta} - \boldsymbol{\eta}^0) \\ &\quad + \frac{1}{2} (\boldsymbol{\eta} - \boldsymbol{\eta}^0)^\mathsf{T} \boldsymbol{H} (\boldsymbol{\eta} - \boldsymbol{\eta}^0) \\ &= \frac{1}{2} \boldsymbol{\eta}^\mathsf{T} \boldsymbol{H} \boldsymbol{\eta} + (\nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H} \boldsymbol{\eta}^0)^\mathsf{T} \boldsymbol{\eta} + \text{const}. \end{aligned}$$

Consequently, the second-order Taylor approximation of $\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ expanded around $\boldsymbol{\eta}^0$ becomes

$$\begin{aligned} &\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) \\ &\approx \frac{1}{2} \boldsymbol{\eta}^\mathsf{T} \boldsymbol{H} \boldsymbol{\eta} + (\nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H} \boldsymbol{\eta}^0)^\mathsf{T} \boldsymbol{\eta} \\ &\quad - \frac{1}{2} \boldsymbol{\eta}^\mathsf{T} \boldsymbol{Q}_\varepsilon \boldsymbol{\eta} + (\boldsymbol{Q}_\varepsilon \boldsymbol{Z} \boldsymbol{v})^\mathsf{T} \boldsymbol{\eta} + \text{const} \\ &= -\frac{1}{2} \boldsymbol{\eta}^\mathsf{T} (\boldsymbol{Q}_\varepsilon - \boldsymbol{H}) \boldsymbol{\eta} + (\boldsymbol{Q}_\varepsilon \boldsymbol{Z} \boldsymbol{v} + \boldsymbol{b})^\mathsf{T} \boldsymbol{\eta} + \text{const}, \end{aligned}$$

where $\boldsymbol{b} = (\nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H} \boldsymbol{\eta}^0)$. This derivation yields a Gaussian approximation with a mean vector

$$(\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1} (\boldsymbol{Q}_\varepsilon \boldsymbol{Z} \boldsymbol{v} + \boldsymbol{b})$$

and covariance matrix $(\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1}$. However, as the vector $\boldsymbol{\eta}^0$ is the mode of the conditional posterior function $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$. the following relation holds:

$$\begin{aligned} &\nabla \log \pi(\boldsymbol{\eta}^0 \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) \\ &= \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{Q}_\varepsilon \boldsymbol{\eta}_0 + (\boldsymbol{Q}_\varepsilon \boldsymbol{Z} \boldsymbol{v})^\mathsf{T} = \boldsymbol{0}. \end{aligned}$$

The mean of the Gaussian approximations becomes

$$(\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1}(\boldsymbol{Q}_\varepsilon \boldsymbol{Z}\boldsymbol{v} + \boldsymbol{b})$$

$$= (\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1}(\boldsymbol{Q}_\varepsilon \boldsymbol{Z}\boldsymbol{v} + \nabla f(\boldsymbol{\eta}^0) - \boldsymbol{H}\boldsymbol{\eta}^0)$$

$$= (\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1}(\boldsymbol{Q}_\varepsilon \boldsymbol{\eta}^0 - \boldsymbol{H}\boldsymbol{\eta}^0)$$

$$= (\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1}(\boldsymbol{Q}_\varepsilon - \boldsymbol{H})\boldsymbol{\eta}^0 = \boldsymbol{\eta}^0.$$

Thus, a Gaussian approximation of the conditional posterior density function $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ evaluated at the mode $\boldsymbol{\eta}^0$ is given by

$$\tilde{\pi}(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\eta} \mid \boldsymbol{\eta}^0, (\boldsymbol{Q}_\varepsilon - \boldsymbol{H})^{-1}). \qquad \square$$

A.5

PROOF. The logarithm of the acceptance ratio given in (2.19) is

$$(\text{A.5}) \qquad r = \log \frac{\pi(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) q(\boldsymbol{\eta}^k)}{\pi(\boldsymbol{\eta}^k \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) q(\boldsymbol{\eta}^*)},$$

where $\pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta})$ is the conditional posterior density function given in (2.17) and $q(\boldsymbol{\eta})$ is the proposal density based on the Gaussian approximation in (2.18). The right-hand side term in (A.5) can be written as

$$\log \pi(\boldsymbol{\eta}^* \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) - \log q(\boldsymbol{\eta}^*)$$

$$- (\log \pi(\boldsymbol{\eta}^k \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) - \log q(\boldsymbol{\eta}^k)).$$

Since the proposal density q is based on the Gaussian approximation in (2.18), the following holds:

$$\log \pi(\boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}) - \log q(\boldsymbol{\eta})$$

$$= f(\boldsymbol{\eta}) - \frac{1}{2}\boldsymbol{\eta}^\top \boldsymbol{Q}_\varepsilon \boldsymbol{\eta} + (\boldsymbol{Q}_\varepsilon \boldsymbol{Z}\boldsymbol{v})^\top \boldsymbol{\eta}$$

$$- \left(\frac{1}{2}\boldsymbol{\eta}^\top \boldsymbol{H}\boldsymbol{\eta} + \boldsymbol{b}^\top \boldsymbol{\eta} - \frac{1}{2}\boldsymbol{\eta}^\top \boldsymbol{Q}_\varepsilon \boldsymbol{\eta} + (\boldsymbol{Q}_\varepsilon \boldsymbol{Z}\boldsymbol{v})^\top \boldsymbol{\eta}\right)$$

$$+ \text{const}$$

$$= f(\boldsymbol{\eta}) - \left(\frac{1}{2}\boldsymbol{\eta}^\top \boldsymbol{H}\boldsymbol{\eta} + \boldsymbol{b}^\top \boldsymbol{\eta}\right) + \text{const}$$

which yields the result in (2.20). $\square$

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

Supplement to "LGM Split Sampler: An Efficient MCMC Sampling Scheme for Latent Gaussian Models" (DOI: 10.1214/19-STS727SUPP; .pdf). This supplement contains results for the LGM split sampler in the case when the data density can be presented as a product of conditionally independent data densities along with proofs of these results.

REFERENCES

CHILÈS, J.-P. and DELFINER, P. (2012). *Geostatistics*: *Modeling Spatial Uncertainty*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2850475 https://doi.org/10.1002/9781118136188

COOLEY, D., NYCHKA, D. and NAVEAU, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.* **102** 824–840. MR2411647 https://doi.org/10.1198/016214506000000780

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1239641 https://doi.org/10.1002/9781119115151

CROCHET, P. (2012). Estimating the flood frequency distribution for ungauged catchments using an index flood procedure. Application to ten catchments in Northern Iceland. Technical Report No. VÍ 2012-005, The Icelandic Meteorological Office.

CROCHET, P., JÓHANNESSON, T., JÓNSSON, T., SIGURÐSSON, O., BJÖRNSSON, H., PÁLSSON, F. and BARSTAD, I. (2007). Estimating the spatial distribution of precipitation in Iceland using a linear model of orographic precipitation. *J. Hydrometeorol.* **8** 1285–1306.

CUNNANE, C. and NASH, J. E. (1971). Bayesian estimation of frequency of hydrological events. *Proc. Warsaw Symp. Math. Models Hydrol.* **1** 47–55.

DAVÍÐSSON, Ó. B. (2015). Bayesian flood frequency analysis using monthly maxima. Master's thesis, Univ. Iceland.

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980 https://doi.org/10.1214/11-STS376

DIGGLE, P. J., TAWN, J. A. and MOYEED, R. A. (1998). Model-based geostatistics. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **47** 299–350. MR1626544 https://doi.org/10.1111/1467-9876.00113

FAHRMEIR, L. and TUTZ, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. *Springer Series in Statistics*. Springer, New York. MR1284203 https://doi.org/10.1007/978-1-4899-0010-4

FERKINGSTAD, E., FRIGESSI, A., RUE, H., THORLEIFSSON, G. and KONG, A. (2008). Unsupervised empirical Bayesian multiple testing with external covariates. *Ann. Appl. Stat.* **2** 714–735. MR2524353 https://doi.org/10.1214/08-AOAS158

FILIPPONE, M., ZHONG, M. and GIROLAMI, M. (2013). A comparative evaluation of stochastic-based inference methods for Gaussian process models. *Mach. Learn.* **93** 93–114. MR3093120 https://doi.org/10.1007/s10994-013-5388-x

GEIRSSON, Ó. P., HRAFNKELSSON, B. and SIMPSON, D. (2015). Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid. *Environmetrics* **26** 339–353. MR3366968 https://doi.org/10.1002/env.2343

GEIRSSON, Ó. P., HRAFNKELSSON, B., SIMPSON, D. and SIGURDARSON, H. (2020). Supplement to "LGM split sampler: An efficient MCMC sampling scheme for latent Gaussian models." https://doi.org/10.1214/19-STS727SUPP.

GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 https://doi.org/10.1111/j.1467-9868.2010.00765.x

GREHYS (1996). Presentation and review of some methods for regional flood frequency analysis. *J. Hydrol.* **186** 63–84.

GUTTORP, P. and GNEITING, T. (2006). Studies in the history of probability and statistics. XLIX. On the Matérn correlation family. *Biometrika* **93** 989–995. MR2285084 https://doi.org/10.1093/biomet/93.4.989

HRAFNKELSSON, B., MORRIS, J. S. and BALADANDAYUTHAPANI, V. (2012). Spatial modeling of annual minimum and maximum temperatures in Iceland. *Meteorol. Atmos. Phys.* **116** 43–61.

KLEIN, N., KNEIB, T., LANG, S. et al. (2013). Bayesian structured additive distributional regression. Technical Report No. 2013-23, Department of Public Finance, Univ. Innsbruck.

KLEIN, N., KNEIB, T., KLASEN, S. and LANG, S. (2015). Bayesian structured additive distributional regression for multivariate responses. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64** 569–591. MR3367789 https://doi.org/10.1111/rssc.12090

KNEIB, T. (2013). Beyond mean regression. *Stat. Model.* **13** 275–303. MR3179527 https://doi.org/10.1177/1471082X13494159

KNORR-HELD, L. and RUE, H. (2002). On block updating in Markov random field models for disease mapping. *Scand. J. Stat.* **29** 597–614. MR1988414 https://doi.org/10.1111/1467-9469.00308

LAWSON, A. B. (2013). *Bayesian Disease Mapping*: *Hierarchical Modeling in Spatial Epidemiology*. *Interdisciplinary Statistics*. CRC Press, Boca Raton, FL. MR2484272

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

MARTINO, S., AAS, K., LINDQVIST, O., NEEF, L. R. and RUE, H. (2011). Estimating stochastic volatility models using integrated nested Laplace approximations. *Eur. J. Finance* **17** 487–503.

MARTINS, T. G., SIMPSON, D., LINDGREN, F. and RUE, H. (2013a). Bayesian computing with INLA: New features. *Comput. Statist.*

Data Anal. **67** 68–83. MR3079584 https://doi.org/10.1016/j.csda.2013.04.014

MARTINS, T. G., SIMPSON, D., ILLIAN, J. B., RUE, H. and GEIRSSON, Ó. P. (2013b). Discussion of 'Beyond mean regression' [MR3179527]. *Stat. Model.* **13** 355–361. MR3179533 https://doi.org/10.1177/1471082X13494527

NEAL, R. M. (1993). Probabilistic inference using Markov chain Monte Carlo methods. Technical Report No. CRG-TR-93-1, Department of Computer Science, Univ. Toronto.

PETTITT, A. N., WEIR, I. S. and HART, A. G. (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Stat. Comput.* **12** 353–367. MR1951708 https://doi.org/10.1023/A:1020792130229

RIGBY, R. A. and STASINOPOULOS, D. M. (2005). Generalized additive models for location, scale and shape. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **54** 507–554. MR2137253 https://doi.org/10.1111/j.1467-9876.2005.00510.x

ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields*: *Theory and Applications. Monographs on Statistics and Applied Probability* **104**. CRC Press/CRC, Boca Raton, FL. MR2130347 https://doi.org/10.1201/9780203492024

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

RUE, H., RIEBLER, A., SØRBYE, .S. H., ILLIAN, J. B., SIMPSON, D. P. and LINDGREN, F. K. (2017). Bayesian computing with INLA: A review. *Ann. Rev. Stat. Appl.* **4** 395–421.

SCHAEFLI, B., TALAMBA, D. B. and MUSY, A. (2007). Quantifying hydrological modeling errors through a mixture of normal distributions. *J. Hydrol.* **332** 303–315.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303 https://doi.org/10.1198/016214504000000241