

RESEARCH ARTICLE

WILEY

A Bayesian spatio-temporal model for short-term forecasting of precipitation fields

S. R. Johnson¹ | S. E. Heaps² | K. J. Wilson¹ | D. J. Wilkinson² 

¹School of Mathematics, Statistics & Physics, Newcastle University, Newcastle upon Tyne, UK

²Department of Mathematical Sciences, Durham University, Durham, UK

Correspondence

D. J. Wilkinson, Department of Mathematical Sciences, Durham University, South Road, Durham, DH1 3LE, UK.
Email: darren.j.wilkinson@durham.ac.uk

Funding information

Natural Environment Research Council

Abstract

With extreme weather events becoming more common, the risk posed by surface water flooding is ever increasing. In this work we propose a model, and associated Bayesian inference scheme, for generating short-term, probabilistic forecasts of localised precipitation on a spatial grid. Our generative hierarchical dynamic model is formulated in discrete space and time with a lattice-Markov spatio-temporal auto-regressive structure, inspired by continuous models of advection and diffusion. Observations from both weather radar and ground based rain gauges provide information from which we can learn the precipitation field through a latent process in addition to unknown model parameters. Working in the Bayesian paradigm provides a coherent framework for capturing uncertainty, both in the underlying model parameters and in our forecasts. Further, appealing to simulation based sampling using MCMC yields a straightforward solution to handling zeros, treated as censored observations, via data augmentation. Both the underlying state and the observations are of moderately large dimension ($\mathcal{O}(10^4)$ and $\mathcal{O}(10^3)$ respectively) and this renders standard inference approaches computationally infeasible. Our solution is to embed the ensemble Kalman smoother within a Gibbs sampling scheme to facilitate approximate Bayesian inference in reasonable time. Both the methodology and the effectiveness of our posterior sampling scheme are demonstrated via simulation studies and by a case study of real data from the Urban Observatory project based in Newcastle upon Tyne, UK.

KEYWORDS

advection-diffusion processes, dynamic spatio-temporal models, ensemble Kalman smoother, high-dimensional statistics, rainfall modeling

1 | INTRODUCTION

Increased urbanization coupled with the effects of global climate change means that cities are becoming increasingly susceptible to localized surface water flooding as a result of intense rainfall events (Barr et al., 2020; Kendon et al., 2014). Such events can result in substantial damage to private properties and businesses, and cause severe disruption to inner city transport systems and infrastructure. Recent computational advances have enabled the development of high-performance

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Environmetrics* published by John Wiley & Sons Ltd.

integrated hydrodynamic modeling systems for rainfall-induced surface water flows in urban areas (Liang & Smith, 2015; Xia & Liang, 2018; Xing et al., 2019). Such models rely on high-dimensional spatio-temporal inputs that characterize the amount of rainfall observed at ground level. Ground-based rain gauges are perhaps the natural choice for obtaining information. However, it is often not practical to deploy these at sufficient spatial resolution. Thus, in practice, inputs are typically obtained from weather radar systems, but these do not necessarily provide an accurate description of the rainfall rates on the ground.

In what follows we propose a physically motivated spatio-temporal statistical model that is capable of jointly modeling observations stemming from a (potentially sparse) rain gauge network and a weather radar, allowing the synthesis of these two different data sources. The basic premise is to leverage information from the weather radar, for example where the rain clouds are and where they are moving, and use the ground-based rain gauge information to calibrate the spatio-temporal field of interest. Data fusion based on joint modeling of multiple spatial data sets has been considered previously in the literature (Chatterjee et al., 2010; Chiu et al., 2013; Manzione & Castrignanò, 2019; Puttaswamy et al., 2014; Rundel et al., 2015), but fusion in the context of spatio-temporal modeling has received much less attention, and we do not know of previous approaches combining weather radar and rainfall gauge data for Bayesian rainfall field nowcasting. We apply our model to real data obtained from Newcastle's Urban Observatory project (James et al., 2014) that collects data from a large range of sensors distributed around the North East of England. This work forms part of a wider consortium of analytics and models developed as part of the NERC funded Flood-PREPARED (Predicting Rainfall Events by Physical Analytics of REaltime Data) project; further details are given by Barr et al. (2020).

The analysis of spatio-temporal data has a rich history in the statistical literature and a wealth of approaches have been proposed; in the context of applications to weather and climate, recent contributions include (Kleiber et al., 2023; Mastrantonio et al., 2019; Mastrantonio et al., 2022; Richards & Wadsworth, 2021), and for a more detailed overview of the field see Liu et al. (2021). The methods adopted by different authors vary substantially, since they are tailored to the available data and the question of inferential interest. Here we appeal to the hierarchical dynamical spatio-temporal modeling framework (Berliner, 2003; Katzfuss et al., 2020; Sigrist et al., 2012; Stroud et al., 2010; Wikle et al., 1998; Wikle & Cressie, 1999) which we find naturally appealing for several reasons. In particular, the modelling procedure can be decomposed into a sequence of conditional model specifications, of which the main components are a *system model* and an *observation model*. The system model describes how the (latent) process of interest evolves over space and time. The observations are modelled conditionally on this process, via the observation model, which provides a straightforward framework for handling observations from numerous data sources. The direct specification of a system model can be a rather daunting task due to the complex spatial and temporal dependencies that need to be captured. Given the physical nature of our problem we choose to model the dynamics of the system model via a generative auto-regressive process, loosely inspired by advection-diffusion-reaction dynamics, since that is the natural description of moisture transport in the atmosphere. This approach avoids the need for direct specification of space-time covariance functions (which can be difficult in practice) and gives rise to a tractable Gaussian model of vector auto-regressive (VAR) form (Cressie & Wikle, 2015; Krainski et al., 2018).

The dimension of both the observations and the spatio-temporal field of interest are moderately large and so solving the inference (inverse) problem is challenging. We consider an efficient Markov chain Monte Carlo (MCMC) scheme for sampling from the posterior distribution of interest. By opting for a simulation based inference approach we are able to straightforwardly handle both missing and zero (censored) observations by appealing to data augmentation (Tanner & Wong, 1987). More specifically, we introduce an appropriate collection of latent variables (Heaps et al., 2015; Rundel et al., 2015; Sansó & Guenni, 2000) that, when coupled with an appropriate prior distribution, gives rise to a tractable dynamic linear model (DLM) for the complete data. For these types of models state inference is typically achieved via the forwards-filtering backwards-sampling (FFBS) algorithm (Carter & Kohn, 1994; Frühwirth-Schnatter, 1994), however the dimension of our state-space renders (exact) simulation from the joint full conditional distribution of the dynamic states computationally infeasible. We therefore consider an MCMC algorithm in which an approximate sample from the joint full conditional distribution of the dynamic states is obtained via the ensemble Kalman smoother (EnKS) proposed by Evensen and Van Leeuwen (2000). The remaining unknown quantities can be sampled from their respective full conditional distributions and so the resulting MCMC algorithm is effectively the Gibbs ensemble Kalman smoother (GEnKS) outlined in Katzfuss et al. (2020), but applied to a model that is rather more complex than the examples considered in that work.

Various properties of the data preclude use of simpler, off-the shelf solutions to the nowcasting problem. First, since the data are collected at high temporal resolution, there are a large proportion of exact zeros. Note that zero includes both “no rain” and “an unmeasurably small amount of rain”, since we are not able to distinguish these in our observations

or our model, and the distinction is not relevant to our intended downstream application of flood forecasting. As such, there is no transformation that we can apply to the data to make a Gaussian observation model appropriate. Second, we are fusing point (rain gauge) data with areal (radar) data, and using the former to calibrate the latter, which necessitates the use of a joint model for the two data streams. In order to accommodate these atypical features, along with the relatively high spatial and temporal resolution of the data, the model we develop is necessarily complex and unavoidably high-dimensional. Correspondingly, the algorithms we develop for model-fitting are highly computationally intensive. Nevertheless, by making a number of pragmatic but principled simplifications during modeling and computational inference, we are able to describe and fit a model that generates realistic forecasts of ground-level precipitation on timescales befitting of the nowcasting problem. The model developed in this paper falls within the class of latent Gaussian models and so an alternative approach to that considered here might try to avoid Monte Carlo methods completely and directly approximate the posterior using methods based on integrated nested Laplace approximations (INLA). Ruiz-Cárdenas et al. (2012) describe a framework that allows the R-INLA software to be used to fit a variety of state space models where the states follow a Gaussian distribution. Although a number of extensions to space-time modeling have followed, the dimension of the state is typically only up to a few hundred (e.g. Laurini, 2017; Sánchez-Balseca & Pérez-Foguet, 2020). In our primary application, the dimension of the state is over a million and so fitting the model using existing INLA software on routinely available hardware is not possible.

The remainder of the paper is structured as follows. In Section 2, we present and justify our auto-regressive system model. The corresponding observation model is outlined in Section 3. In Section 4, we define the complete Bayesian model together with our prior specification, and also discuss parameter identifiability. Our Bayesian approach to inference is discussed in Section 4.2. Simulation studies illustrating the effectiveness of our posterior sampling scheme are given in Section 4 of the supplementary material and are also summarized briefly in Section 4.2 of this paper. Section 5 illustrates the use of our proposed model in a real data setting, where we consider observations obtained from the Urban Observatory project based in Newcastle upon Tyne, UK. Section 6 offers some conclusions.

2 | SYSTEM MODEL

We suppose that precipitation intensity at ground level is represented by a $(2 + 1)$ -dimensional scalar-valued latent field $\theta(\mathbf{s}, t)$ that takes values on the (unrestricted) real line. Intuitively, positive values can be thought of as a (possibly transformed) rainfall rate, with negative values indicating no rainfall; this notion is formalised via our observation model in Section 3. We suppose that the spatial domain of interest S is described by a regular lattice of dimension $N = n \times n$, that is, we consider a finite collection of equally spaced locations $\mathbf{s}_{ij} = (x_i, y_j) \in S$ for $i, j = 1, \dots, n$. The temporal domain \mathcal{T} is also discretized by considering equally spaced time points and, without loss of generality, we consider $t \in \{1, 2, \dots, T\}$. Let $\theta_{ij}^t \equiv \theta(\mathbf{s}_{ij}, t)$ be the state of the system at location \mathbf{s}_{ij} and time t .

Since rainfall intensity is related to atmospheric moisture-content, and the natural physical model of moisture-content transport and evolution is an advection-diffusion-reaction process, we can use a simple finite difference discretization of an appropriate stochastic partial differential equation (SPDE) model in order to motivate the form of the discrete time evolution equation we propose. A detailed description is provided in Sections 1 and 2 of the supplementary materials. For more background on the use of SPDEs to motivate the form of discrete system evolution equations, see the discussion in sections 6.3 and 6.4 of Cressie and Wikle (2015) and in section 5.3 and appendix D of Wikle et al. (2019).

The resulting evolution equation is

$$\begin{aligned} \theta_{ij}^{t+1} = & \alpha \left[(1 - 4\beta)\theta_{ij}^t + (\beta - v_x^t)\theta_{i+1,j}^t + (\beta + v_x^t)\theta_{i-1,j}^t \right. \\ & \left. + (\beta - v_y^t)\theta_{i,j+1}^t + (\beta + v_y^t)\theta_{i,j-1}^t \right] + S_{ij}^t + \epsilon_{ij}^t, \end{aligned} \quad (1)$$

for $t = 1, \dots, T$. The quantity $\mathbf{v}_t = (v_x^t, v_y^t)'$ represents the (potentially unknown) velocity field. Given the relatively compact spatial domain in our application, it is assumed to be spatially homogeneous. Both S_{ij}^t and ϵ_{ij}^t in (1) result from including auxiliary source-sink and error processes that are defined over the same spatial and temporal domains as $\theta(\mathbf{s}, t)$. Scalar β is a diffusion parameter controlling the degree of spatial dependence in the process, with $\beta = 0$ corresponding to no spatial dependence. Similarly, α governs the temporal dependence in the process, with $\alpha = 0$ corresponding to no temporal dependence.

Incorporating both the source-sink and error processes is a modelling choice, as they each allow for instantaneous injections/decay to and from the system. If both processes are assumed to be uncorrelated in time then the source-sink process can be absorbed into the error process ($\epsilon(\mathbf{s}, t) \rightarrow S(\mathbf{s}, t) + \epsilon(\mathbf{s}, t)$) and, if desired, spatially correlated injections/decay can be achieved by imposing a spatially dependent covariance kernel on the error process; see, for example, Sigrist et al. (2015). In our application, it is plausible to assume that injections to the system, for example the formation of a rain cloud, are likely to be both spatially and temporally correlated. Specifically, if the rainfall rate is increasing at location \mathbf{s} at time t ($S(\mathbf{s}, t) > 0$) then we think it is more likely that it is also increasing within the neighborhood of \mathbf{s} and also at time $t + dt$. This assumption can be embedded in the dynamics in one of two ways. First, the source-sink process could be temporally correlated and spatially uncorrelated, with the spatial dependence instead embedded in the error process; this approach was taken by Stroud et al. (2010) when considering the discretized advection-diffusion model. Alternatively, and the approach we consider here, the source-sink process can be both spatially and temporally correlated and coupled with a simple (spatially and temporally uncorrelated) error process. We favour this approach because it keeps the form of the error process simple, thereby avoiding direct specification of a covariance kernel, and also allows us to motivate the spatial and temporal interactions via a physical process. More specifically, motivated by similar considerations to those already discussed, the evolution of the source-sink is inspired by a finite difference solution to the (isotropic) diffusion-reaction process. Here we assume isotropy for reasons of parsimony and we omit an advection term because this process is intended to capture effects independent of wind-transport. This also aids identifiability of the model. The evolution equation corresponding to the source-sink process is then

$$S_{ij}^{t+1} = \alpha^* \left[(1 - 4\beta^*) S_{ij}^t + \beta^* (S_{i+1,j}^t + S_{i-1,j}^t + S_{i,j+1}^t + S_{i,j-1}^t) \right] + \tilde{\epsilon}_{ij}^t, \quad (2)$$

for $t = 1, \dots, T$, for appropriate scalar parameters α^* and β^* .

Together (1) and (2) specify the evolution of the system and are well defined for the majority of the points within the spatial domain. However, issues arise along the edge of the system as not all neighboring locations exist. Thus we need to enforce an appropriate boundary condition on the spatial domain. We consider periodic boundary conditions and, whilst this may provoke concerns about spurious periodicity, we note that this can be alleviated by constructing the problem such that the forecast domain of interest, say S^p , is relatively small in comparison to the model domain S and so the effect of any spurious periodicity is likely to be negligible within S^p (Cressie, 1991; Lawson et al., 1999; Liu et al., 2021).

2.1 | Spatial dependence structure

For a particular location (i, j) the relationship between the value of the processes at time $t + 1$ and that at the previous time point, t , that is, the evolution defined between the square braces ([]) of (1) and (2), can be conveniently expressed by the five-point stencils (Milne, 1953)

$$\begin{bmatrix} - & \beta - v_y^t & - \\ \beta + v_x^t & 1 - 4\beta & \beta - v_x^t \\ - & \beta + v_y^t & - \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} - & \beta^* & - \\ \beta^* & 1 - 4\beta^* & \beta^* \\ - & \beta^* & - \end{bmatrix}, \quad (3)$$

respectively. These stencils highlight the first-order spatial neighborhood structure; this evolution is akin to that defined by a (first-order) STAR(1₁) model (Cressie, 1991; Garside & Wilkinson, 2003; Pfeifer & Deutsch, 1980). Of course, higher order neighborhoods could also be considered. For example, assuming anisotropic diffusion would introduce dependence on the second-order (corner) neighbors, at the cost of additional unknown parameters. Further, an arbitrary number of spatial neighbors could be introduced by considering higher order (spatial) approximations in the finite difference solution.

A natural approach to determining the spatial order would be to pose this as a model selection problem, but this approach is not straightforward. Solving the inference problem is challenging, especially in the Bayesian paradigm, and requires a bespoke approach for each different model. Thus, rather than considering a collection of models, we instead propose to address the issue indirectly by considering the system model given by (1) and (2) and “imputing time-steps” between the observation times. Put another way, the latent processes (potentially) operate on a different temporal scale to the observed process. The basic premise is that increasing the number of imputed time-steps is akin to considering

a larger spatial neighborhood when viewed marginally from the observation times. For example, with observations at time t and $t + 1$, from (1) we have that the process at time $t + 1$ is spatially dependent on the first order neighborhood at time t (neighbors no more than one grid square from the node of interest). However, if we introduce an intermediate time $t + \frac{1}{2}$, then propagate $\theta_t \rightarrow \theta_{t+\frac{1}{2}} \rightarrow \theta_{t+1}$ via (1), keeping the same first-order neighborhood structure for each imputed time point, then information from higher order spatial neighbors is passed from time t to time $t + 1$ via the intermediate time $t + \frac{1}{2}$ (so the node of interest will now be influenced by neighbors no more than two grid squares away). As a result, increasing the number of imputed time-steps allows for the process to transition across numerous grid squares between the observation times; without imputation the first-order neighborhood structure (3) would only allow for transitions across a single grid square per time-step. The benefit of considering imputed time-steps as opposed to a collection of different models/evolution stencils is that this approach has minimal impact on our solution to the Bayesian inference problem, is parsimonious (does not introduce additional dependence parameters), and exhibits desirable behavior such as the decaying influence of a node on the node of interest with spatial distance. It is therefore straightforward to consider varying levels of imputation and use standard model selection methods to determine which model is best; this is discussed further in Section 5.

2.2 | State space representation

The evolution Equations (1) and (2) can be straightforwardly rewritten as vector autoregressive relationships. Let $\theta_t = (\theta_{11}^t, \dots, \theta_{n1}^t, \theta_{12}^t, \dots, \theta_{n2}^t, \dots, \theta_{1n}^t, \dots, \theta_{nn}^t)'$ denote the state of the process at time t , with S_t, ϵ_t and ϵ_t^* similarly defined. We can then construct *evolution* matrices $G(\mathbf{v}_t)$ and G^* , where row i contains the neighbor weights (for location i) as prescribed by the stencils (3) for θ_t (left) and S_t (right). We also allow autocorrelation in the time series for the velocity vector \mathbf{v} via a first-order (vector) autoregression. Finally, subject to certain restrictions on the parameter values (e.g., $\alpha \in (0, 1)$) that we impose via the prior distribution, the system has a *steady state* solution. It follows that as $t \rightarrow \infty$ the forecast function tends to its marginal mean, which is zero. The assumption of a zero-mean process is not physically justifiable and so we also introduce an additional (spatially and temporally homogeneous) mean parameter μ .

The dynamics of the system can then be expressed as

$$\begin{aligned}\theta_t - \mu \mathbf{1}_N &= G(\mathbf{v}_{t-1})(\theta_{t-1} - \mu \mathbf{1}_N) + S_{t-1} + \epsilon_t, \\ S_t &= G^* S_{t-1} + \epsilon_t^*, \\ \mathbf{v}_t &= \alpha_v \mathbf{v}_{t-1} + \epsilon_t^v,\end{aligned}\tag{4}$$

for $t = 1, \dots, T$, where $\mathbf{1}_N$ is an N -length column vector containing ones. The initial value/condition of the system θ_0, S_0 and \mathbf{v}_0 is assumed to be unknown and assigned a prior distribution; full details are given in Section 4.

3 | THE OBSERVATION MODEL

In general, we suppose that all quantities in (4) are unknown and conduct inference based on observed data. We suppose that both the radar and rain gauges provide noisy and potentially biased observations of the latent rain field. Further, both of these observation sources provide readings in units of mm/h and are therefore bounded below at zero. We therefore suppose that these data provide censored (at zero) observations of the latent precipitation intensity field (that exists on the real line). The relationship between the observations and the latent process is modeled via a (Type I) Tobit model (Tobin, 1958) where, for a zero-censored variable, if $Y \sim \text{Tobit}(\mu, \phi)$ then Y has a mixed discrete-continuous distribution with likelihood contribution $\pi(Y|\mu, \phi) = [\sqrt{\phi} \varphi\{\sqrt{\phi}(y - \mu)\}]^{\mathbb{I}(Y>0)} [1 - \Phi\{\sqrt{\phi}(\mu - 0)\}]^{\mathbb{I}(Y=0)}$. Here φ denotes the standard normal probability density function and Φ is the corresponding cumulative distribution function. This approach to modelling rainfall using an exceedance over a threshold of a latent Gaussian process is a well established technique in the literature (Allard & Bourotte, 2015; Allcroft & Glasbey, 2003; Mastrantonio et al., 2022). In what follows we parameterize the relationship between the observations and the latent field of interest (θ) in a general framework. Full details about the transformations and how these data are acquired for our application are given in Section 5.

3.1 | Information provided by the gauges

We suppose that we have observations from N^g rain gauges, where gauge g provides a single observation $G_{tg} \geq 0$ at each time t . Let $\mathcal{D}^g = \{G_{tg}; g = 1, \dots, N^g, t = 1, \dots, T\}$ denote the collection of observed gauge data. These observations are censored at zero and typically exhibit a long right tail; let $\tilde{Y}_{tg}^g = \mathcal{M}(G_{tg})$ denote a suitably transformed observation, where $\mathcal{M} \in \mathcal{F}$ is from the set of functions $\mathcal{F} = \{f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \text{ s.t. } f(0) = 0\}$. Yeo and Johnson (2000) propose a family of power transformations of the form

$$\mathcal{M}(x; \lambda) = \begin{cases} \log(x+1) & \lambda = 0, \\ ((x+1)^\lambda - 1)/\lambda & \text{otherwise.} \end{cases}$$

For our application we choose parameter $\lambda = 0$, and so $\mathcal{M}(x) = \log(x+1)$.

Each gauge is located within a single cell of our regular lattice grid. Let $\ell_g \in \{1, \dots, N\}$ denote the location of gauge g for $g = 1, \dots, N^g$ and note that several gauges may reside at the same location, that is we may have $\ell_g = \ell_{g'}$ for $g \neq g'$. Conditional on the quantities in (4), we assume

$$\tilde{Y}_{tg}^g | \cdot \sim \text{Tobit}(\theta_{t\ell_g}, \phi_g) \quad (5)$$

for $g = 1, \dots, N^g$ and $t = 1, \dots, T$, where $\phi_g = \sigma_g^{-2}$ is a (potentially) unknown precision parameter governing the observation error.

3.2 | Information provided by the radar

For the purpose of model specification we suppose that the radar provides observations of rainfall intensity (mm/h) at each location within our spatial domain; further details about how these data are obtained are provided in our case study in Section 5. Following Section 3.1, we let $\mathcal{D}^r = \{R_{ti}; i = 1, \dots, N, t = 1, \dots, T\}$ denote the collection of observed radar data where $R_{ti} \geq 0$ is the radar observation from location i at time t and $\tilde{Y}_{ti}^r = \mathcal{M}(R_{ti})$ denotes the corresponding transformed observation. Similarly, conditional on the quantities in (4), we assume

$$\tilde{Y}_{ti}^r | \cdot \sim \text{Tobit}(\theta_{ti} + \mu_r, \phi_r) \quad (6)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. Here an additional bias parameter $\mu_r \in \mathbb{R}$ is introduced to allow for potential miscalibration of the radar and $\phi_r = \sigma_r^{-2}$ is a (potentially) unknown precision parameter governing the observation error.

4 | THE BAYESIAN FORMULATION

We are now in a position to define our complete Bayesian model that combines the observation models (5) and (6) together with the system model (4). More specifically

$$\begin{aligned} \tilde{Y}_{tg}^g | \cdot &\sim \text{Tobit}(\theta_{t\ell_g}, \phi_g), & g = 1, \dots, N^g, \\ \tilde{Y}_{ti}^r | \cdot &\sim \text{Tobit}(\theta_{ti} + \mu_r, \phi_r), & i = 1, \dots, N, \\ \theta_t - \mu \mathbf{1}_N &= G(\mathbf{v}_{t-1})(\theta_{t-1} - \mu \mathbf{1}_N) + \mathbf{S}_{t-1} + \epsilon_t, \\ \mathbf{S}_t &= G^* \mathbf{S}_{t-1} + \epsilon_t^*, \\ \mathbf{v}_t &= \alpha_v \mathbf{v}_{t-1} + \epsilon_t^v, \end{aligned} \quad (7)$$

for $t = 1, \dots, T$. As noted in Section 2, given that both spatial and temporal dependence are allowed for in the source-sink process (S), we propose to keep the form of the errors simple. In particular we assume (i.i.d) zero-mean Gaussian innovation errors and so $\epsilon_t \stackrel{\text{indep}}{\sim} N_N(\mathbf{0}, \phi_\theta^{-1} \mathbb{I}_N)$ and $\epsilon_t^* \stackrel{\text{indep}}{\sim} N_N(\mathbf{0}, \phi_s^{-1} \mathbb{I}_N)$ for $t = 1, \dots, T$. Similarly we take the innovation error on the “velocities” to be $\epsilon_t^v \stackrel{\text{indep}}{\sim} N_2(\mathbf{0}, \phi_v^{-1} \mathbb{I}_2)$ for $t = 1, \dots, T$. These assumptions, coupled with an appropriate prior

distribution for the initial values of the dynamic states ($\theta_0, \mathbf{S}_0, \mathbf{v}_0$) and the unknown static quantities of interest, will allow an appropriate collection of latent variables to be introduced that in turn facilitate the full conditional distributions for all unknown quantities to be written down in closed form.

4.1 | Prior specification

The problem of specifying a suitable prior distribution has been given considerable attention in the Bayesian literature (Bernardo & Smith, 1994). Given the Gaussian form of our observation model and, with tractability in mind, we choose to place Gaussian distributions over the initial values of the system. More specifically we take $\theta_0 | \mu \sim N_N(\mu \mathbf{1}_N, C_\theta)$, $\mathbf{S}_0 \sim N_N(\mathbf{m}_s, C_s)$ and $\mathbf{v}_0 \sim N_2(\mathbf{m}_v, C_v)$, where the mean vectors (\mathbf{m}_\cdot) and covariance matrices (C_\cdot) are specified *a priori*. The static mean and bias parameters are also assumed to follow Gaussian distributions $\mu \sim N(m_\mu, v_\mu)$ and $\mu_r \sim N(m_{\mu_r}, v_{\mu_r})$. Our prior specification is completed by choosing appropriate distributions for the static parameters in the evolution matrix G_t ; we take $\beta \sim N(m_\beta, v_\beta)$ where, to encourage positive evolution coefficients, m_β and v_β are to be chosen so that $\Pr(\beta \in (0, 0.25)) \simeq 1$ *a priori*. To ensure a stationary solution we take $\alpha \sim \text{TN}(m_\alpha, v_\alpha, 0, 1)$, that is, a truncated Gaussian distribution ensuring $\alpha \in (0, 1)$.

The remaining static quantities in (7) are considered to be fixed (known) constants, although we note that this is not a limitation of our approach. For example, appropriate (Gamma) prior distributions could also be placed on the innovation precision parameters ($\phi_\theta, \phi_s, \phi_v$), though we found that these parameters are only weakly identified in practice. Similar issues arise for the observation precision parameters ϕ_r and ϕ_g . This is further complicated given that, in our application, the radar provides a much larger number of observations (and therefore, information) in comparison to the rain gauges ($N \gg N^g$). It follows that, with ϕ_g unknown, there is a potential risk of the gauge observations being explained by large observation error ($\phi_g \simeq 0$), and this is not desirable given that the gauges provide the most accurate information about the precipitation intensity field at ground level; this issue can be alleviated by an appropriate choice of $\phi_g \gg \phi_r$. Of course, choosing suitable values for these static quantities poses its own challenge, although such choices can be guided by inspection of the prior predictive distribution. The choice of state innovation error ϕ_θ from a finite set of possible values is posed as a model selection problem.

4.2 | Posterior computation

The Tobit likelihood(s) together with the prior distribution given in Section 4.1 leads to a posterior distribution that is not available in closed form. We therefore propose a simulation based inference approach, specifically MCMC, to obtain draws from the posterior distribution of interest. The design of the sampling algorithm can be simplified by the introduction of appropriate latent variables that give rise to tractable full conditional distributions (FCDs) for all unknown quantities of interest. However, the dimension of our state-space renders (exact) simulation from the joint full conditional distribution of the dynamic states computationally infeasible. A potential, relatively straightforward, solution to this issue would be to sample each of the latent states one-at-a-time from its full conditional distribution. Unfortunately this approach is likely to lead to poor mixing in the MCMC scheme and therefore require an infeasible number of iterations to obtain a reasonable number of near uncorrelated posterior draws. Of course, samples from the posterior distribution could be obtained via Metropolis-Hastings (MH) steps within an MCMC scheme. However, whilst it is fairly straightforward to conceive of a sensible proposal distribution for the static parameters, it is less clear how to construct a suitable joint proposal for the dynamic states. In particular, the high-dimension of the state-space is likely to lead to low acceptance rates, which is not appealing. We therefore propose to use the ensemble Kalman smoother (EnKS) to obtain an approximate sample from the joint full conditional distribution of the dynamic states. The EnKS sampling step is embedded within a Gibbs sampling algorithm (in place of its exact counterpart) to facilitate approximate posterior sampling in reasonable time. Section 4.3 introduces an appropriate collection of latent variables and the resulting complete data model is given in Section 4.4. The EnKS algorithm is outlined in Section 4.5 and a discussion of our sampling algorithm is given in Section 4.6.

4.3 | Latent variables

The inference problem can be made more tractable by appealing to data augmentation (Tanner & Wong, 1987). The basic idea is to introduce an appropriate collection of *latent observations* such that the complete data likelihood, that is the joint

distribution of the latent observations and the observed data, is of a convenient form. An equivalent way to arrive at the observation model (7) is to directly define (partially) latent vectors that denote the complete data, \mathbf{Y}_t^g and \mathbf{Y}_t^r , and assume $Y_{ig}^g | \cdot \sim N(\theta_{t\ell_g}, \phi_g)$ and $Y_{it}^r | \cdot \sim N(\theta_{ti} + \mu_r, \phi_r)$ for $g = 1, \dots, N^g$, $i = 1, \dots, N$ and $t = 1, \dots, T$. Then, conditional on the complete data (and all other unknown quantities), we can let the observed data $\tilde{Y}_{ig}^g = Y_{ig}^g \mathbb{I}(Y_{ig}^g > 0)$ and $\tilde{Y}_{it}^r = Y_{it}^r \mathbb{I}(Y_{it}^r > 0)$, from which it follows that the observed data likelihood contributions, for example $\pi(\tilde{Y}_{ig}^g | \cdot) = \int \pi(\tilde{Y}_{ig}^g | Y_{ig}^g) \pi(Y_{ig}^g | \cdot) dY_{ig}^g$, are as outlined in Section 3. Note that, given a collection of observed radar and gauge observations $\mathcal{D} = \{\mathcal{D}^g, \mathcal{D}^r\}$, the full conditional distributions of the quantities in the (partially) latent vectors are straightforward to obtain and are $Y_{ig}^g | \mathcal{D}^g, \cdot \sim \text{TN}(\theta_{t\ell_g}, \phi_g, -\infty, 0)$ if $\tilde{y}_{ig}^g = 0$ and $Y_{ig}^g = \tilde{y}_{ig}^g$ otherwise; similarly $Y_{it}^r | \mathcal{D}^r, \cdot \sim \text{TN}(\theta_{ti} + \mu_r, \phi_r, -\infty, 0)$ if $\tilde{y}_{it}^r = 0$ and $Y_{it}^r = \tilde{y}_{it}^r$ otherwise.

Let $\mathbf{Y}_t = (\mathbf{Y}_t^r, \mathbf{Y}_t^g)$ be the $(N + N^g)$ -length complete data vector that contains the positive (radar and gauge) observations when available, and the latent (truncated Gaussian) variables otherwise. Further, for notational simplicity, let $\rho = (\mu, \mu_r, \alpha, \beta)$ denote the unknown static quantities and $\mathbf{q}_{1:T} \equiv \{\mathbf{q}_1, \dots, \mathbf{q}_T\}$ for any vector valued quantity \mathbf{q} . From above it follows that the complete data likelihood $\pi(\mathbf{Y}_{1:T} | \theta_{0:T}, \mathbf{S}_{0:T}, \mathbf{v}_{0:T}, \rho)$ is a $T \times (N + N^g)$ -dimensional multivariate Gaussian density. Crucially this allows us to write down a dynamic linear model for the complete data from which the full conditional distributions of all remaining unknown quantities can be straightforwardly derived; this is the topic of the next section.

4.4 | Complete data model

The *complete data model* is given by our original state model coupled with that for the latent observations $\mathbf{Y}_{1:T}$. With notational simplicity in mind, let $\mathbf{x}_t = (\theta_t', \mathbf{S}_t')'$ denote the joint vector containing both of the dynamic processes at time t . It follows that the unknown quantities of interest are $\mathbf{Y}_{1:T}$, $\mathbf{x}_{0:T}$, $\mathbf{v}_{0:T}$ and ρ and the density of all stochastic quantities is $\pi(\mathcal{D}, \mathbf{Y}_{1:T}, \mathbf{x}_{0:T}, \mathbf{v}_{0:T}, \rho) = \pi(\mathcal{D} | \mathbf{Y}_{1:T}) \pi(\mathbf{Y}_{1:T} | \mathbf{x}_{0:T}, \mathbf{v}_{0:T}, \rho) \pi(\mathbf{x}_{1:T} | \mathbf{x}_0, \mathbf{v}_{0:T}, \rho) \pi(\mathbf{x}_0 | \rho) \pi(\mathbf{v}_{0:T} | \rho) \pi(\rho)$, from which the full conditional distributions can be obtained in closed form. Note that the full conditional distributions for the latent variables $\mathbf{Y}_{1:T}$ are those outlined in Section 4.3. Conditional on the latent observations (and the observed data), we can then construct the conditional distribution $\pi(\mathbf{x}_{0:T}, \mathbf{v}_{0:T}, \rho | \mathbf{Y}_{1:T}, \mathcal{D}) \propto \pi(\mathbf{Y}_{1:T} | \mathbf{x}_{0:T}, \mathbf{v}_{0:T}, \rho) \pi(\mathbf{x}_{1:T} | \mathbf{v}_{0:T}, \rho) \pi(\mathbf{x}_0 | \rho) \pi(\mathbf{v}_{0:T} | \rho) \pi(\rho)$. Crucially, given the prior specification outlined in Section 4.1 and the velocities, $\mathbf{v}_{0:T}$, this is simply the density corresponding to the dynamic linear model (DLM),

$$\begin{aligned} \mathbf{Y}_t &= F\mathbf{x}_t + \mu_r \mathbf{L}^r + \mathbf{v}_t, & \mathbf{v}_t &\sim N_{N+N^g}(\mathbf{0}, V), \\ (\mathbf{x}_t - \mu \mathbf{L}) &= \tilde{G}_t(\mathbf{x}_{t-1} - \mu \mathbf{L}) + \mathbf{w}_t, & \mathbf{w}_t &\sim N_{2N}(\mathbf{0}, W), \\ \mathbf{v}_t &= \alpha_v \mathbf{v}_{t-1} + \epsilon_t^v, \end{aligned} \quad (8)$$

for $t = 1, \dots, T$. Here $\mathbf{L} = (\mathbf{1}_N', \mathbf{0}_N')'$ and $\mathbf{L}^r = (\mathbf{1}_N', \mathbf{0}_{N^g}')'$ are indicator vectors that map the mean and bias parameters to the respective process and observation layer. The matrices are of the form

$$F = \begin{bmatrix} \mathbb{I}_N & 0_{N \times N} \\ \mathbf{I}^g & 0_{N^g \times N} \end{bmatrix}, \tilde{G}_t = \begin{bmatrix} G(\mathbf{v}_{t-1}) & \mathbb{I}_N \\ 0_{N \times N} & G^* \end{bmatrix}, V = \begin{bmatrix} \phi_r^{-1} \mathbb{I}_N & 0_{N \times N^g} \\ 0_{N^g \times N} & \phi_g^{-1} \mathbb{I}_{N^g} \end{bmatrix} \text{ and } W = \begin{bmatrix} \phi_\theta^{-1} \mathbb{I}_N & 0_{N \times N} \\ 0_{N \times N} & \phi_s^{-1} \mathbb{I}_N \end{bmatrix},$$

where \mathbf{I}^g maps the latent rain field (θ) at location ℓ_g to the observation from gauge g ; specifically an $(N^g \times N)$ dimensional matrix where row g contains a 1 in column ℓ_g and zeros otherwise.

The full conditional distributions for $\mathbf{x}_{0:T}$, $\mathbf{v}_{0:T}$ and ρ can now straightforwardly be derived from $\pi(\mathbf{x}_{0:T}, \mathbf{v}_{0:T}, \rho | \mathbf{Y}_{1:T}, \mathcal{D})$. In particular, since the joint distribution of $\mathbf{x}_{0:T}$ and $\mathbf{Y}_{0:T}$ can be expressed through a dynamic linear model, it follows that a joint sample from $\pi(\mathbf{x}_{0:T} | \cdot)$ can be obtained via the FFBS algorithm (Frühwirth-Schnatter, 1994). Thus, in principle, it is straightforward to construct a Gibbs sampler in which we draw a joint sample of the dynamic states (via the FFBS algorithm) and then proceed to draw from the full conditional distributions of the remaining parameters (including the latent observations) in turn. Whilst this strategy is naturally appealing, unfortunately the FFBS algorithm becomes computationally infeasible in large state dimension. In our application \mathbf{x}_t has length $2N = 2 \times 72^2 = 10368$ and so analytically computing the moments of the required (Gaussian) distributions is challenging; in particular the required covariance matrices are of dimension 10368×10368 . Our solution to this problem is to instead use the EnKS to obtain an approximate draw from $\pi(\mathbf{x}_{0:T} | \cdot)$; this is the topic of the next section.

4.5 | State sampling via the ensemble Kalman smoother

The EnKS is a method for obtaining approximate draws from $\pi(\mathbf{x}_{0:T}|\cdot)$. In essence, rather than analytically computing the moments of the required (Gaussian) distributions (as is done in the FFBS algorithm), we instead consider an ensemble of N^e (equally weighted) particles that are draws from the distributions of interest. These particles are propagated and sequentially updated as observations “arrive”, where the (cross-)covariance matrices required to compute the update are approximated from the ensemble. For dynamic linear models this method provides an exact sample from $\pi(\mathbf{x}_{0:T}|\cdot)$ in the limit $N^e \rightarrow \infty$ (Katzfuss et al., 2020). The EnKS is a generalisation of the more well known ensemble Kalman filter (EnKF) and numerous variants of the smoother can be found in the literature (Bocquet & Sakov, 2014; Evensen & Van Leeuwen, 2000; Khare et al., 2008). Here we consider the EnKS of Evensen and Van Leeuwen (2000); the basic idea is that the smoothed states can be obtained by applying the standard EnKF update to the augmented state that includes all the previous history. It follows that, at each time t , the entire state history $(\mathbf{x}_0, \dots, \mathbf{x}_t)$ is updated to incorporate the information in the observations \mathbf{y}_t ; see Katzfuss et al. (2020) for further details. In what follows, the notation $\mathbf{x}_{n|m}$ represents a draw from the distribution $\pi(\mathbf{x}_n|\mathbf{Y}_{1:m}, \cdot)$, that is, a sample of the state vector at time n given observations up to and including at time m . For the model given by (8), an approximate sample $\mathbf{x}_{0:T|T}$ from $\pi(\mathbf{x}_{0:T}|\cdot)$ is obtained as follows.

Initialize: draw $\mathbf{x}_{0|0}^j \stackrel{\text{indep}}{\sim} N(\mathbf{m}_0, C_0)$ for $j = 1, \dots, N^e$. Then, for $t = 1, \dots, T$,

1. Forecast step: for $j = 1, \dots, N^e$,
2. Let $\mathbf{x}_{t|t-1}^j = \tilde{G}_t(\mathbf{x}_{t-1|t-1}^j - \mu\mathbf{L}) + \mu\mathbf{L} + \mathbf{w}_t^j$ where $\mathbf{w}_t^j \stackrel{\text{indep}}{\sim} N(\mathbf{0}, W)$.
 - (a) Smoothing step: for $j = 1, \dots, N^e$,
 - (b) Generate pseudo-observations $\hat{\mathbf{y}}_{t|t-1}^j = F\mathbf{x}_{t|t-1}^j + \mu_r\mathbf{L}^r + \mathbf{v}_t^j$ where $\mathbf{v}_t^j \stackrel{\text{indep}}{\sim} N(\mathbf{0}, V)$.
 - (c) For $\ell = 0, \dots, t$, compute (the smoothed state) $\mathbf{x}_{\ell|t}^j = \mathbf{x}_{\ell|t-1}^j + \hat{K}_{\ell t}(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}^j)$.

A sample of the state vector is obtained by drawing $j \in \{1, \dots, N^e\}$ uniformly at random and letting $\mathbf{x}_{0:T|T} = \mathbf{x}_{0:T|T}^j$. Here $\hat{K}_{\ell t} = \Sigma_{\ell t|t-1}^{\text{xy}} \left(\Sigma_{tt|t-1}^{\text{yy}} \right)^{-1}$ is an approximation of the *optimal* Kalman gain computed from the ensemble. More specifically $\Sigma_{\ell t|t-1}^{\text{xy}} = \text{Cov}(\mathcal{X}_{\ell|t-1}, \mathcal{Y}_{t|t-1})$ is the sample cross-covariance of the ensemble of states $\mathcal{X}_{\ell|t-1} = (\mathbf{x}_{\ell|t-1}^1, \dots, \mathbf{x}_{\ell|t-1}^{N^e})$ and pseudo observations $\mathcal{Y}_{t|t-1} = (\hat{\mathbf{y}}_{t|t-1}^1, \dots, \hat{\mathbf{y}}_{t|t-1}^{N^e})$; similarly $\Sigma_{tt|t-1}^{\text{yy}}$ is the sample covariance of the ensemble of pseudo observations $\mathcal{Y}_{t|t-1}$.

From a practical standpoint, updating (smoothing) the entire state history in Step 2 is computationally burdensome, even for modest length time-series. Thus in practice it is typically only feasible to consider a “fixed-lag smoother” in which only the states at the previous τ time points are updated given the observations at each time t . More formally, to only consider $\ell = t^*, \dots, t$, where $t^* = \max(0, t - \tau)$ in Step 2 of the algorithm above; the EnKF is a special case and is recovered by taking $\tau = 0$. Finally, we note that if there are no observations available at time t , for example if \mathbf{x}_t is the state vector at an imputed time-step, then the smoothing step (Step 2) becomes trivial and we simply let $\mathbf{x}_{\ell|t}^j = \mathbf{x}_{\ell|t-1}^j$ for $\ell = 0, \dots, t$ and $j = 1, \dots, N^e$.

Key to the successful implementation of the EnKS algorithm is the effective estimation of the (cross-)covariance matrices required to compute $\hat{K}_{\ell t}$, especially when the ensemble size (N^e) is small relative to the dimension of the states and/or observations. Several methods have been proposed for improving the accuracy of the covariance approximations in this setting, most commonly *covariance tapering* (Houtekamer & Mitchell, 2001) and *covariance inflation* (Anderson & Anderson, 1999). However there is little guidance on how to choose an appropriate tapering function and/or inflation factor. The approach we favor, and also that which we find works well in practice, relies on approximating the matrices from the so-called *deterministic ensemble* and then using analytic results to obtain the required matrix. More specifically we introduce $\tilde{\mathbf{x}}_{t|t-1}^j = \tilde{G}_t\mathbf{x}_{t-1|t-1}^j$ in Step 1 of the EnKS algorithm and let $\mathbf{x}_{t|t-1}^j = \tilde{\mathbf{x}}_{t|t-1}^j + \mathbf{w}_t^j$ for $j = 1, \dots, N^e$. The Kalman gain matrix can then be written as

$$\hat{K}_{\ell t} = \begin{cases} \tilde{\Sigma}_{\ell t|t-1} F' (F \tilde{\Sigma}_{tt|t-1} F' + F W F' + V)^{-1}, & \text{for } \ell < t, \\ (\tilde{\Sigma}_{tt|t-1} + W) F_t' (F \tilde{\Sigma}_{tt|t-1} F' + F W F' + V)^{-1}, & \text{for } \ell = t. \end{cases}$$

Here $\tilde{\Sigma}_{t|t-1}$ is the sample covariance of the deterministic ensemble $\tilde{\mathcal{X}}_{t|t-1} = (\tilde{\mathbf{x}}_{t|t-1}^1, \dots, \tilde{\mathbf{x}}_{t|t-1}^{N^e})$ and $\tilde{\Sigma}_{\ell|t-1}$ is the cross-covariance matrix computed from the ensembles $\mathcal{X}_{\ell|t-1}$ and $\tilde{\mathcal{X}}_{t|t-1}$. Additional discussion is provided in Section 3 of the supplementary material and further guidance on the practical implementation of these algorithms is given by Evensen (2003).

4.6 | MCMC sampling algorithm

With all full conditional distributions available in closed form we can straightforwardly implement a Gibbs sampling scheme to obtain draws from the desired posterior distribution (Geman & Geman, 1984; Smith & Roberts, 1993; Tierney, 1994). Our MCMC algorithm is effectively the Gibbs ensemble Kalman smoother (GEnKS) algorithm outlined in Katzfuss et al. (2020) applied to our data fusion model. In particular we draw a joint sample of the dynamic states $\mathbf{x}_{0:T}$ via the EnKS algorithm outlined in Section 4.5, then proceed to draw a sample of each quantity in ρ is from its corresponding (univariate) FCD. Samples of the latent observations ($\mathbf{Y}_{1:T}$) are drawn from their FCDs given in Section 4.3 and finally the “velocity” parameters are sampled from $\pi(\mathbf{v}_t|\cdot)$ for $t = 0, \dots, T$. The latent variable method of Damien and Walker (2001) is used to facilitate numerically stable sampling from truncated (Gaussian) distributions. A joint sample of $\mathbf{v}_{0:T}$ could be obtained via the FFBS algorithm which may promote better mixing of the MCMC chain. However, in general we find that our chains mix well, and so do not consider this here.

Although Gibbs samplers are inherently serial algorithms, the total execution time can often be reduced by appealing to parallel computing in each of the sampling steps. In particular, many of the computations required within the EnKS algorithm can be effectively performed in parallel; see, for example, Houtekamer et al. (2014). Further, the large matrix operations, including those required to compute the moments of the full conditional distributions, can be straightforwardly computed in parallel, for example via routines from the Eigen library (Guennebaud & Jacob, 2010). All computations described here can be carried out on routinely available hardware (standard Linux servers with at least 16GB RAM). The full MCMC algorithms run in a few hours on a reasonable processor (e.g., an Intel i7), and nowcasting based on inferred parameters can be carried out in a few seconds, allowing practical application to the intended downstream flood forecasting application. C++ code to run our algorithm can be found at the GitHub repository <https://github.com/srjresearch/BayesianRainfallModelling>.

4.7 | Simulation study

To investigate the effectiveness of the posterior sampling scheme outlined in Section 4.6, we analyze several synthetic datasets for which the values of the underlying parameters are known. Details of the data generating mechanism and additional discussion is given in Section 4 of the supplementary material. The synthetic datasets considered closely resemble that used in our real data application, which follows in Section 5, and consider an $N = 72 \times 72$ spatial grid with $N^g = 15$ rain gauges over $T = 72$ observation times. The choice of ensemble size and the length of the smoothing window to consider are primarily limited by computational resource. From a practical standpoint we would like to be able to make inferences in a reasonable amount of time (hours as opposed to days) and to this end we let $N^e = 100$ and consider two different choices of smoothing window; in Analysis 1 we back smooth the latent states up to and including the previous $\tau = 3$ observation times. Analysis 2 considers a larger smoothing window with $\tau = 6$. In both analyses the number of imputed time-steps is fixed at $\tilde{\tau} = 4$.

Figure 1 shows boxplots of the marginal posterior distributions of the unknown static parameters α, β, μ and μ_r for both analyses 1 and 2; the crosses (\times) highlight the value from which these data were generated and the triangles (Δ) denote the (marginal) posterior mean. From Figure 1 (right) we see that both μ and μ_r are well recovered in each case, that is there is reasonable posterior support for the synthetic values used to generate these data. Interestingly both analyses struggle to recover the synthetic value of $\beta = 0.18$ with $E(\beta|D) = 0.213, 0.212$ for Analysis 1 and 2, respectively. That said, it is pleasing to see that the velocity components \mathbf{v}_t are generally well recovered, with the majority of the synthetic values looking plausible under their marginal posterior distributions for both analyses. Figure 2 shows the marginal posterior means, the 95% credible regions, and also the values of the velocity components v_x (top) and v_y (bottom) from which these data were simulated for both Analysis 1 and 2, left and right, respectively.

The result of reducing the smoothing window in the EnKS step is perhaps most likely to manifest in the inferences for the latent process θ_t . However, the results from Analysis 1 and Analysis 2 suggest that reducing the smoothing window

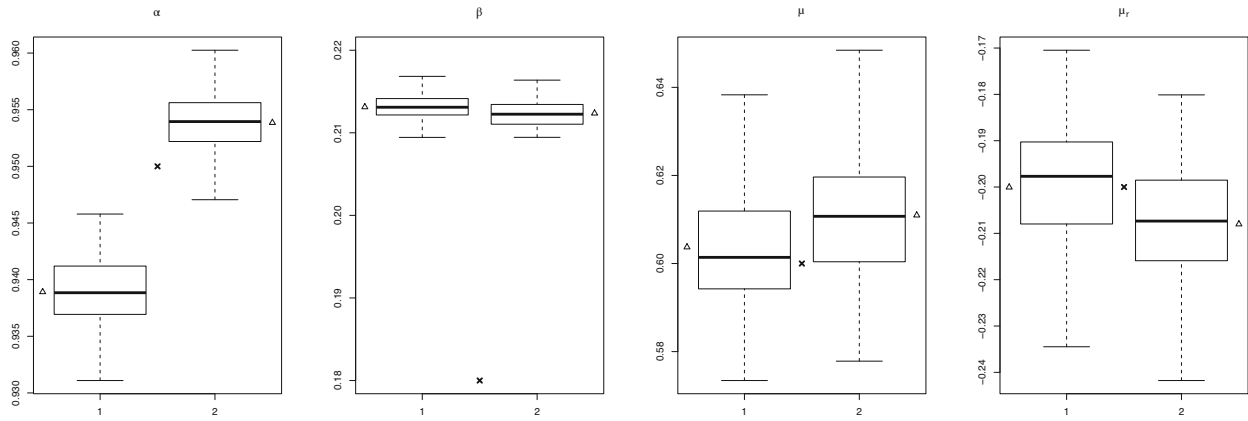


FIGURE 1 Boxplots of the marginal posterior distributions of the unknown static parameters for analyses 1 and 2; crosses (x) highlight the value from which these data were simulated, triangles (Δ) denote the posterior means.

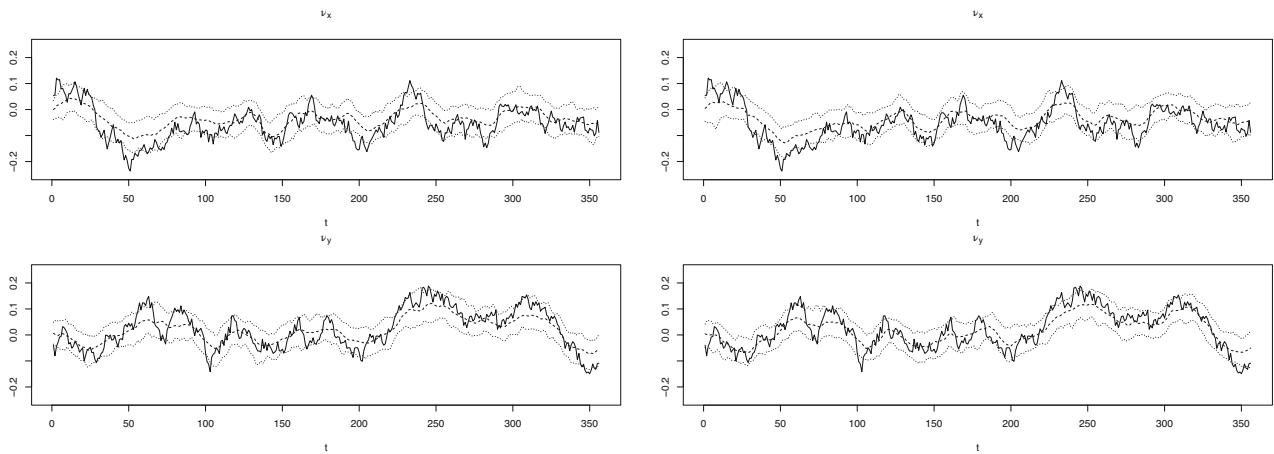


FIGURE 2 Marginal posterior means (---), 95% credible regions (···), and the synthetic values (—) for the velocity components v_x (top) and v_y (bottom) under Analysis 1 (left) and Analysis 2 (right) for $t = 1, \dots, \bar{T}$.

from 6 to 3 has little impact on the state inference. In particular, similar spatial clusters of positive (and negative) surface water rates are recovered under both analyses; corresponding figures can be found in Section 4 of the supplementary materials.

Our model (7) is governed by numerous parameters in addition to multiple layers of latent variables and this may give rise to parameter identifiability issues. However, as discussed above, the simulation studies suggest that the underlying parameters are identifiable and generally well recovered, despite the bias introduced by the approximation in the EnKS step. It is perhaps unsurprising that inferences for the static parameters α and β that govern the system evolution are those most affected by EnKS approximation; this may also be an artefact of information loss as these data are only partially observed. Crucially, our simulation studies suggest that we are able to make plausible inferences about the latent process of interest θ_t , with spatial clusters of positive (and negative) precipitation intensities being well recovered. Finally we note that the observed data look plausible under their posterior predictive distribution and so, based on these simulation studies, we suggest that we are able to make reasonable inferences given the choice of $N^e = 100$ and $\tau = 3$.

5 | REAL DATA APPLICATION

The Urban Observatory (James et al., 2014) collects data from a large number of sensors distributed around the North East of England. We consider observations obtained from a WR-10X radar situated on Claremont Tower, Newcastle upon

Tyne, UK (Lat 54.9804, Long -1.6146) in addition to a collection of tipping bucket rain gauges distributed around the city centre. We consider a 12 h period from the August 12, 2018, when moderately large rainfall rates were observed, and for which a reasonable amount of rain gauge data are available. Our modelling domain S is taken to be an 18 km^2 area centered at the location of the radar (Claremont Tower). This domain is discretized as outlined in Section 2 by imposing a regular lattice grid of dimension $N = 72 \times 72$, and so each grid cell is of dimension 500 m^2 .

The radar provides, at 10 min time intervals, 240×360 raw observations z that are given in units of dBZ (decibel relative to Z). The corresponding rainfall rates R (mm/h) are obtained via the Marshall and Palmer (1948) relationship. Additionally we suppose that values of $z < 0$ should correspond to zero rainfall and so we let the z - R relationship be $R(z) = (10^{z/10}/200)^{0.625} \mathbb{I}(z > 0)$. Note that $(1/200)^{0.625} = 0.04$ and so only rainfall rates less than 0.04 mm/h are truncated to zero. The radar records observations at 150 m intervals along “beams” that are 36 km in length; 360 beams cover azimuth angles of 1 – 360° (at an angular resolution of 1°). It follows that these observations are irregularly spaced over the modelling domain S and so we take R_{it} to be the average rainfall rate when numerous observations fall within the same grid square; observations that fall outside of our modelling domain are discarded. Additionally we obtain observations from $N^g = 15$ tipping bucket rain gauges that reside within S . Each gauge is located within a single grid square and provides rainfall accumulations over a fixed time domain and, where appropriate, we sum over the appropriate time domains to obtain the 10 min accumulation before converting into units of mm/h (multiplying by a factor of six). It follows that the length of the time series is $T = 72$ and the number of the observations (at each time-step) is $N + N^g = 5199$. These data (in units of mm/h) are reproduced in the associated GitHub repository.

Our ultimate interest lies in inferring the (latent) precipitation intensity field at ground level and we believe that the rain gauges provide reasonably accurate information about this process (in comparison to the radar observations). Based on our understanding of the devices, we make the specifications $\phi_g = 100$ and $\phi_r = 2$, which allows the gauge observations to strongly influence the state inference; these values were also used in the simulation studies discussed in Section 4.7. Although in principle ϕ_g and ϕ_r could be inferred from the data, in practice, due to the huge imbalance between the number of radar and gauge observations, there is a tendency for the model to want to discount the gauge observations as “noise”. We consider the prior distribution outlined in Section 4.1 with $\theta_0 | \mu \sim N_N(\mu \mathbf{1}_N, 2^2 \mathbb{I}_N)$, $S_0 \sim N_N(\mathbf{0}, 0.5^2 \mathbb{I}_N)$, $\nu_0 \sim N_2(\mathbf{0}, 0.1^2 \mathbb{I}_2)$, $\mu, \mu_r \sim N(0, 1)$, $\alpha \sim \text{TN}(0.8, 250^{-1}, 0, 1)$ and $\beta \sim N(0.1, 500^{-1})$. The parameters governing the evolution of the source-sink process, $\alpha^* = 0.85$, $\beta^* = 0.15$ and $\phi_s = 20$, are considered fixed and were chosen via visual inspection of the prior predictive distribution. We suppose that the components of the velocity vector \mathbf{v} should exhibit large temporal dependence and, to this end, we let $\alpha_v = 0.95$ and $\phi_v = 2000$, so that each component is modelled via a stationary AR(1) process with stationary standard deviation of 0.07 . Finally we pose the choice of the innovation precision parameter ϕ_θ as a model selection problem. More specifically, we consider multiple values of ϕ_θ and select the value which gives rise to the lowest deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002); the effective number of parameters is approximated by (two times) the sample variance of the log observed data likelihood (Gelman et al., 2014). Other model selection methods could also be considered, however DIC is particularly appealing as it is computationally cheap to evaluate and does not require saving the MCMC draws of all unknown quantities, which is prohibitive in large scale applications.

Attempts to directly fit the model (7) to these data (with various values of ϕ_θ) gave rise to poor model fit and predictive performance (results not reported here). Inspection of the radar observations reveals that rain clouds move fairly quickly across the spatial domain, that is they transition across numerous grid squares between a single time-step. Thus, following the discussion in Section 2.1, we insert \tilde{t} time-steps between the observation times to allow our model to capture this

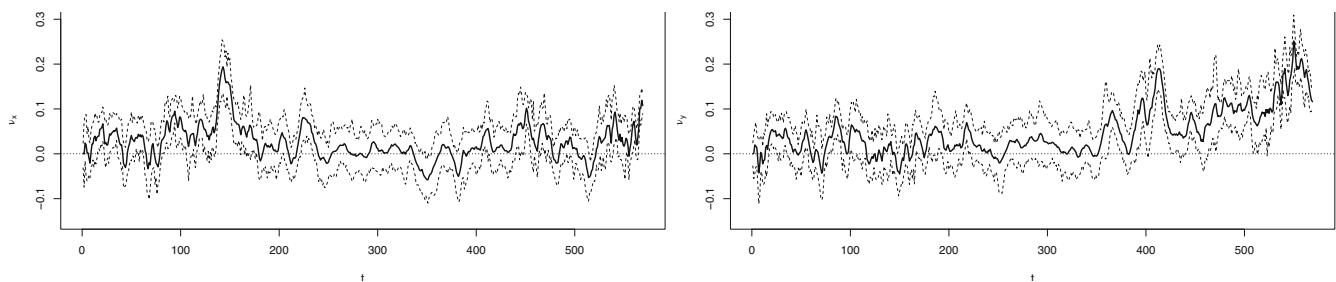


FIGURE 3 Marginal posterior means (—) and the 95% credible regions (---) for the velocity components v_x (left) and v_y (right) for $t = 1, \dots, \tilde{T}$. Horizontal line (· · ·) drawn at $v_x, v_y = 0$.

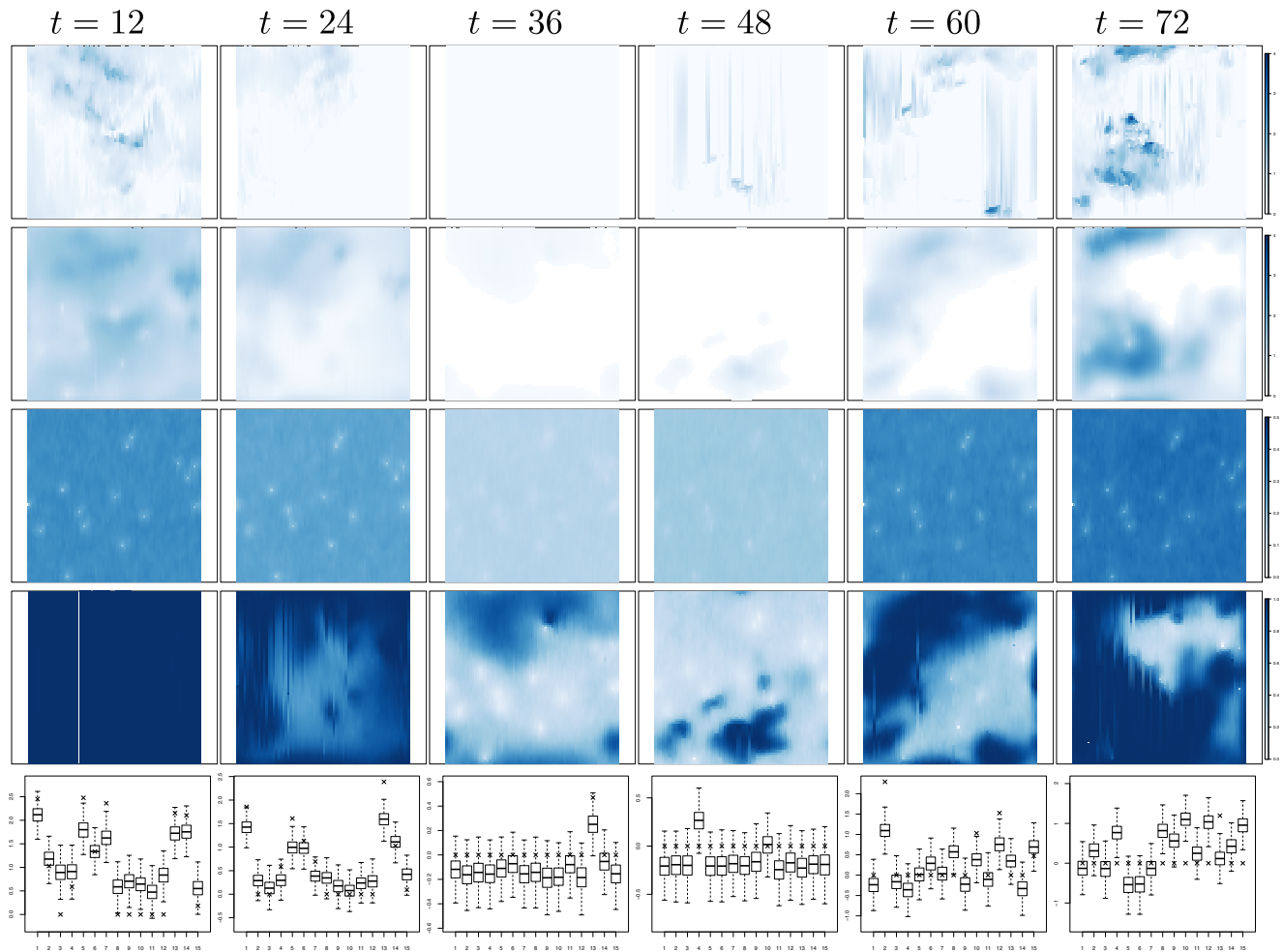


FIGURE 4 Image plots of the (transformed) radar observations \mathbf{y}_t^r (top row), marginal posterior means $E(\theta_{ii}|D)$ and standard deviations $SD(\theta_{ii}|D)$ of the precipitation field (second and third row, respectively). Darker shades of blue indicated higher values. Posterior probabilities $\Pr(\theta_{ii} > 0|D)$ (fourth row) and boxplots of the marginal posterior distributions of the precipitation field at the rain gauge locations; crosses show the (transformed) gauge observations \mathbf{y}_t^g (fifth row). Note that all real data images correspond to a 36×36 km² latitude-longitude-aligned square region (North at top) centred on 54.98°N , 1.61°W . See Figure 5 of the supplementary materials for a map of the data locations.

behavior. This helps because the first order model implicitly assumes that velocity is small relative to the grid size and time step. Introducing latent observations shortens the time step, making this assumption more plausible. Note that the dynamic processes of interest are $\mathbf{x}_{0:\tilde{T}}$ and $\mathbf{v}_{0:\tilde{T}}$ where the length of the augmented time series is $\tilde{T} = \tilde{t}(T - 1) + T$. Further, to only introduce similar levels of stochastic noise in to the latent process between the observation times, we scale the innovation (precision) matrix and let $W^{-1} \rightarrow W^{-1} \times (\tilde{t} + 1)$. We consider $\phi_\theta \in \{20, 40, 60\}$ and $\tilde{t} \in \{0, 1, \dots, 10\}$. Posterior draws are obtained via the MCMC algorithm outlined in Section 4.6, where we perform 2000 iterations with the first 1000 discarded as burn-in. Convergence is assessed by monitoring the trace of the (log) complete data likelihood in addition to the static parameters and randomly selected dynamic states.

These analyses reveal that, amongst the values of ϕ_θ and \tilde{t} considered, the model which was optimal according to the DIC set $\phi_\theta = 40$ and used $\tilde{t} = 7$ imputed time-steps. Inspection of the marginal posterior distributions of the static parameters reveals that there is high correlation between the latent states at successive time points with $E(\alpha|D) = 0.99$ ($SD(\alpha|D) = 3.87 \times 10^{-4}$); this is perhaps not surprising given the modestly large number of imputed time-steps. Interestingly, we find that $E(\beta|D) = 0.24$ ($SD(\beta|D) = 6.50 \times 10^{-4}$). When coupled with the velocities v_x, v_y , this suggests that the (first-order) neighbors have a large influence with relatively little weight $(1 - 4\beta)$ assigned to the same location at the previous time-step. The analysis also reveals that the radar typically provides lower rainfall rates than those observed on the ground with $E(\mu_r|D) = -1.09$ ($SD(\mu_r|D) = 0.03$); if the radar was calibrated with the gauge observations we would expect to see $\mu_r \approx 0$.

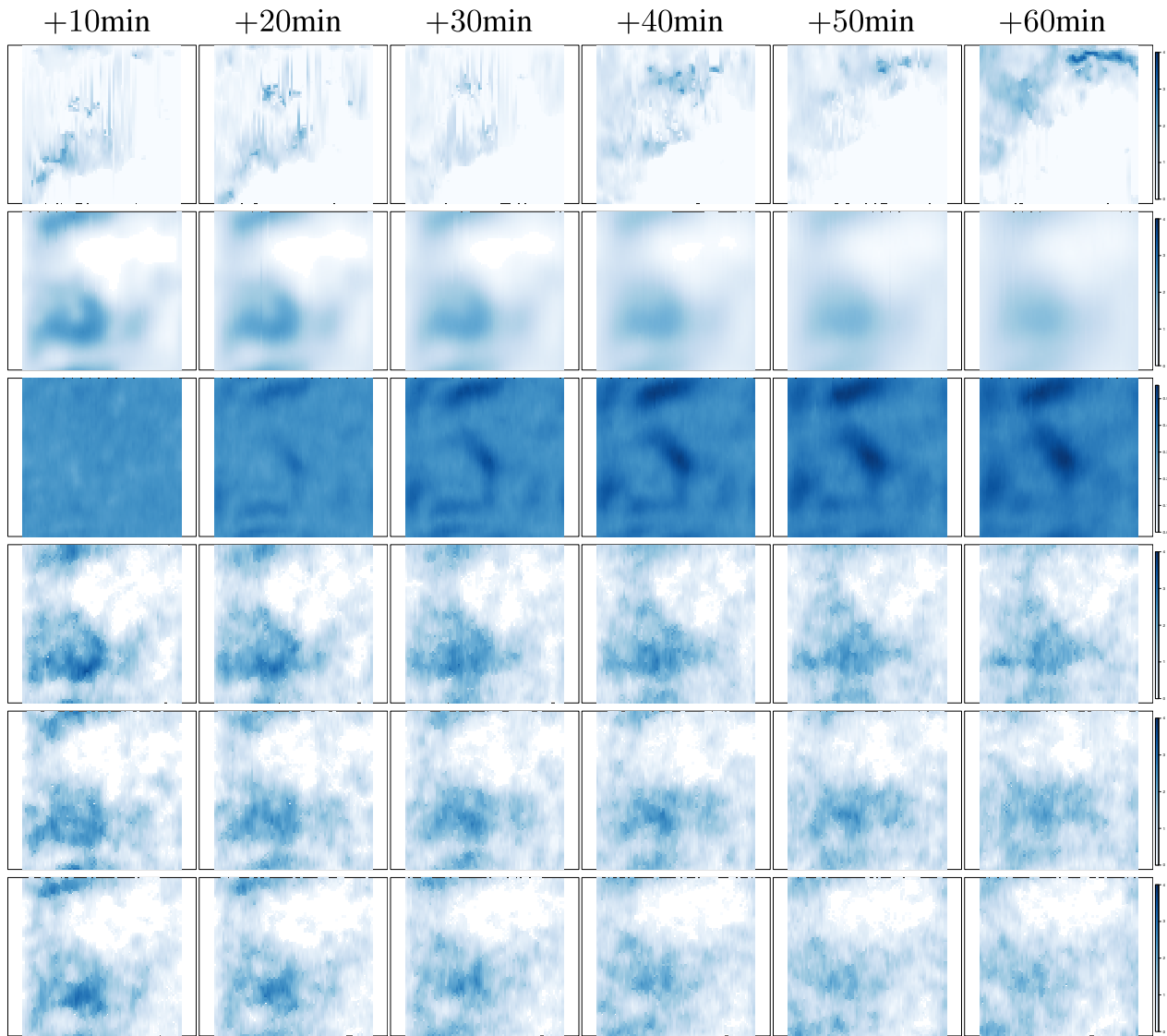


FIGURE 5 Image plots of the (transformed) radar observations at the subsequent six observation times (top row), marginal posterior predictive means and standard deviations of the corresponding precipitation field (second and third row, respectively). Rows four, five, and six each show a typical predictive draw of the precipitation field.

Turning to the dynamic states, Figure 3 shows the marginal posterior means along with the 95% credible regions for the velocity components v_x (left) and v_y (right) for $t = 1, \dots, \tilde{T}$. Inspection of the radar observations reveals that the rain clouds generally drift in a north-easterly direction and so it is pleasing to see that this feature is also observed in the inferred velocity components, with both v_x and v_y typically larger than zero. Figure 4, top row, shows image plots of the (transformed) radar observations \tilde{y}_t^r at observation times $t \in \{12, 24, 36, 48, 60, 72\}$; the corresponding plots for all time-steps are given in Section 5 of the supplementary material. The marginal posterior means ($E(\theta_{ti}|\mathcal{D})$, $i = 1, \dots, N$) of the precipitation field of interest are shown in the second row of Figure 4, from which we can see that the areas in which we expect to observe rainfall at ground level ($E(\theta_{ti}|\mathcal{D}) > 0$) correlate with the positive rainfall rates observed on the weather radar. The third row of Figure 4 shows the marginal standard deviations of the precipitation field ($SD(\theta_{ti}|\mathcal{D})$, $i = 1, \dots, N$) and from this we can see the reduced uncertainty around the rain gauge locations, which is perhaps unsurprising given the increased information at these locations. This analysis reveals that our model is able to capture the locations/times where we are unlikely to observe any precipitation; the fourth row of Figure 4 shows $\Pr(\theta_{ti} > 0|\mathcal{D})$. Finally, the bottom row of Figure 4 contains boxplots of the marginal posterior distributions of the precipitation field at the rain gauge locations, $\pi(\theta_{t,\ell_g}|\mathcal{D})$ for $g = 1, \dots, 15$; crosses show the (transformed) gauge observations \tilde{y}_t^g .

Our dynamical model (7) allows us to straightforwardly generate predictions for future time-steps by applying the model recursions; Figure 5, top row, shows image plots of the (transformed) radar observations at the next six observation times. The second row shows the marginal expectation of the precipitation field at each location, and the corresponding standard deviations are shown in row 3. The final three rows in Figure 5 each show a typical predictive draw of the precipitation field. These suggest that our model is capable of producing plausible short-term forecasts (“nowcasts”) up to around time $t + 30$ min; recalling that the radar typically provides lower rainfall rates than those observed at ground level. We note in passing that forecasts are also available on a finer temporal resolution than that of the observed data as a result of introducing the imputed time-steps. At time $t + 40$ min a large cluster of rain clouds form in the northern region of our domain, which highlights the unpredictability of weather systems even at modest forecast horizons. That our model is not able to predict that event with much certainty is perhaps not surprising, although we note that we are able to capture an increased amount of uncertainty in the region where that event is taking place; see Figure 5, third row. Thus, as noted by other authors, although both mid-range and long-range forecasts are straightforward to obtain, these should be treated with caution in radar-based modelling applications as weather systems can develop (and decay) quickly.

6 | CONCLUSION

We have considered the challenging problem of modelling spatio-temporal precipitation fields for short-term forecasting. In particular we have outlined a physically motivated multi-layered auto-regressive process model. In contrast to other approaches in the literature our observation model enables information from both ground-based rain gauges and weather radar to be used to infer the (latent) precipitation field at ground level, which is of interest when modeling rainfall-induced surface water flows in urban areas. Posterior sampling is achieved via an efficient MCMC scheme where the EnKS is embedded within a Gibbs sampling scheme that facilitates inference in reasonable time. We considered a previously unanalysed dataset from the Urban Observatory project (Newcastle upon Tyne, UK) that highlights how ground-based rain gauges can be effectively used to calibrate radar observations to infer likely rainfall rates at ground level.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in BayesianRainfallModelling at <https://github.com/srjresearch/BayesianRainfallModelling>.

ORCID

D. J. Wilkinson  <https://orcid.org/0000-0003-0736-802X>

REFERENCES

- Allard, D., & Bourotte, M. (2015). Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. *Stochastic Environmental Research and Risk Assessment*, 29, 453–462.
- Allcroft, D. J., & Glasbey, C. A. (2003). A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 52, 487–498.
- Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12), 2741–2758.
- Barr, S., Johnson, S. R., Ming, X., Peppas, M., Dong, N., Wen, Z., Robson, C., Smith, L., James, P., & Wilkinson, D. J. (2020). Flood-PREPARED: A nowcasting system for real-time impact adaption to surface water flooding in cities. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 6, 9–15.
- Berliner, L. M. (2003). Physical-statistical modeling in geophysics. *Journal of Geophysical Research: Atmospheres*, 108(D24).
- Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. Wiley.
- Bocquet, M., & Sakov, P. (2014). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140(682), 1521–1535.
- Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81(3), 541–553.
- Chatterjee, A., Michalak, A. M., Kahn, R. A., Paradise, S. R., Braverman, A. J., & Miller, C. E. (2010). A geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol optical thickness. *Journal of Geophysical Research: Atmospheres*, 115(D20).
- Chiu, G., Lehmann, E., & Bowden, J. (2013). A spatial modelling approach for the blending and error characterization of remotely sensed soil moisture products. *Journal of Environment Statistics*, 4(9).
- Cressie, N. (1991). *Statistics for spatial data*. John Wiley & Sons.
- Cressie, N., & Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

- Damien, P., & Walker, S. G. (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, 10(2), 206–215.
- Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343–367.
- Evensen, G., & Van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128(6), 1852–1867.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2), 183–202.
- Garside, L., & Wilkinson, D. J. (2003). Dynamic lattice-Markov spatio-temporal models for environmental data. *Bayesian Statistics*, 7, 535–542.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). CRC Press.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Guennebaud, G., & Jacob, B. (2010). Eigen v3. <http://eigen.tuxfamily.org>
- Heaps, S. E., Boys, R. J., & Farrow, M. (2015). Bayesian modelling of rainfall data by using non-homogeneous hidden Markov models and latent Gaussian variables. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 64(3), 543–568.
- Houtekamer, P. L., He, B., & Mitchell, H. L. (2014). Parallel implementation of an ensemble Kalman filter. *Monthly Weather Review*, 142(3), 1163–1182.
- Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1), 123–137.
- James, P. M., Dawson, R. J., Harris, N., & Jonczyk, J. (2014). Urban observatory environment. <https://doi.org/10.17634/154300-19>
- Katzfuss, M., Stroud, J. R., & Wikle, C. K. (2020). Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. *Journal of the American Statistical Association*, 115(530), 866–885.
- Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., & Senior, C. A. (2014). Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 4(7), 570–576.
- Khare, S. P., Anderson, J. L., Hoar, T. J., & Nychka, D. (2008). An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems. *Tellus A: Dynamic Meteorology and Oceanography*, 60(1), 97–112.
- Kleiber, W., Sain, S., & Harr, P. (2023). Stochastic tropical cyclone precipitation field generation. *Environmetrics*, 34(1), e2766.
- Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., & Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. CRC Press.
- Laurini, M. P. (2017). A continuous spatio-temporal model for house prices in the USA. *The Annals of Regional Science*, 58, 235–269.
- Lawson, A., Biggeri, A., & Dreassi, E. (1999). Edge effects in disease mapping. In A. B. Lawson, A. Biggeri, & D. Böhning (Eds.), *Disease Mapping and Risk Assessment for Public Health* (pp. 85–97). Wiley.
- Liang, Q., & Smith, L. S. (2015). A high-performance integrated hydrodynamic modelling system for urban flood simulations. *Journal of Hydroinformatics*, 17(4), 518–533.
- Liu, X., Yeo, K., & Lu, S. (2021). Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *Journal of the American Statistical Association*, 117(539), 1482–1499.
- Manziona, R. L., & Castrignanò, A. (2019). A geostatistical approach for multi-source data fusion to predict water table depth. *Science of the Total Environment*, 696, 133763.
- Marshall, J. S., & Palmer, W. M. K. (1948). The distribution of raindrops with size. *Journal of Meteorology*, 5(4), 165–166.
- Mastrantonio, G., Jona Lasinio, G., Pollice, A., Capotorti, G., Teodonio, L., Genova, G., & Blasi, C. (2019). A hierarchical multivariate spatio-temporal model for clustered climate data with annual cycles. *The Annals of Applied Statistics*, 13(2), 797–823.
- Mastrantonio, G., Jona Lasinio, G., Pollice, A., Teodonio, L., & Capotorti, G. (2022). A Dirichlet process model for change-point detection with multivariate bioclimatic data. *Environmetrics*, 33(1), e2699.
- Milne, W. E. (1953). *Numerical solution of differential equations* (Vol. 19). Wiley.
- Pfeifer, P. E., & Deutsch, S. J. (1980). A three-stage iterative procedure for space-time modeling. *Technometrics*, 22(1), 35–47.
- Puttaswamy, S. J., Nguyen, H. M., Braverman, A., Hu, X., & Liu, Y. (2014). Statistical data fusion of multi-sensor AOD over the continental United States. *Geocarto International*, 29(1), 48–64.
- Richards, J., & Wadsworth, J. L. (2021). Spatial deformation for nonstationary extremal dependence. *Environmetrics*, 32(5), e2671.
- Ruiz-Cárdenas, R., Krainski, E. T., & Rue, H. (2012). Direct fitting of dynamic models using integrated nested Laplace approximations – INLA. *Computational Statistics and Data Analysis*, 56, 1808–1828.
- Rundel, C. W., Schliep, E. M., Gelfand, A. E., & Holland, D. M. (2015). A data fusion approach for spatial analysis of speciated PM2.5 across time. *Environmetrics*, 26(8), 515–525.
- Sánchez-Balseca, J., & Pérez-Foguet, A. (2020). Modelling hourly spatio-temporal PM2.5 concentration in wildfire scenarios using dynamic linear models. *Atmospheric Research*, 242, 104999.
- Sansó, B., & Guenni, L. (2000). A nonstationary multisite model for rainfall. *Journal of the American Statistical Association*, 95(452), 1089–1100.
- Sigrist, F., Künsch, H. R., & Stahel, W. A. (2012). A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *The Annals of Applied Statistics*, 6(4), 1452–1477.
- Sigrist, F., Künsch, H. R., & Stahel, W. A. (2015). Stochastic partial differential equation based modelling of large space-time data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1), 3–33.
- Smith, A. F., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(1), 3–23.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639.

- Stroud, J. R., Stein, M. L., Lesht, B. M., Schwab, D. J., & Beletsky, D. (2010). An ensemble Kalman filter and smoother for satellite data assimilation. *Journal of the American Statistical Association*, 105(491), 978–990.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), 528–540.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4), 1701–1728.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, 26(1), 24–36.
- Wikle, C. K., Berliner, L. M., & Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5(2), 117–154.
- Wikle, C. K., & Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86(4), 815–829.
- Wikle, C. K., Zammit-Mangion, A., & Cressie, N. (2019). *Spatio-temporal statistics with R*. Chapman and Hall/CRC Press.
- Xia, X., & Liang, Q. (2018). A new efficient implicit scheme for discretising the stiff friction terms in the shallow water equations. *Advances in Water Resources*, 117, 87–97.
- Xing, Y., Liang, Q., Wang, G., Ming, X., & Xia, X. (2019). City-scale hydrodynamic modelling of urban flash floods: The issues of scale and resolution. *Natural Hazards*, 96(1), 473–496.
- Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954–959.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Johnson, S. R., Heaps, S. E., Wilson, K. J., & Wilkinson, D. J. (2023). A Bayesian spatio-temporal model for short-term forecasting of precipitation fields. *Environmetrics*, 34(8), e2824. <https://doi.org/10.1002/env.2824>