Data Assimilation for Precipitation Nowcasting using Bayesian Inference

Remi Barillec, Dan Cornford

Neural Computing Research Group, Aston University, Birmingham, B4 7ET, UK

Abstract

This work introduces a new variational Bayes data assimilation method for the stochastic estimation of precipitation dynamics using radar observations for short term probabilistic forecasting (nowcasting). A previously developed spatial rainfall model based on the decomposition of the observed precipitation field using a basis function expansion captures the precipitation intensity from radar images as a set of 'rain cells'. The prior distributions for the basis function parameters are carefully chosen to have a conjugate structure for the precipitation field model to allow a novel variational Bayes method to be applied to estimate the posterior distributions in closed form, based on solving an optimisation problem, in a spirit similar to 3D VAR analysis, but seeking approximations to the posterior distribution rather than simply the most probable state. A hierarchical Kalman filter is used to estimate the advection field based on the assimilated precipitation fields at two times. The model is applied to tracking precipitation dynamics in a realistic setting, using UK Met Office radar data from both a summer convective event and a winter frontal event. The performance of the model is assessed both traditionally and using probabilistic measures of fit based on ROC curves. The model is shown to provide very good assimilation characteristics, and promising forecast skill. Improvements to the forecasting scheme are discussed.

1. Introduction

Short term forecasts of precipitation fields are of critical importance to hydrologists who rely on them to prevent floods [40] or manage sewage systems in real time [56]. Because such applications are usually restricted to specific areas (typically an urbanized zone or the small scale catchment of a river) and deal with events which develop over short periods of time, the forecasting community has to provide hydrologists with precipitation forecasts at high spatial and temporal resolutions, typically referred to as *nowcasts*. According to [21], precipitation run-off models could require, in theory, forecasts down to 0.5 km spatial and 1 min temporal resolutions.

Traditional numerical weather prediction (NWP) models replicate the physics of the atmosphere, by solving a system of governing differential equations together with physics based parameterisations and extrapolating into the future. This requires heavy computational power, and considerable computing time. With the advent of ever faster computers, NWP models are starting the be run at resolutions down to 1-4 km[15; 30]. However, two main issues arise when applying NWP models to nowcasting: the existence of a spin up phase during which the incomplete representation of initial conditions affects the model's performance, and the fact that at the scales of interest, our understanding of convective processes is still very limited. Recent research projects aimed at increasing that understanding[6] will hopefully lead to a better formulation of NWP models at small scales.

Because of the reasons mentionned above, it is well acknowledged that NWP models are not particularly suited to the nowcasting problem[20; 21; 33]. Therefore alternative approaches have been developed. Constrained to run fast and at high resolution, nowcasting systems cannot rely on physics the way NWP mod-

els do. The physical equations driving the process of interest have to be approximated in some way. There are many different ways to model precipitation, ranging from image processing based techniques to purely stochastic approaches. A short review of some models found in the literature is given in Section 2.

An immediate consequence of the approximations applied in nowcasting models is their limited forecast skill. For instance, in advection based nowcasting systems such as the one presented in this work, the assumption that the rainfall evolution is dominated by motion is only valid up to a certain time, beyond which the growth and decay of cells become a non negligible source of change. The speed at which forecast skill is lost depends on the nature of the rainfall, with reasonable skill being retained for a few minutes for fast developing convective storms up to a few hours for slowly evolving fronts.

In order to ensure the nowcasting model remains coherent with the evolution of the true precipitation field, observations of the precipitation field are used to recalibrate the model (i.e. adjust its parameters) in real time. The process of incorporating observed information into the model is often called *data assimilation*.

With the advent of weather radar, frequent measurements (one observation every 5 to 15 minutes) with high spatial resolution (1 to 5 km) have become available to the forecasting community [8]. Such advantages quickly made radar data the favourite source of information for nowcasting models, greatly influencing the range of techniques developed. Although there are other sources of precipitation measurement available (raingauges, airborne sensors, satellites), this paper only discusses the problem of radar-based precipitation nowcasting, however the framework could easily incorporate other observations in the future.

In the following sections, we introduce a radar-based precipitation model which provides a fully probabilistic framework for precipitation nowcasting, with emphasis on data assimilation . The discussion is organised as follows. Section 2 gives a brief review of recent advances in precipitation nowcasting, while Section 3 outlines the concepts of probabilistic data assimilation. Section 4 then describes the model and its implementation. Results are presented and analysed in Sections 6, 7 and discussed in Section 8.

2. Precipitation nowcasting systems

There have been two main trends in the development of nowcasting methods. On the one hand, *extrap-*

olation methods predict the evolution of the observed rainfall field using object tracking and advection-based techniques, while on the other hand, storm generating methods focus on the birth, growth and dissipation of storms. The latter category can be divided further into *point process models* and *multifractal models*. Both extrapolation and storm generating methods involve, in one way or another, the following components: a spatial representation of the precipitation field, a spatial-temporal model for its evolution and a means to incorporate observation data into the model. We present an outline review of such existing methods. The reader is referred to [59; 7; 45] for more extensive reviews.

2.1. *Extrapolation methods*

Radar-based extrapolation methods rely on the assumption that, at the time scales considered, the evolution of the precipitation field is governed by motion primarily. The UK NIMROD nowcasting system [20; 21] identifies clusters of rainy pixels on radar scans and propagates them forward in time using advection estimates obtained either by cross-correlation techniques (at short lead times) or NWP wind field forecasts (at longer lead times), with a smoother merging function used to switch between these. Other systems such as the American TITAN [14] and the model of [38] also rely on object tracking methods. The precipitation field is decomposed into a set of elliptic storms, which are then matched from one image to the following. The optimal trajectory of each storm is determined by minimising the total mismatch error. Both systems provide support for the merging and splitting of storms. TITAN also estimates the growth and decay of storms using linear trends. Another correlation-based method, the TREC/COTREC systems [47; 32], partitions radar scans into smaller regions and matches these regions across subsequent scans using a maximum correlation criterion. Motion vectors are then estimated from the regions' displacement.

The NCAR Auto-Nowcast system[41] is a complex nowcasting system incorporating a variety of datasets including: radar, lightning and satellite data, wind profiles and NWP model outputs. The Auto-Nowcast system makes use of several algorithms to process these datasets into "predictor fields" (e.g. fields of reflectivity, storm growth/decay, accumulated precipitation...). Amongst the algorithms involved, the TREC method is applied to wind field retrieval from radar data and the TITAN algorithm is used to detect storms and estimate trends in their evolution. A boundary detection

algorithm, coupled with a cloud detection system, allows for convergence lines, where storms are expected to occur, to be identified. A physically-based model using fuzzy logic algorithms merges the various predictor fields into a dimensionless likelihood field which is then post-processed to obtain a final prediction estimate.

Bringing further the concept of identifying physical components of the precipitation field, the GANDOLF [46; 21] model considers individual precipitation cells with a conceptual life cycle. The precipitation objects grow and decay according to the life cycle model of [26], while being advected by wind fields estimated from an NWP model. NIMROD, TITAN, GANDOLF and other operational nowcasting systems are discussed more in detail in the context of the Sydney 2000 forecasting project in [45].

2.2. *Point process models*

Point process models put the emphasis on estimating the internal dynamics of precipitation fields using statistical representations based on point process models. The models of [48; 11; 50] follow the formalism of [31] in which the occurrence (in space and/or time) of precipitating objects (cells or storms) is governed by Poisson point processes. Storms are defined as areas of constant random precipitation, life-time[48] and velocity [11; 50]. The contributions of overlapping cells are added in both space and time. Estimation of the model's parameters is achieved by fitting to existing records of precipitation using a method of moments.

Two types of Poisson process for the generation of cells can be found in the literature: the Neyman-Scott process and the Bartlett-Lewis process. In the Neyman-Scott process, storms origins are generated from a Poisson process. Each storm is given a random number of corresponding precipitation cells, with their origin (in time) distant from the storm's by a delay drawn from an exponential distribution. Examples using the Neyman-Scott process and derived approaches include [10; 16; 43]. In the Bartlett-Lewis process, storms also originate from a Poisson process in pspace, but a second Poisson process dictates, for each storm, the time at which cells are initiated. This cell-generating process is terminated after a sample duration drawn from an exponential distribution. [44; 29; 51] are few of the many applications of the Bartlett-Lewis process to rainfall modelling. A comparative review of some of these models can be found in [57].

We note here that these models are usually not developed with forecasting or nowcasting in mind rather

being developed providing a valid statistical description of precipitation fields, which might be used for unconditional simulation purposes. In this work we are more interested in conditioning the modelled precipitation field on observations, that is data assimilation.

2.3. *Multifractal models*

In the 1980s, empirical studies have shown that rainfall fields exhibit statistical invariance with respect to the scale at which they are observed [34; 25] and, as such, could be modelled using random cascade models. In such models, the field at a given scale (i.e. spatial or temporal resolution) can be decomposed into a field at lower scale (higher resolution) by splitting unit regions into equal subregions. The intensity of the field at these subregions is determined through a random scaling which conserves the overall statistical characteristics of the field. Various options for the distribution of the scaling generators are given in [52].

The cascade methodology is naturally expressed using multifractal theory[52; 35] and presents several advantages over object tracking methods: it models the rainfall field at different scales, allows seamless incorporation of data at various resolutions (fine for radar, coarser for NWP estimates) and in the universal multifractal framework described in [52], only require a very small number of parameters to characterise their statistics. Dynamics in these models can be modelled by including time (along with space) in the cascade [13; 4] or using advection based propagation as systems like S-PROG[49] and STEPS[5].

However, a study by [55] underlines the limitations of multifractal models and shows, using several datasets, that assuming the scaling factors to be independent of the scale leads to unrealistic simulation results. The authors conclude that "rainfall does not behave like a multifractal process" and that multifractal models in which the scaling generators are identically distributed are "inadequate", and suggest directions for a new generation of multifractal models. We note that the data assimilation method we develop and apply in this paper is equally applicable to cascade like models, where the precipitation field and advection could be computed at each level in the cascade using our variational Bayesian methods.

2.4. *Other methods*

Other techniques have been applied to the problem of precipitation forecasting. [60] apply to pre-

cipitation nowcasting the probabilistic framework of integro-differential equations developed by [58]. In this framework, a redistribution kernel specifies how each pixel's intensity evolves in time as a linear combination of the previous pixels' intensities. Parameter estimation is performed using Markov Chain Monte Carlo, after projection onto the spectral domain (using Fourier decomposition), and dimension reduction (only the principal components are considered). The use of Monte Carlo methods implies a significant computational requirements making the method less suited to nowcasting. [24] tried to model the evolution of the precipitation field using a back-propagation neural network trained on real data, but showed that this method had little advantage over the more traditional advection schemes. [23] and [28] applied traditional variational techniques to estimate the parameters of a cloud model and the parameters in the Z-R relationship respectively, i.e. techniques which seek for the optimal "trajectory" (in a least squares formulation) of the parameters over a given time window covering several observations. A variational approach is also found in [17] to compute wind field estimates from sequential radar scans, which are then used to propagate the rainfall field using a semi-Lagrangian scheme (precipitation regions are advected along the estimated wind stream). Simple probabilistic forecasts at a given location can then be generated by considering the spatial structure of the neighbouring predicted field, as described in [18].

2.5. *Single point versus probabilistic estimates*

Nowcasting models provide forecasts which fall in either of the two following categories: *single point forecasts*, in which a single "best" estimate is issued and *probabilistic forecasts*, which provide a mean prediction and a measure of the prediction's uncertainty.

The motivation for probabilistic models arises from the fact that neither the data (radar) nor the models provide an exact representation of the true processes they represent. Both observation and modelling are subject to errors or approximations from many sources, as discussed in Section 3. It is thus important to be able not only to predict the evolution of precipitation events, but also to quantify how uncertain about our prediction we are.

The model described in this paper follows the stochastic approach, and extends the work of [9] with a novel data assimilation method that makes the implementation more stable and scalable, and adds a new birth and death processes for precipitation cells.

The method is tested on two large scale precipitation scenarios from convective and frontal rainfall events. Before introducing the model, we review the basic concepts of probabilistic data assimilation.

3. Data assimilation

Data assimilation denotes a set of methods used to track the state of some physical system using two sources of information: a numerical model for the physical process and observations of the true process. The observations are assumed to be available in discrete time, and are denoted $\mathbf{y}_t$. Thus, between each observation time, the initial step evolves the state from time k to time $k+1$ using the numerical model (*prediction step*), thus giving a *forecast state*. In a sequential (filtering) data assimilation method, the best possible estimate of the state at $k+1$ is obtained using the forecast state and the newly available observations (*assimilation* or *update step*), thus giving an *updated state* (also called the *analysis* state). When necessary, the forecast and analysis will be denoted using superscripts 'f' and 'a' respectively: $\mathbf{x}^f$, $\mathbf{x}^a$.

3.1. *Evolution of the state*

In a generic framework, making the often used and relatively weak assumption that the numerical model is Markovian in character , the evolution of the state is given by:

$$\mathbf{x}_{t+1} = m_t(\mathbf{x}_t) + \eta_t. \tag{1}$$

m_t is the non-linear evolution operator and η_t is a noise term accounting for the divergence of the model from the true process. Such divergence can be due to:
– an incomplete or inaccurate physical representation of reality;
– an incomplete, simplified formulation of the physical equations;
– localisation issues;
– numerical errors: approximations, linearisation, discretisation...
The noise term is commonly referred to as *model error*. We assume that model error is additive and Gaussian with mean zero and diagonal covariance $\mathbf{Q}_t$ (unbiased model): $\eta_t \sim \mathcal{N}(0, \mathbf{Q}_t)$ although arguably a state dependent, or multiplicative noise term might be more realistic, and would have little affect on the data assimilation method presented.

3.2. *Mapping to observation space*

In most practical problems, the state of the system cannot be observed directly. Instead, some other more tractable, physical quantity that relates to it is measured. The *measurement* (or *observation*) *operator* maps from state space to observation space as follows:

$$\mathbf{y}_t = h_t(\mathbf{x}_t) + \varepsilon_t. \tag{2}$$

h_t is the measurement operator, and ε_t a noise term accounting for the imperfections inherent in the observation process. This noise is commonly referred to as *observation error*. The most common error sources usually associated with radar data are [1; 22]:

– Blocking and clutter: the radar beam can be obstructed by hills or mountains (blocking) or detect ground objects (buildings, trees...) or flying objects (insect clouds, birds) which will be identified as precipitation
– Attenuation: the beam's intensity is weakened beyond stormy areas as these absorb much of the beam's energy.
– Overshooting: Due to the angle between the beam and the ground, measurements at a large distance from the radar sometimes capture precipitation aloft which does not reach the ground (evaporation) and sometimes fail to capture precipitation from low clouds (overshooting).
– Anomalous propagation: in some particular situations, the radar beam can refracted towards the ground by a layer of air, resulting in precipitation being (wrongly) detected. This is most common at early hours, when a layer of warm air lies above a layer of cooler air.
– Bright band: snow melting into rain in the atmosphere has higher reflectivity than rain and can be misinterpreted as heavy rain.
– Incorrect conversion model from observed reflectivity ($\mathbf{z}$) to rain intensity ($\mathbf{y}$). Nimrod uses the Marshall-Palmer relationship $\mathbf{z} = 200\,\mathbf{y}^{1.6}$ [19].

The NIMROD radar data undergoes intensive processing prior to being released, in order to correct for noise, clutter, occultation, anomalous propagation, attenuation, range, bright band and orographic enhancement [54]. It is worth pointing out that, although the resulting products are more accurate than the raw data, they are not error-free (e.g. overshooting errors are not corrected).

Similarly to model error, we assume observation error to be Gaussian, uncorrelated in time, with mean zero and diagonal covariance $\mathbf{R}_t$: $\varepsilon_t \sim \mathcal{N}(0, \mathbf{R}_t)$. This is a very simplified error model, as radar errors are known to be non Gaussian and spatially correlated. The choice here is motivated by the computation of the KL divergence (Section 5.2) which for a Gaussian likelihood can be computed exactly. Other, more realistic noise models will be investigated in future developments.

Note that m_t and h_t are assumed non-linear a priori.

3.3. *Probabilistic formulation*

The presence of errors in both the evolution and the assimilation steps leads naturally to a probabilistic formulation of the problem, where the interest focuses not only on the state $\mathbf{x}$, but on its probability density function (pdf) given the observations seen to that time (the filtering distribution): $p(\mathbf{x}_t|\mathbf{Y}_t) = p(\mathbf{x}_t|\mathbf{y}_t, \mathbf{y}_{t-1}, ..., \mathbf{y}_1)$.

The evolution of the state's pdf is achieved by incorporating the uncertainty due to the model's imperfection:

$$p(\mathbf{x}_t|\mathbf{Y}_{t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1})\, d\mathbf{x}_{t-1}. \tag{3}$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is determined using (1). This step is the *evolution step*.

Given a new observation vector $\mathbf{y}_t$, the probability density function of $\mathbf{x}$ is updated using Bayes' rule:

$$p(\mathbf{x}_t|\mathbf{Y}_t) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{Y}_{t-1})}{p(\mathbf{y}_t|\mathbf{Y}_{t-1})}. \tag{4}$$

where $p(\mathbf{y}_t|\mathbf{x}_t)$ is the observation's *likelihood* (measuring how likely the observation is given the model state). This step is often referred to as the *assimilation* (or *update*) *step*, and the updated distribution as the *posterior distribution*.

$p(\mathbf{x}_t|\mathbf{Y}_{t-1})$ is obtained from (3) and is typically called the *prior*. We assume that the initial prior in the absence of observation, $p(x_0|z_0) = p(x_0)$, is known, but vague. The normalising constant $p(\mathbf{y}_t|\mathbf{Y}_{t-1})$ is called the evidence, or marginal likelihood, and is defined as:

$$p(\mathbf{y}_t|\mathbf{Y}_{t-1}) = \int p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{Y}_{t-1})\, d\mathbf{x}_t. \tag{5}$$

Unfortunately, in most problems, the evidence cannot be computed and one has to revert to approximation. A common approximation consists in finding the value of $\mathbf{x}_t$ which maximises the posterior distribution (this is known as the maximum a posteriori – or MAP – estimation), and can be related to the traditional 3D and 4D variational data assimilation methods . An alternative is to approximate the posterior using sampling, in (Markov Chain) Monte Carlo approaches.

4. The Basis Function (BF) model

4.1. *Spatial representation*

We consider the following operator h to map the 2-dimensional spatial domain Ω to the precipitation intensity space. In our case, Ω is a grid with resolution 5×5 km, and total dimensions of a few hundred km. We denote M the number of pixels in Ω and $\mathbf{x}$ the input vector made of all pixels, so that $\mathbf{s} = (\mathbf{s}_1, ..., \mathbf{s}_j, ..., \mathbf{s}_M)$.

For a given observation $\mathbf{y}$, each pixel $\mathbf{s}_j$ has the associated observed precipitation intensity y_j, hence $\mathbf{y} = (y_1, ..., y_j, ..., y_M)$.

The precipitation field is modelled by a set of K 2-dimensional Gaussian-like basis functions (cells) with parameter vector $\mathbf{x}_k = (\mathbf{c}_k, w_k, r_k)$, where $\mathbf{c}_k$ is the kth cell's centre, w_k its width and r_k its central intensity (Figure 1). A cell in this context is a Gaussian-shaped precipitation volume with

Fig. 1. Rain cell

precipitation rate decreasing from the centre. Such volumes are added up in order to reproduce the observed precipitation landscape. Thus, cells are expected to overlap. Note that this differs from the traditional definition of a rain cell as an area of constant precipitation rate.

The modelled precipitation intensity at pixel $\mathbf{s}_j$ for cell k is:

$$h(\mathbf{s}_j, \mathbf{x}_k) = r_k \exp\left(\frac{\| \mathbf{c}_k - \mathbf{x}_j \|^2}{2w_k} \right) \qquad (6)$$

The total precipitation intensity at a given pixel is the sum of the intensities from all cells:

$$h(\mathbf{s}_j, \mathbf{x}) = \sum_{k=1}^{K} h(\mathbf{s}_j, \mathbf{x}_k), \qquad (7)$$

where $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_K)$ is the parameter vector for all cells.

We similarly define the precipitation field over the space domain Ω for a given cell:

$$h(\mathbf{s}, \mathbf{x}_k) = (h(\mathbf{s}_1, \mathbf{x}_k), ..., h(\mathbf{s}_M, \mathbf{x}_k)), \qquad (8)$$

and the total precipitation field over the image:

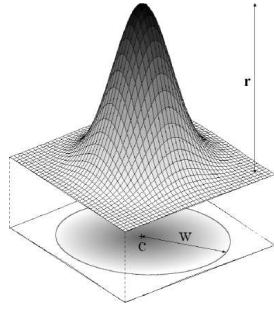$$h(\mathbf{s}, \mathbf{x}) = \sum_{k=1}^{K} h(\mathbf{s}, \mathbf{x}_k). \qquad (9)$$

The basis functions considered are continuous in space with infinite support, and thus differentiable. Although this makes gradient computations (and thus gradient-based optimisation) possible, it is an unrealistic feature for precipitation cells, which are known to cover a limited spatial area only.

The above model provides a visually satisfying static approximation to the observed precipitation field (see Figure 2).

4.2. *Dynamics*

The precipitation field is propagated in time and space according to the advection equation:

$$\frac{\partial h}{\partial t} + \mathbf{u}.\nabla h \approx 0, \qquad (10)$$

where $\mathbf{u}$ denotes the velocity *or advection* vector for the precipitation field. The equation relates the evolution of the precipitation in time and space assuming the precipitation field is approximately conserved while being advected. The advection is treated as a spatial field and is specified at each cell's location, so that $\mathbf{u} = (\mathbf{u}_1, ..., \mathbf{u}_K)$ with spatial vector $\mathbf{u}_k$ denoting the advection at cell k; thus the advection can vary spatially.

5. Data assimilation in the BF model

In this section, we detail the specific implementation of the Bayesian formalism from Section 3.3.

The observation's likelihood is defined to be Normally distributed:

$$\mathrm{p}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{M}{2}} |\mathbf{R}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{y}-h(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-h(\mathbf{x}))} \qquad (11)$$

where $\mathbf{y}$ is the observed precipitation field vector (i.e. radar), $h(\mathbf{x})$ is the model predicted precipitation field given the model state, $\mathbf{x}$, and $\mathbf{R}$ is the covariance matrix for the observation error.

The computation of the posterior distribution is made simpler by choosing prior distributions which are *conjugate* for a particular likelihood, i.e. the posterior distribution has the same distribution as the prior (e.g. if the prior is Gaussian, the posterior will be also be Gaussian too, although with different parameters). In [9] the priors were chosen for convenience, however in this work conjugate priors are used. Given a conjugate prior, the posterior distribution can be derived analytically up to its parameters, which can then be evaluated using an optimisation procedure. This *variational Bayesian* approach to the data assimilation problem is novel and
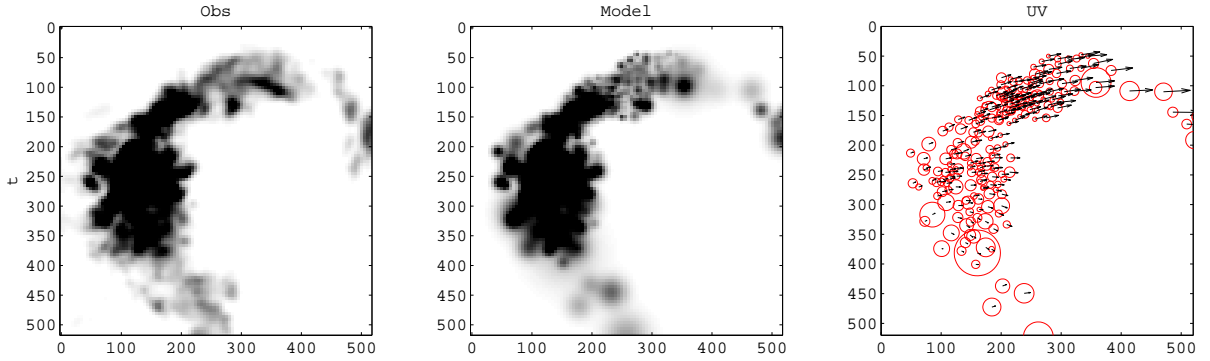
Fig. 2. A radar observation of the precipitation field (left) and the model's corresponding representation (centre), with $K = 200$ cells. The principal cell contours and advection vectors are plotted on the right.

has several benefits, the primary one being that an inference problem is converted to an optimisation problem. We stress here that we are seeking a distribution over the posterior of the parametrised precipitation field and advection field, which we define to be the state of the system.

The priors over the precipitation cell parameters have been chosen as follows. The cell widths have an Inverse-Gamma distribution:

$$\mathrm{p}(w_k) = \mathrm{IGa}(\alpha_k, \beta_k), \qquad (12)$$

the cell centres follow a Bivariate Normal distribution conditioned on the widths, up to a scale factor ξ_k (so that centres and widths together follow a Normal-Gamma distribution):

$$\mathrm{p}(\mathbf{c}_k|w_k) = \mathcal{N}(\bar{\mathbf{c}}_k, \xi_k w_k), \qquad (13)$$

and the cell intensities are Gamma distributed:

$$\mathrm{p}(r_k) = \mathrm{Ga}(\gamma_k, \delta_k). \qquad (14)$$

This choice of prior is motivated by the following considerations:
– The priors over the heights and widths should ideally have strictly positive support,
– The larger the cell, the less confident one is about the exact position of the centre, hence the conditioning on the width,
– For the normal likelihood, the Normal-Gamma prior on the centres and inverse-widths (which is equivalent to a Normal-Inverse Gamma prior on the centres and widths) is conjugate.

The distribution over the different groups of parameters are assumed uncorrelated, so that:

$$\mathrm{p}(\mathbf{x}) = \prod_{k=1}^{K} \mathrm{p}(\mathbf{c}_k|w_k)\mathrm{p}(w_k)\mathrm{p}(r_k) \qquad (15)$$

The conditioning of the centres on the widths seems a relevant assumption as locating the centre of a small cell is easier than locating the centre of a wide cell. Whether other correlations between parameters are worth capturing is open to discussion. The authors believe that a cell's intensity and width are unlikely to be strongly correlated a priori, and neither the width nor the intensity should depend (here again, a priori) on the cell's location. However, the assumption that cells are uncorrelated to each other is certainly a simplification, as we do expect cells to be correlated, at least locally. The motivation for this assumption lies in the simplifications it brings to the computation in the assimilation step (below).

The advection $\mathbf{u}$ follows a 2-dimensional Gaussian process with a Matern correlation function, with fixed order equating to a twice mean-square differentiable stochastic process [9].

5.1. *Prediction step*

The precipitation is evolved forward in time using the advection equation (10). Assuming discrete time indexed by t, the evolution of h for the pixel at location $\mathbf{s}$ is:

$$h(\mathbf{s}, \mathbf{x}_{t+1}) = h(\mathbf{s}, \mathbf{x}_t) - \mathbf{u}_t . \nabla h(\mathbf{s}, \mathbf{x}_t). \qquad (16)$$

In the BF model, the cells are assumed to have no dynamics other than motion. This is an unrealistic feature, as we know that precipitation cells also undergo modifications due to internal dynamics, resulting in growth and decay phenomena. This limitation is discussed later, however we note that in practice it has been difficult to show strong benefits for including internal cell dynamics [46] and that we do allow for additional model uncertainty due to the omission of specific internal dynamics.

Following (1), Gaussian noise is added to the model parameter distributions to account for the model imper-

7

fections. The cell heights and widths are left unchanged in mean but their covariances increase. The centres are modified according to the following:

$$p(\mathbf{c}_{t+1}|w_{t+1}) = \mathcal{N}(\bar{\mathbf{c}}_{t+1}, \Sigma_{\mathbf{c}_{t+1}}) \qquad (17)$$

where $\Sigma_{\mathbf{c}_{t+1}}$ is the full diagonal covariance matrix for the centres, i.e. $\Sigma_{\mathbf{c}_{t+1}}(k,k) = \xi_{k,t+1} w_{k,t+1}$, and

$$\bar{\mathbf{c}}_{t+1} = \bar{\mathbf{c}}_t + \Delta_t \mathbf{u}_t \qquad (18)$$

$$\Sigma_{\mathbf{c}_{t+1}} = \Sigma_{\mathbf{c}_t} + (\Delta t)^2 \Sigma_{\mathbf{u}_t} + \mathbf{Q}_t \qquad (19)$$

Similarly to the widths and heights, the advection $\mathbf{u}_t$ is assumed constant during the prediction. This is a realistic assumption at the time scales considered, where the evolution of the precipitation field is faster than that of the advection field. Again, a small amount of system error is added to reflect the inaccuracies in the model's dynamics.

5.2. *Assimilation step*

The model described thus far is able to model a precipitation field and propagate it through time in a generative manner that can be used for simulation purposes. For it to work in a forecasting context, it needs to be tied to the underlying real process via real-time observations. Section 3 described the general concepts of Bayesian data assimilation, and the need for some numerical method in order to evaluate the posterior distribution of the variable of interest. The initial model developed by [9] used a maximum a posteriori method to determine the optimal state $\mathbf{x}^*$, and a Laplace approximation about $\mathbf{x}^*$ to estimate the posterior uncertainty via the covariance matrix.

The maximum a posteriori method consists of finding the value $\mathbf{x}^*$ of $\mathbf{x}$ which maximises the posterior probability (or, as is usually done in practice, the negative logarithm of the posterior probability). The priors on the centres, widths and heights were chosen to be convenient to compute and the posterior up to a normalising constant was computed analytically; the optimum was determined using the scaled conjugate gradient optimisation method [42].

The Laplace approximation (see for example [36]) uses the fact that at its maximum, the log-posterior can be expanded using Taylor expansion as follow:

$$\ln p(\mathbf{x}) \approx \ln p(\mathbf{x}^*) - \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \mathbf{H}(\mathbf{x} - \mathbf{x}^*) \qquad (20)$$

where $\mathbf{H}$ is the Hessian of $\ln p$ at point $\mathbf{x}^*$. Then, if we denote Z the, typically unknown, normalising constant, p can be approximated at point $\mathbf{x}^*$ by the Gaussian distribution:

$$q(\mathbf{x}) = \frac{1}{Z} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \mathbf{H}(\mathbf{x} - \mathbf{x}^*)\right]. \qquad (21)$$

The Laplace assumption is only valid about the maximum in the posterior, it is important for accurate results to ensure $\mathbf{x}^*$ is as close as possible to this maximum. As the number of cells considered increases, reaching the required accuracy becomes computationally expensive, and slows down the model considerably. The approximation is also a local one, and is likely to underestimate the posterior variance, especially if the posterior distribution is multi-modal, although we have not found strong evidence for multi-modal posteriors in this problem.

In order to improve the assimilation of radar data, a new approach is presented here which relies on Variational Bayesian techniques. Variational Bayes was introduced by [27] in a paper which showed that the true posterior distribution can be approximated by a Gaussian distribution using a deterministic algorithm. The method can be applied to other types of distributions, and involves minimising the Kullback-Leibler (KL) divergence between the true posterior and some simpler, parametric, approximating distribution, for which parameters are estimated. The KL divergence, or relative entropy, between the posterior p and the approximating distribution q, is defined as follows:

$$\mathrm{KL}(q \parallel p) = -\int q(\mathbf{x}_t|\mathbf{Y}_t)\ \ln \frac{p(\mathbf{x}_t|\mathbf{Y}_t)}{q(\mathbf{x}_t|\mathbf{Y}_t)}\ d\mathbf{x} \qquad (22)$$

Given the conjugate nature of the priors chosen, the q distribution has the same structure as the prior. Given the prior (eq. 15) and the likelihood (eq. 11), the expression for $p(\mathbf{x}_t|\mathbf{Y}_t)$ can be obtained, up to a constant, by applying Bayes' rule (Eq. 4), leading to:

$$\mathrm{KL}(q \parallel p) \propto -\int q(\mathbf{x}_t|\mathbf{Y}_t)\ \ln \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{Y}_{t-1})}{q(\mathbf{x}_t|\mathbf{Y}_t)}\ d\mathbf{x} \qquad (23)$$

Expanding this yields:

$$\mathrm{KL}(p \parallel q) \propto -\left\langle \ln p(\mathbf{y}_t|\mathbf{x}_t) \right\rangle_{q(\mathbf{x}_t|\mathbf{Y}_t)}$$
$$-\left\langle \ln \frac{p(\mathbf{x}_t|\mathbf{Y}_{t-1})}{q(\mathbf{x}_t|\mathbf{Y}_t)} \right\rangle_{q(\mathbf{x}_t|\mathbf{Y}_t)} \qquad (24)$$

where the cross-entropy of p with respect to q is defined as:

$$\left\langle \ln p(\mathbf{x}) \right\rangle_{q(\mathbf{x})} = -\int q(\mathbf{x})\ \ln p(\mathbf{x})\ d\mathbf{x} \qquad (25)$$

The likelihood term arising from the new radar image being assimilated expands, after computation, and omitting the time index for readability, to:

$$-\left\langle \ln p(\mathbf{y}|\mathbf{x}) \right\rangle_{q(\mathbf{x}|\mathbf{Y})} = \frac{1}{2\sigma^2} \sum_{j=1}^{M} \left[\left(\sum_{k=1}^{N} E_{k,j} - y_j \right)^2 \right.$$
$$\left. - \sum_{k=1}^{N} \left(E_{k,j} \right)^2 + \sum_{k=1}^{N} F_{k,j} \right] \tag{26}$$

with

$$E_{k,j} = \frac{\gamma_k}{\delta_k(1+\xi_k)} \left(1 + \frac{(\bar{\mathbf{c}}_k - \mathbf{s}_j)^{\mathrm{T}}(\bar{\mathbf{c}}_k - \mathbf{s}_j)}{2\beta_k(1+\xi_k)} \right)^{-\alpha_k} \tag{27}$$

and

$$F_{k,j} = \frac{\gamma_k(\gamma_k+1)}{\delta_k^2(1+2\xi_k)} \left(1 + \frac{(\bar{\mathbf{c}}_k - \mathbf{s}_j)^{\mathrm{T}}(\bar{\mathbf{c}}_k - \mathbf{s}_j)}{\beta_k(1+2\xi_k)} \right)^{-\alpha_k} \tag{28}$$

Further computation leads to the following results for the prior term:

$$-\left\langle \ln \frac{p(\mathbf{x}|\mathbf{Y})}{q(\mathbf{x}|\mathbf{Y})} \right\rangle_{q(\mathbf{x}|\mathbf{Y})} =$$
$$\sum_{k=1}^{N} \left[\gamma_k \ln \delta_k - \gamma_k' \ln \delta_k' - \ln \frac{\Gamma(\gamma_k)}{\Gamma(\gamma_k')} \right.$$
$$+ (\gamma_k - \gamma_k') \left[\Psi(\gamma_k) - \ln \delta_k \right] - \gamma_k \left(1 - \frac{\delta_k'}{\delta_k} \right)$$
$$+ \alpha_k \ln \beta_k - \alpha_k' \ln \beta_k' - \ln \frac{\Gamma(\alpha_k)}{\Gamma(\alpha_k')} \tag{29}$$
$$+ (\alpha_k - \alpha_k') \left[\Psi(\alpha_k) - \ln \beta_k \right] - \alpha_k \left(1 - \frac{\beta_k'}{\beta_k} \right)$$
$$+ \ln \frac{\xi_k'}{\xi_k} + \frac{\xi_k}{\xi_k'} + \frac{1}{2\xi_k'} \frac{\alpha_k}{\beta_k} (\bar{\mathbf{c}}_k - \bar{\mathbf{c}}_k')^{\mathrm{T}}(\bar{\mathbf{c}}_k - \bar{\mathbf{c}}_k')$$
$$\left. \right] - \frac{N}{2}$$

in which the prime is used to distinguish the prior parameters from the equivalent parameters in the posterior. The Γ and Ψ operators are the standard Gamma function and its first derivative (Digamma) . We omit the full derivation of these results which requires a series of non-trivial integrals. These will be made available in a future report. The main point to note is that the use of the variational Bayes framework allows us to repose the estimation of the posterior distribution by a minimisation of Equation 24. The minimisation is achieved using a scaled conjugate gradient algorithm [3; 42]. Since the q distribution is an approximation to the posterior distribution we are able to take relatively few optimisation steps and still have reasonable estimates of the posterior uncertainty about the locations of the precipitation cells in particular, which is crucial in obtaining realistic estimates of the advection field, and our uncertainty in this.

5.3. Forecast

The predicted distribution at a given lag L is given by:

$$p(\mathbf{x}_{t+L}|\mathbf{Y}_t) = \int p(\mathbf{x}_{t+L}|\mathbf{x}_t) \, p(\mathbf{x}_t|\mathbf{Y}_t) \, d\mathbf{x}_t. \tag{30}$$

Since the model is assumed Markov, the predicted distribution in the integral factorises as the product of all intermediate predicted estimates, giving:

$$p(\mathbf{x}_{t+L}|\mathbf{Y}_t) = \int .. \int p(\mathbf{x}_{t+L}|\mathbf{x}_{t+L-1}) \ldots p(\mathbf{x}_{t+1}|\mathbf{x}_t)$$
$$p(\mathbf{x}_t|\mathbf{Y}_t) \, d\mathbf{x}_{L-1} \ldots d\mathbf{x}_t. \tag{31}$$

Although this formulation allows for the full distribution of the state to be forecast forward in time according to Equation (1), it relies heavily on a good specification of the model error. Because this model error has been chosen Gaussian, which we know is unrealistic, an alternative approach based on Monte Carlo approximation is preferred. A number of realisations are sampled from the state's distribution $p(\mathbf{x}_t|\mathbf{Y}_t)$ and propagated to time $t+L$ using the model. The forecast distribution is then approximated by the ensemble of forecasts, from which statistics can be estimated. This method is expected to better capture the variability introduced by the model and compensate for the crude error model chosen.

The forecast is obtained by propagating the cells linearly given their initial advection estimate. This is a very simple advection model, and it is expected to give reasonable forecasts at short times only (up to 1h), where the linear approximation holds. At longer lead times, a better forecasting scheme would have the cells propagated according to the estimated advection at their predicted location, rather than at their initial location. This has not been been implemented in the current work, but will be the primary focus of upcoming developments.

5.4. Extensions to the model

In order to apply the model to real data, a number of additional extensions to the model in [9] were required. It is beyond the scope of this paper to discuss these in detail, but it is worth mentioning them:
– Support for the deletion of "dead" cells (cells that have collapsed or moved out of the observable area).
– Support for the detection of new cells (cells that have appeared as a result of in situ convective development or that have been advected into the observable area).
– Localisation of each cell's precipitating area: by design, the cells have infinite support. It is thus important to limit their effect to a certain radius from

the centre. A cut-off of 0.8 mm.h^{-1} is applied, beyond which the effect of the cell is ignored. This also speeds up the computation considerably.
These extensions make is possible to run the model on large radar images, over long time periods.

6. Preliminary results

As a sanity check, we first test the ability of the model to correctly locate a single precipitation cell, on simulated data. The data is generated by propagating a single precipitation cell with constant, horizontal advection of 6 m.s^{-1}. The cell is 42 km wide and its central intensity is 33 mm.h^{-1}. Observations are taken every 900 s and corrupted with Gaussian white noise with variance 4.0 mm^2.h^{-2}. The noise is applied locally, i.e. only to those pixels where the precipitation rate exceeds some threshold (0.8 mm.h^{-1}). This is believed to be a realistic noise model for simulated radar-like data, since it is empirically observed that radar is usually well able to identify geographical regions without precipitation, with the errors occurring mainly in the regions where precipitation is detected [39]. Other noise models were tested (global additive noise and multiplicative noise), but they did not lead to any significant differences in the results.
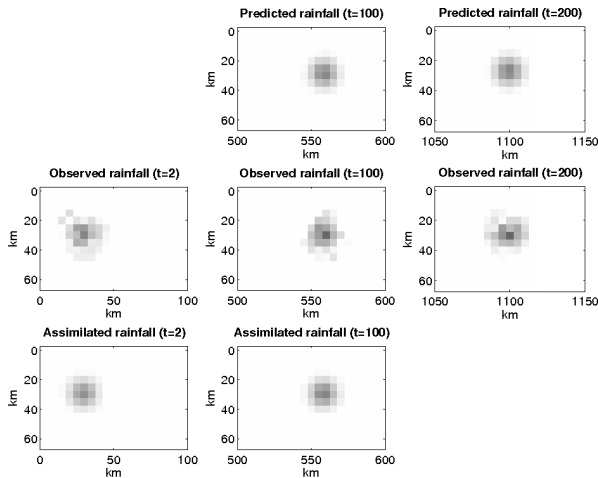
Fig. 3. Single precipitation cell: movement in time – The top row shows the predicted precipitation cell (top), the observed precipitation cell (middle) and the assimilated precipitation cell (bottom) at 3 different times: initial guess (left), assimilation (centre) and end of forecast (right).

The cell parameters are estimated over 180 time steps (equating to 45 hours), and then forecast over 12 steps (3 hours). The initial variances over the centres, widths and heights are set to 10.0, representing weak knowledge of the initial state at the start of the assimilation. The propagation error variance (also known as the model error component) is set to 25 km^2.h^{-1} for the centres, 0.1 mm^2.h^{-1} for the central intensity, 0.1 km^2.h^{-1} for the cell widths and 1.6 m^2.s^{-2}.h^{-1} for the advection field. These priors are chosen to be characteristic of the typical model errors we might expect for real data, and have been derived by analysis of radar image sequences.

Figure 3 shows the predicted, observed and assimilated cells from the idealised model at 3 different times. At the end of the assimilation period, the cell is propagated forward in time using the last estimate of the parameters. There is no noticeable difference between the cell's characteristics at the end of the forecast and the observed cell, which confirms the model managed to track the "true" cell on this very simple validation example.

7. Results on real data

In order to test the model on real data, the model is run against radar observations from a convective precipitation event in the summer of 2006 and to contrast on a frontal precipitation event in the winter of 2005. Both experiments use NIMROD radar data provided by the UK Met Office, accessed through the British Atmospheric Data Centre [54].

The observations were pre-processed using a Gaussian filter with radius 10 km to improve the estimation of the model precipitation field. The spatial domain was restricted to 500×500 km (100×100 pixels) and the model was run with a limited number of 250 cells, in order to keep computation times sufficiently low (less than 5 minutes per observation processed). Furthermore, the optimisation of the KL divergence was limited to 200 iterations, which proved sufficient in practice to achieve a good fit of the posterior distribution while remaining computationally efficient.

The assimilation was carried out over a sequence of 672 observations, taken every 15 minutes, which amounts to a weeks worth of data (168 hours). Figure 5 show the state of the model for the convective event, after about 67 hours of data (270 observations) have been assimilated. 4 hourly estimates are shown. This corresponds to a peak in the complexity of the precipitation field, and thus to a peak in the Root Mean Square Error (RMSE) with respect to the mean of the posterior distribution, as shown in Figure 8. Reading from top to bottom Figure 5 shows: the radar observations of precipitation, the corresponding estimated precipitation after assimilation of the observed radar, the cell contours
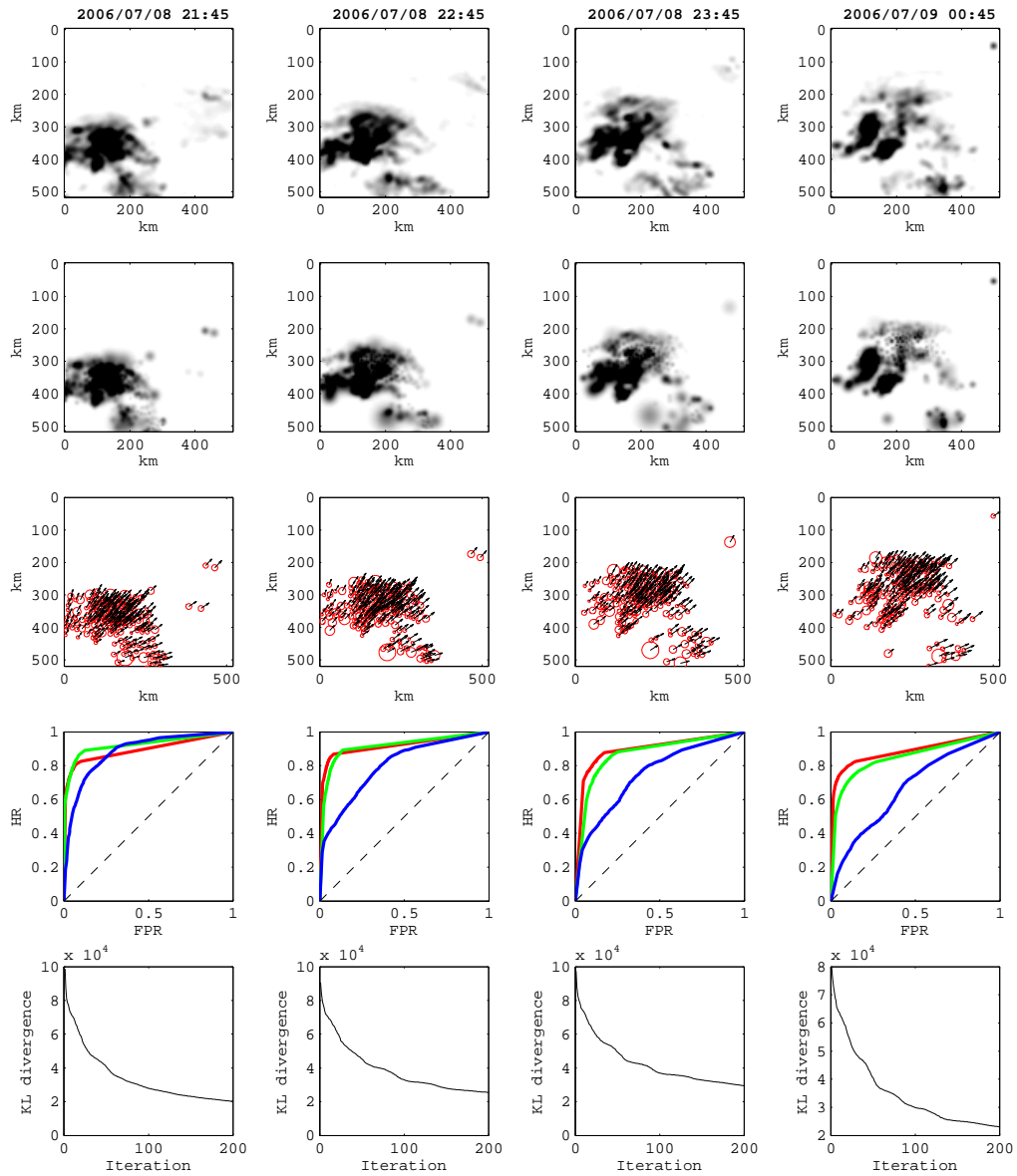
Fig. 4. Convective event – This plot shows the evolution of the observed (a) and estimated (b) rainfall over a 4-hour period (from left to right). The main cells and their associated advection are shown on row (c). ROC curves for a fixed threshold of 1 mm.h^{-1} are shown on row (d) for 3 different forecast lead times: 30 min (top curve), 60 min (middle curve) and 180 min (bottom curve). The dashed line correspond to a non informative forecast (no skill). Row (e) shows the optimisation of the KL divergence over 200 optimisation steps.

Fig. 5. Estimation of convective precipitation (July 2006)

11

along with their advection vectors (only the major cells are displayed for clarity), the ROC curves for a precipitation threshold $R_{min} = 1$ mm.h^{-1} with 3 curves correspond, from the upper one to the lower one, to the forecasts at t+30 min, t+60 min and t+180 min, and, on the bottom row, the KL convergence curves for each of the 4 time steps considered.

The quality of the probabilistic forecasts was assessed based on [2]. At each 15 minute time step, 100 realisations of the stochastic model were generated from the parameters posterior distribution and then propagated forward in time, to provide an Monte Carlo (or large ensemble) estimate of the probability distribution of the precipitation rates over the region at times from t+15min to t+180 min. Given a precipitation threshold R_{min}, the model is considered to detect rain at a given location if at least N forecasts out of the 100 predict rain exceeding R_{min} at that location. The same threshold is applied to the observed rainfall and the Hit Rate (ratio of correctly detected wet locations) is plotted against the False Prediction Rate (ratio of incorrectly detected dry locations). The ROC curve is generated by having N vary from 0 to 100. Ideally the all forecasts for a given N will have 100% hit rates, and zero false alarms, corresponding to a line along the top of the plot. A random forecast should plot along the diagonal line. Thus a perfect probabilistic forecast will result in the area between the curve and the diagonal [0,1] being 0.5, whereas an area close to zero indicates poor skill (zero being equivalent to a random forecast). The area under the curve is computed as the sum of the "slices" between any two sequential points a_{k-1} and a_k on the curve, joined by a straight line (trapezoidal Riemann sum). The total area is thus given by:

$$\frac{1}{2} \sum_{k=2}^{N} [FPR(a_k) - FPR(a_{k-1})] \times \atop [HR(a_k) + HR(a_{k-1})] \tag{32}$$

Figure 5 shows that after seeing 67 hours of data the model is able to assimilate the radar data to estimate the precipitation field while also jointly estimate advection vectors for the precipitation field. It can be noticed that the appears somewhat 'spotty' in some parts. We believe this relates to the partial optimisation of the KL-divergence; as can be seen from the bottom line of the figure, the KL-divergence has not converged full in the optimisation, much like the 3D VAR cost function is only partially optimised in classical data assimilation. This might also be related to the convective nature of the event, with the birth / death processes of the 'precipitation cells' making it very difficult for the model

to track specific precipitation features, and resulting in cells being frequently added and removed. The advection vectors show a clear storm motion from south-west to north-east, but there are small variations in the advection over space, which are likely to reflect differential development and possibly steering of the precipitation field. The ROC curves (4th line) for the four times show that at moderate lead times t+30 min and t+60 min the probabilistic predictions have considerable skill, however for this convective event at t+180 min the skill is much reduced, but still considerably better than random.

Figure 7, shows the same information as Figure 5 but for the January 2005 frontal rainfall event, starting this time after 60 hours of data (240 observations) have been assimilated. We note that the assimilation method quickly converges and thus could be shown after only 2 cycles of assimilation and would show similar results. We again see a good fit the the assimilated precipitation field, but again note a problem with some rather 'spotty' behaviour in the assimilated estimate of precipitation. This problem appears to be most severe in region with strong dynamic changes to the advected precipitation field. The advection vectors from this example show rather complex structure. Initially this was felt to be a problem with the model, however it appears that there is strong apparent differential advection in this storm, probably related to embedded precipitation elements within the frontal zone, particularly early in the time window show here. At the later times the advection seems more uniform across the domain considered. In contrast to the convective scenario the the ROC curves show that forecast skill is retained far longer, with quite respectable skill shown even at t+180 min as might be expected in the more strongly dynamically forced, larger scale processes typical in frontal precipitation. Finally we note that for computational reasons we truncated the optimisation of the KL-divergence at 200 iterations, but it is clear that the system has not fully converged.

8. Discussion

To further evaluate the performance of the model, two statistical validation methods have been applied. The quality of the assimilation was validated using the traditional Root Mean Square Error (RMSE) while the forecasting ability was assessed using ROC curve based verification statistics over the full assimilation window.

The RMSE measures the average distance (over the spatial domain) between the model's estimate and the 'true' observed precipitation field as measured by the
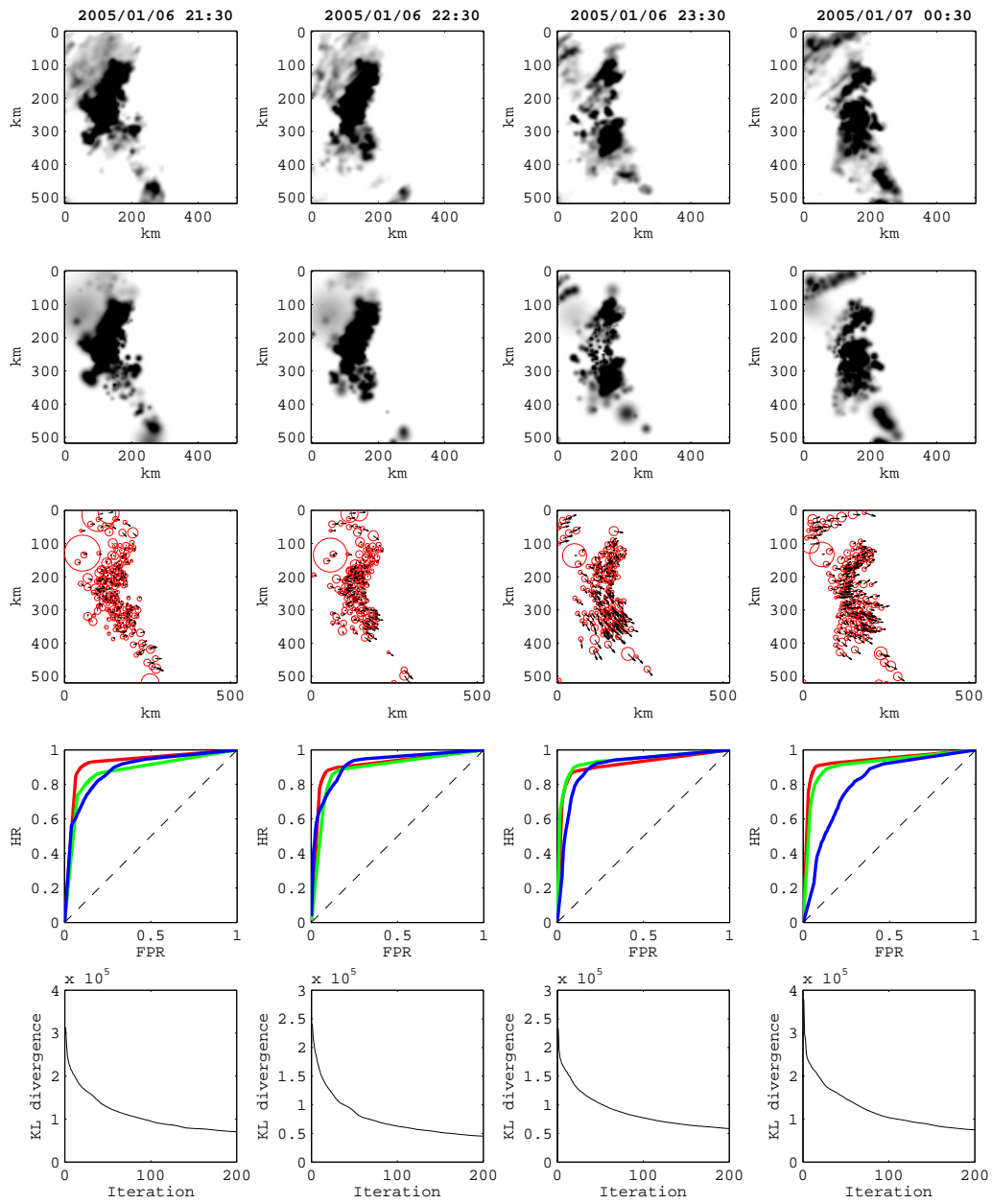
Fig. 6. Frontal event – This plot shows the evolution of the observed (a) and estimated (b) rainfall over a 4-hour period (from left to right). The main cells and their associated advection are shown on row (c). ROC curves for a fixed threshold of 1 mm.h^{-1} are shown on row (d) for 3 different forecast lead times: 30 min (top curve), 60 min (middle curve) and 180 min (bottom curve). The dashed line correspond to a non informative forecast (no skill). Row (e) shows the optimisation of the KL divergence over 200 optimisation steps.

Fig. 7. Estimation of frontal precipitation (January 2005)
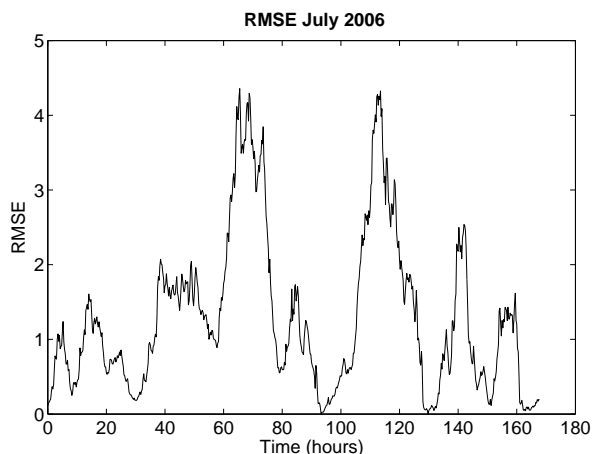
Fig. 8. Convective event: Root Mean Square Error
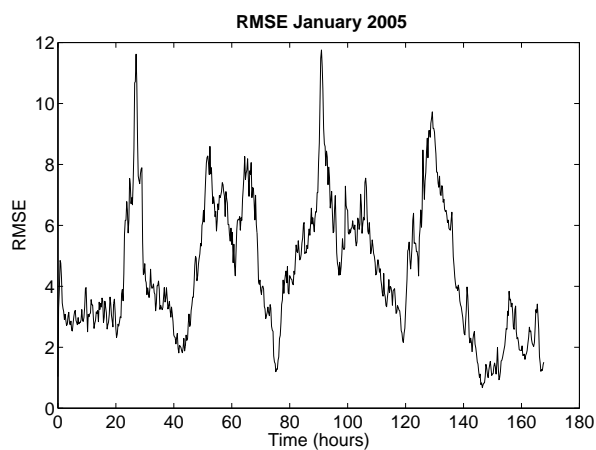
Fig. 9. Frontal event: Root Mean Square Error
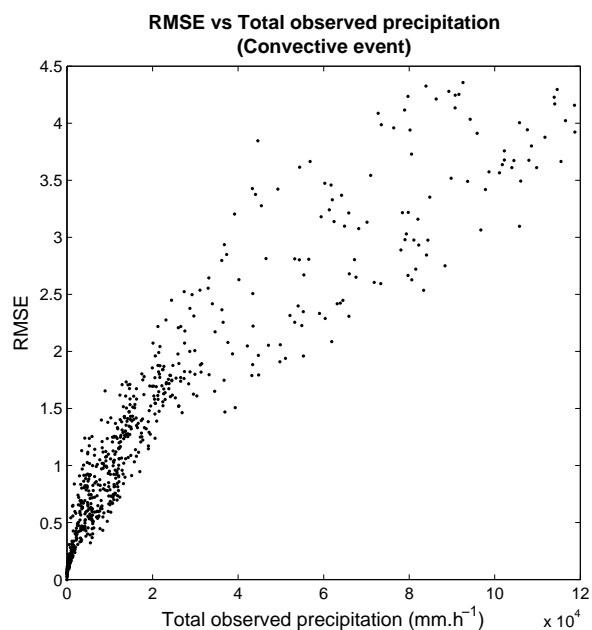
Fig. 10. Scatter plot of RMSE vs Total observed precipitation (July 2006) – The Root Mean Square Error between the estimated rainfield and the observation is plotted against the total observed precipitation for each of the 672 observations in the convective experiment.
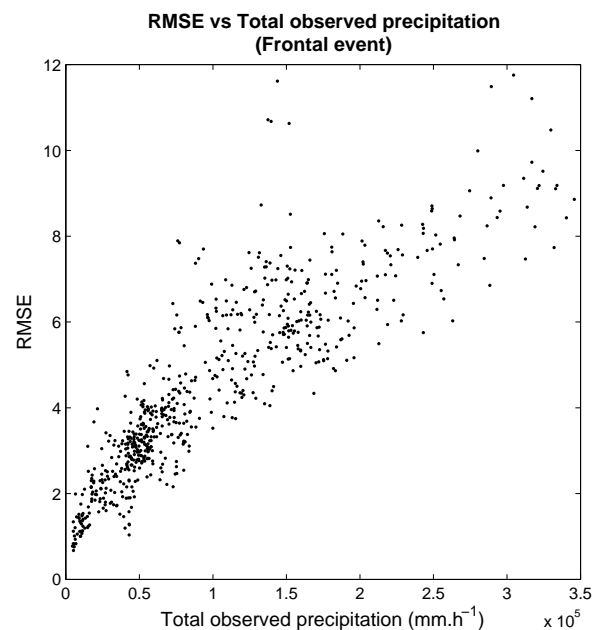
Fig. 11. Same as Figure 10 for the frontal event (January 2005).

radar. Figures 8 and 9 show the evolution of the RMSE in time over the assimilation period. The assimilation appears to give better results for the convective event (Figure 8) than for the frontal event (Figure 9). This is to be related to the spatial complexity of the precipitation field, which is greater in the winter event, probably due to the higher overall rain rates, and the greater part of the domain covered by precipitation compared to the summer event. Figures 10 and 11 show scatter plots of the RMSE against the total observed precipitation

(summed over the spatial domain) for the convective and frontal cases respectively. This confirms, in both cases, the correlation between the complexity of the precipitation field and the quality of the corresponding estimate.

For both cases the same, limited, number of cells (250) was used which is probably not overly realistic. In practice it would be very desirable to be able to estimate an appropriate complexity for the model, which could adapt to the complexity of the observations. This was
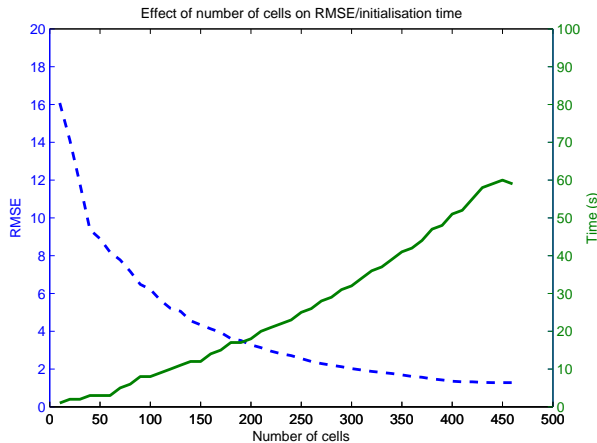
14

Fig. 12. Trade off between number of cells and initialisation speed. This plot shows the effect of the number of cells used on the quality of the fit and computation time. The dashed curve (left y-axis) is the root mean square error between the observation and the model's estimated rainfall field. The solid curve (left y-axis) is the initialisation time in seconds.

Fig. 13. Fit to the data with increasing number of cells. This plot shows the model's realisation as the number of cells increases. The actual observation is shown in the top left corner.

not implemented in this version of the code but we believe that it might be possible to incorporate a 'sparsity' prior in the model, in a similar manner to the treatment of the relevance vector machine [53]. The model is this better able to represent the simple, *localised* precipitation patterns which make up most of the convective data compared to the complex, diffuse precipitation patterns from the winter data. This also explains the considerable variations in the RMSE for each event, where quiet, dry(er) periods alternate with stormy phases.

The fit to the data can be improved by increasing the number of cells, but this comes at the price of computational time, which needs to be kept below the time increments between two observations. Figure 12 shows how the Root Mean Square Error (dashed line) decreases as more and more cells are allowed into the model. Computation time (solid line) increases almost linearly with the number of cells. A limit of 250 cells is sufficient to discard more than 90% of the misfit while keeping computation time below 30 seconds. Figure 13 shows how the fit to the data improves as the number of cells increases. In this work the complexity was chosen to keep the computational time to less than 5 minutes per assimilation cycle on a 1.6 GHz single core desktop PC. It is theoretically possible to parallelise many of the computations in the algorithm to exploit future multi-core machines or dedicated parallel architectures if the method were taken into operational use.

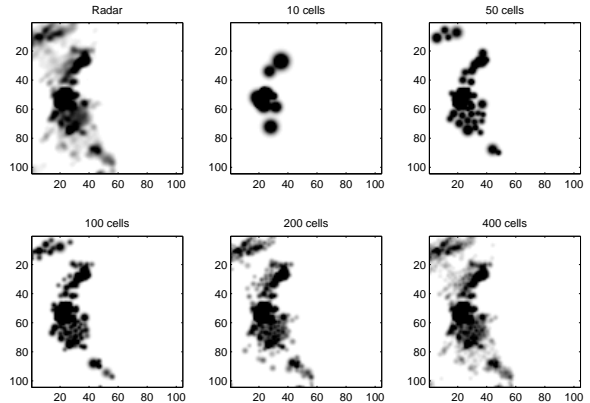The RMSE results only tell half the story, since this

model is designed to be probabilistic. The most widely used diagnostic for probabilistic models in precipitation forecasting is the area under the ROC curve. Figures 14 and 15 show the evolution of the area under the ROC curve as a function of time. Results are plotted, from top to bottom, for the 3 precipitation intensity thresholds (1, 5, and 10 mm.h^{-1}). Each plot shows the evolution of the area for the 3 forecast lead times (t+30 min, t+60 min and t+180 min).

Note that the absence of rain (or heavy rain) during some periods can result in erroneous ROC curves. If no rain is observed, the Hit Rate cannot be computed (division by zero). This results in missing points on the curve as can be seen on the bottom plot around t = 130h. In the case where no rain cell is left in the model (as a result from a dry observation being assimilated), all of the model's realisations will predict a dry forecast, hence all points will coincide with the bottom left corner ($HR = FAR = 0$), resulting in a curve aligned with the diagonal and an area of exactly 0.5. Several such cases can be observed in the two lower plots, where the detection thresholds is higher. These cases should ideally be discarded as they don't give a correct account of the model's prediction skill.

Figure 15 shows similar information to Figure 14 for the frontal event. It can be observed that the prediction skill varies more smoothly that in the convective case, due to the larger scale and slower nature of the precipitative developments. Two regimes can be identified,
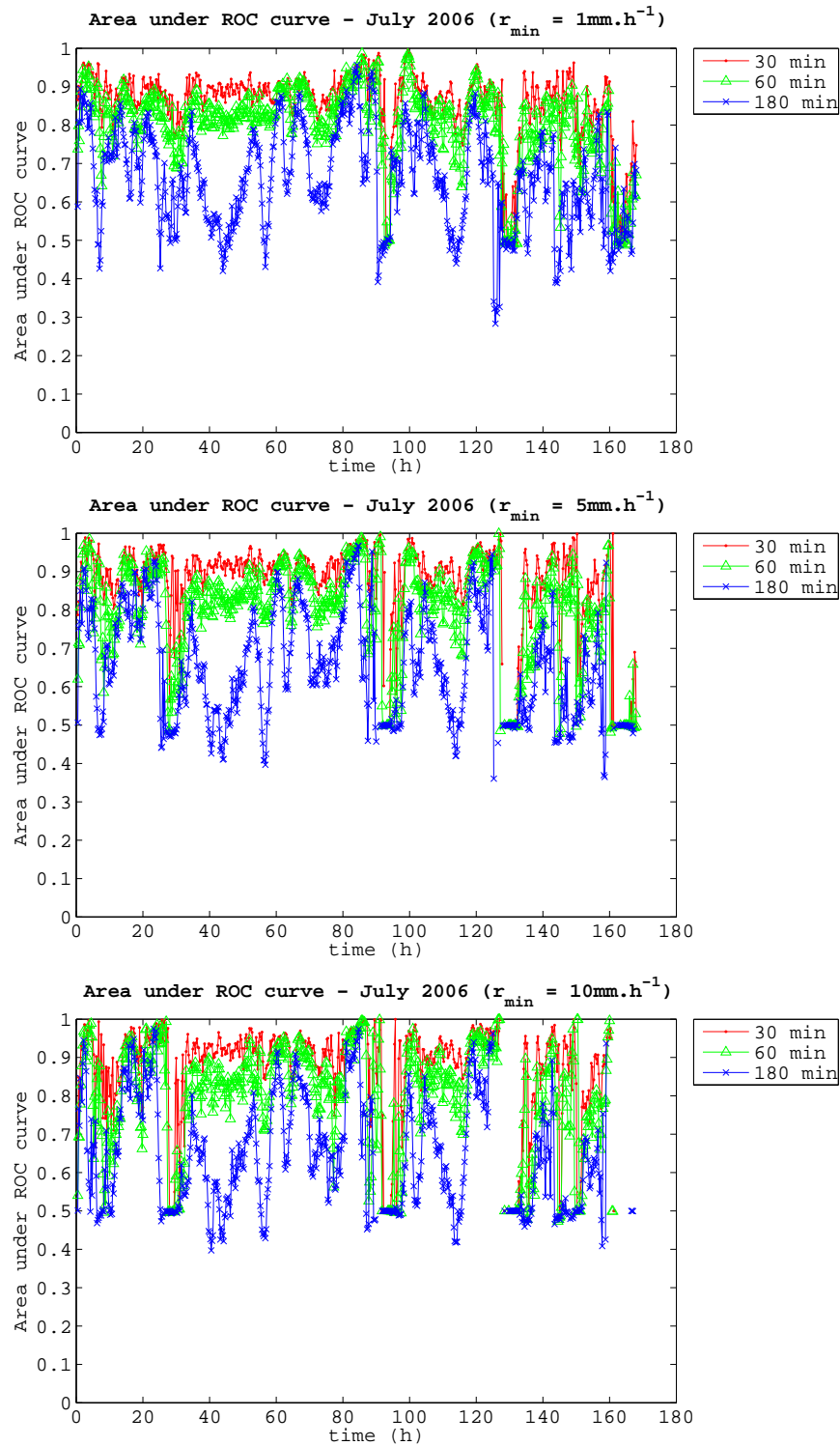
15

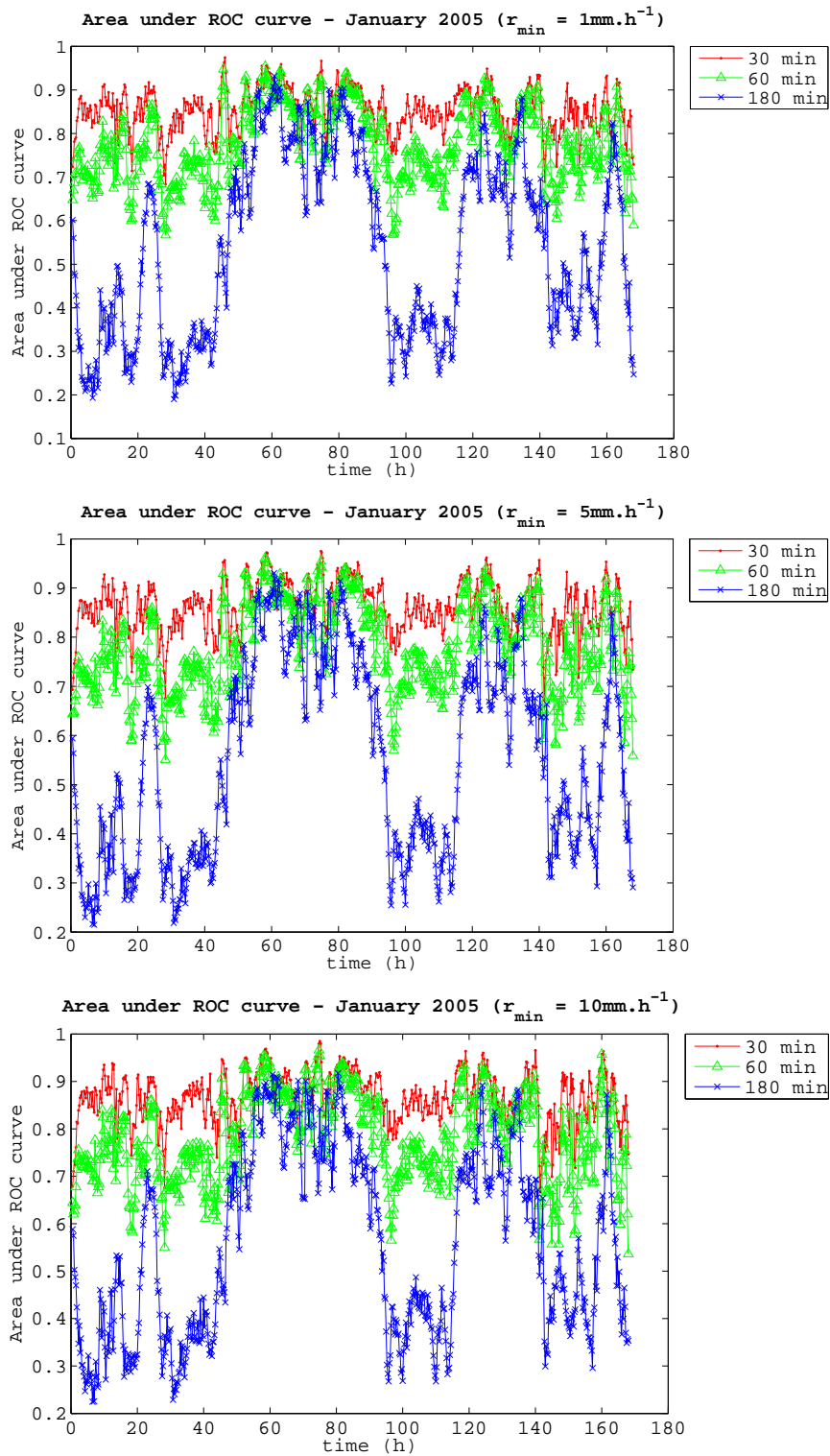Fig. 14. Evolution of the area under ROC curve (July 2006)

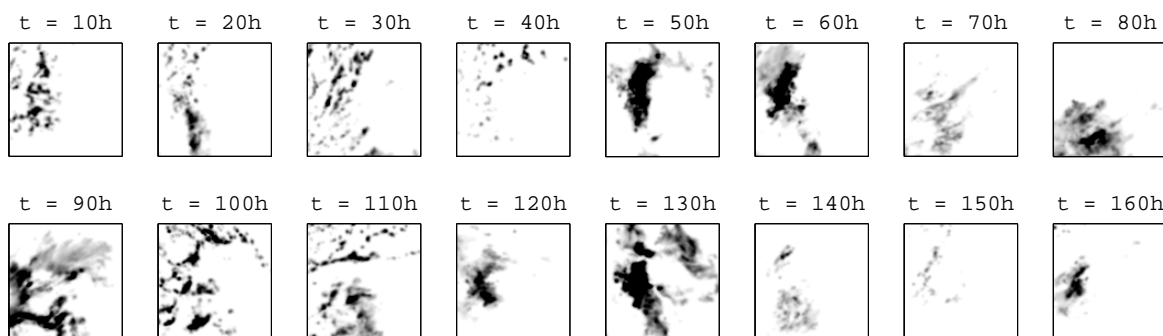Fig. 15. Evolution of the area under ROC curve (January 2005)

Fig. 16. Evolution of the rainfall field during the frontal event (January 2005, 10h snapshots).

one in which the prediction skill decreases quickly (t=0-40, t=100-120, t=140-160) and one where the prediction skill is retained much longer, to the point that even 180min forecasts still show some good skill (t=60-90, t=120-140).

Qualitative analysis has shown that these two regimes can be related to the nature of the observed rainfall field. Figure 16 displays the observed rainfall field at 10h intervals. Heavily clustered fields of high precipitation intensity seem to result in good forecasts while sparse, localised precipitation fronts correspond to poor forecasts. This could be due to the linear forecasting scheme, which is likely to perform better when the rain cells are clustered as this ensures their advection field is smooth. Another possible explanation is the assumption, in advection-based forecasts, that motion is the primary factor of change, and that internal development (growth/decay) can be neglected. It is clear that dissipation phenomena are less noticeable, in proportion, for large areas of intense precipitation than for smaller isolated cells.

Figures 17 and 18 summarise the distributions, for the whole experiment, of the areas under the ROC curve. As expected, the model skill decreases on average as the forecast lead time increases, with very little skill in any of the forecasts after 3 hours. This is due partly to the simplicity of the precipitation field representation, and partly to the linear nature of the forecast scheme applied. Each cell is advected linearly given the advection at its centre, an assumption which remains relevant only for shorter forecast lead times. A better advection scheme would involve projecting the advection field onto a fixed grid and advecting the cells according to the advection at their future location rather than at their initial location. At short time scales, particular at t+30 min the model has more skill when forecasting

heavy precipitation (>10 mm.h^{-1}) than light precipitation (1 mm.h^{-1}). This is a useful feature, since for most flood forecasting applications the heavy precipitation is the most important to predict well. This is a common feature of many advection models, since the heavier precipitation tends to be more temporally persistent.

Figure 19 shows the variogram [12; 37] and for the observed and assimilated rainfall field (top left) at time $t = 70$ h into the experiment. Variograms for the corresponding forecasts from times t-30 min (bottom left), t-60 min (top right) and t-180 min (bottom right) are shown, along with 4 random sample forecast realisations. The predicted rainfall field at +30 min provides a visually satisfying estimate of the observed rainfall field, with sufficient detail. The +60 min forecast still captures the overall precipitation structure, although inaccurately predicts rain in the south-eastern part of the domain, most certainly due to previous observed precipitation which has since then dissipated. The +180 min forecast is rather poor, due to the linear advection scheme which does not maintain, at longer lead times, the cohesion of the field (as observed from the "spotty" nature of the forecast). However, the variograms show that the model is able to retain the overall structure at short scales (< 100 km), but quickly diverges at larger scales in that example.

9. Conclusions

This paper presents a new probabilistic data assimilation algorithm which can be applied to nowcasting using a simple advection based precipitation forecasting model. The algorithm has several desirable features, in particular the ability to estimate the posterior distribution of the model state using optimisation methods, which provides control over the computational com-
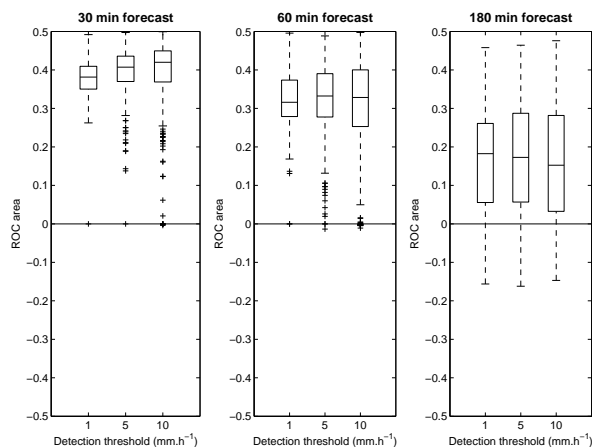
18

Fig. 17. Convective event: ROC areas

Fig. 18. Frontal event: ROC areas

plexity. While the intial derivation is quite mathematically demanding, the implementation can be employed within any optimisation framework, which forms the basis for most traditional variational assimilation methods.

The new method is tested on two large events characterised by convective and frontal dominated rainfall. The results show the model is numerically robust, although further testing is necessary to assess its applicability to operational nowcasting . The ROC curves show probabilistic skill at all forecast horizons, but it is clear that skill is lost rapidly, which is typical of such advection / extrapolation based systems. Future work should ideally compare the results of the variational Bayes methods with other approaches. This would be greatly facilitated by a suit of standard test cases and diagnostics that could be agreed up by the nowcasting community to allow model and method comparison.

There are several areas in which the algorithm could be improved. Parallelisation would allow the algorithm to be optimised for more cycles resulting in an optimal KL-divergence based fit, and the number of cells used in the approximation to the precipitation field could also be increased. The inclusion of an automatic method to select the complexity of the model, using methods similar to those used in the relevance vector machine [53] would further improve the robustness and scaleability of the algorithm. The advection process representation is also rather crude and could be improved with a better representation of the advection field based on a fixed grid and non-linear forecasting methods. It must also be stated that the model, or propagation, error parameters have been set with reference to other studies and tuning on short data sequences and these are almost certainly not optimised. It would be possible to use the marginal
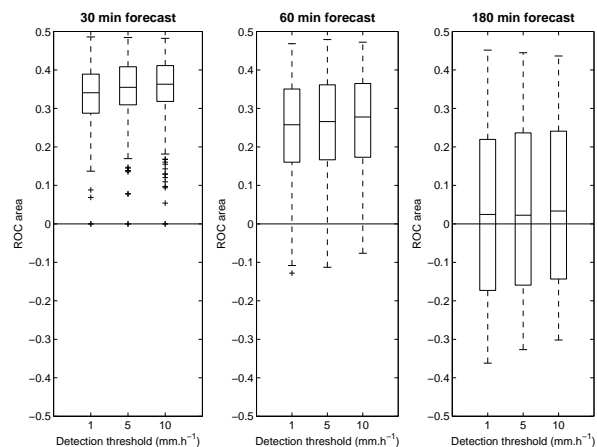
likelihood approximation, derived from the variational Bayes analysis to optimise these parameters as part of the data assimilation method, indeed these could be made adaptive since it is likely that the model error will be state dependent in this application.

More speculatively it would be interesting to attempt to include satellite imagery to track the evolution of the cloud field to better estimate the advection field where precipitation has yet to begin, but where clouds are present. Assimilation of other observation types, including doppler lidar and other more direct measurements of advection would further improve the estimation in the model. Further work could consider a hybrid approach that combines the knowledge of the physical system embodied in high resolution numerical models [15] but is sufficiently simple to be run on the short assimilation cycles required for short range forecasts. Since the model formulation is probabilistic and the uncertainty represents the model uncertainty reasonably well, Bayesian model averaging could be applied to merge smoothly into a more physics based forecast at longer lead times, so long as the uncertainty on both were well characterised. It is the belief of the authors that more effort should be placed on accurately quantifying uncertainty than is currently the case in many forecasting activities. This paper provides a new mechanism for attempting to achieve this in the nowcasting context.
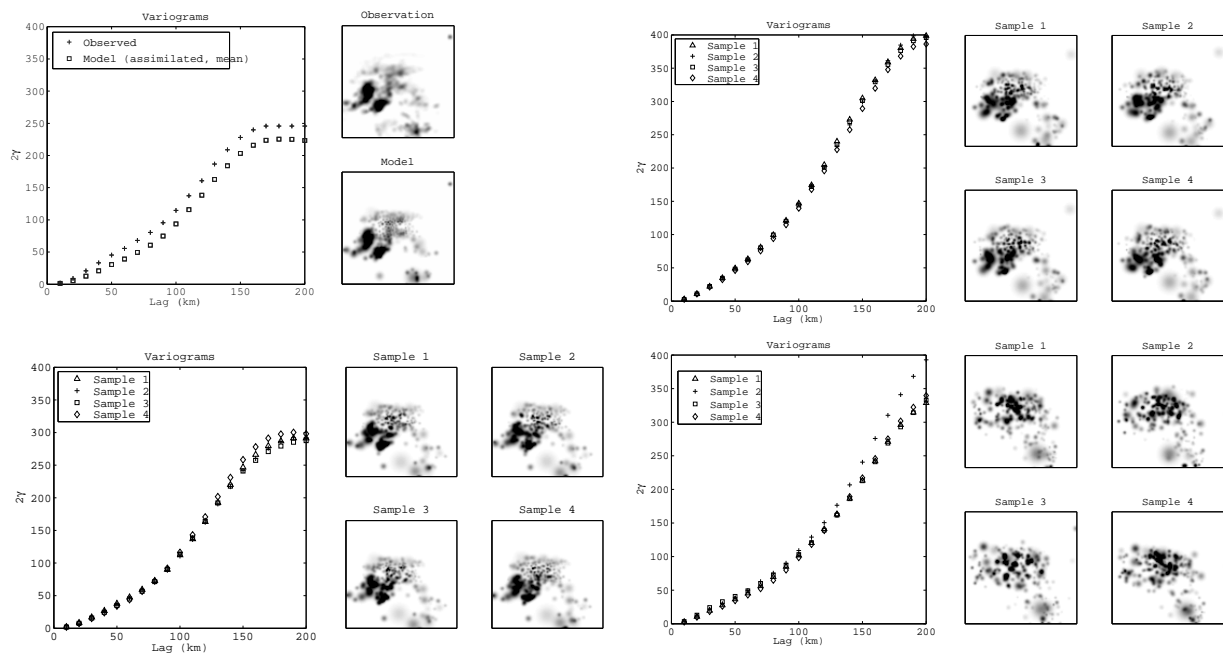
10. Acknowledgements

19

Fig. 19. Variograms of the observed and forecast precipitation fields (July 2006, t=70h). Variograms and rainfall fields correspond to observed and assimilated (top left), 30 min (bottom left), 60 min (top right) and 180 min (bottom right) forecasts. Four forecast realisations are plotted.

References

[1] P. Alberoni, V. Ducrocq, G. Gregoric, G. Haase, I. Holleman, M. Lindskog, B. Macpherson, M. Nuret, A. Rossa, Quality and assimilation of radar data for NWP - a review, Tech. rep., COST-717 (2003).

[2] F. Atger, Verification of intense precipitation forecasts from single models and ensemble prediction systems, Nonlinear Processes in Geophysics 8 (2001) 401–417.

[3] C. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1996.

[4] D. Bocchiola, R. Rosso, The use of scale recursive estimation for short term quantitative precipitation forecast, Physics and Chemistry of the Earth 31 (18) (2006) 1228–1239.

[5] N. Bowler, C. Pierce, A. Seed, STEPS: a probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP, Quarterly Journal of the Royal Meteorological Society 132 (2006) 2127–2155.

[6] K. Browning, A. Blyth, P. Clark, U. Corsmeier, C. Morcrette, J. Agnew, S. Ballard, D. Bamber, C. Barthlott, L. Bennett, et al., The Convective Storm Initiation Project, Bulletin of the American Meteorological Society 88 (12) (2007) 1939–1955.

[7] A. Burton, P. O'Connell, Model based algorithms for short lead time nowcasting, Tech. Rep. 5.3, MUSIC project report (2003).

[8] V. Collinge, The development of weather radar in the UK, Wiley, 1987.

[9] D. Cornford, A Bayesian state space modelling approach to probabilistic quantitative prediction forecasting, Journal of Hydrology 288 (1-2) (2004) 92–104.

[10] P. Cowpertwait, P. O'Connell, A. Metcalfe, J. Mawdsley, Stochastic point process modelling of rainfall. I. Single-site fitting and validation, Journal of Hydrology 175 (1-4) (1996) 17–46.

[11] D. Cox, V. Isham, A simple spatial-temporal model of rainfall, Proceedings of the Royal Society of London A415 (1988) 317–328.

[12] N. Cressie, Statistics for Spatial Data, Wiley and Sons, 1993.

[13] R. Deidda, Rainfall downscaling in a space-time multifractal framework, Water Resources Research 36 (7) (2000) 1779–1794.

[14] M. Dixon, G. Wiener, Titan: Thunderstorm identification, tracking, analysis, and nowcasting - a radar-based methodology, Journal of Atmospheric and Oceanic Technology 10 (6) (1993) 785–797.

[15] J. M. Done, P. A. Clark, G. C. Craig, M. E. B. Gray, S. L. Gray, Mesoscale simulations of organised convection: Importance of convective-equilibrium, Quarterly Journal of the Royal Meteorological Society 132 (2006) 737–756.

[16] A. Favre, A. Musy, S. Morgenthaler, Unbiased parameter estimation of the Neyman–Scott model for rainfall simulation with related confidence interval, Journal of Hydrology 286 (1-4) (2004) 168–178.

[17] U. Germann, I. Zawadzki, Scale-Dependence of the Predictability of Precipitation from Continental Radar Images. Part I: Description of the Methodology, Monthly Weather Review 130 (12) (2002) 2859–2873.

[18] U. Germann, I. Zawadzki, Scale Dependence of the Predictability of Precipitation from Continental Radar Images. Part II: Probability Forecasts, Journal of Applied Meteorology 43 (1) (2004) 74–89.

[19] M. Gibson, Report on further European radar data for use in Nimrod, Tech. rep., Forecasting Research Technical Report (2001).

[20] B. Golding, Nimrod: A system for generating automated very short range forecasts, Meteorological Applications 5 (1) (1998) 1–16.

[21] B. Golding, Quantitative precipitation forecasting in the UK, Journal of Hydrology 239 (2000) 286–305.

[22] C. Golz, T. Einfalt, M. Gabella, U. Germann, Quality control algorithms for rainfall measurements, Atmospheric Research 77 (1-4) (2005) 247–255.

[23] M. Grecu, W. Krajewski, Rainfall forecasting using variational assimilation of radar data in numerical cloud models, Advances in Water Ressources 24 (2001) 213–224.

[24] M. Grecu, W. F. Krajewski, A large-sample investigation of statistical procedures for radar-based short-term quantitative precipitation forecasting, Journal of hydrology 239 (2000) 69–84.

[25] V. Gupta, E. Waymire, Multiscaling properties of spatial rainfall and river flow distributions, Journal of Geophysical Research 95 (D3) (1990) 1999–2009.

[26] W. Hand, B. Conway, An Object-Oriented Approach to Nowcasting Showers, Weather and Forecasting 10 (2) (1995) 327–341.

[27] G. Hinton, D. van Camp, Keeping neural networks simple by minimizing the description length of the weights, Proceedings of the Sixth Annual Conference on Computational Learning Theory (1993) 5–13.

[28] R. Hogan, A Variational Scheme for Retrieving Rainfall Rate and Hail Reflectivity Fraction from Polarization Radar, Journal of Applied Meteorology and Climatology 46 (10) (2007) 1544–1564.

[29] D. Koutsoyiannis, C. Onof, Rainfall disaggregation using adjusting procedures on a Poisson cluster model, Journal of Hydrology 246 (1-4) (2001) 109–122.

[30] H. Lean, P. Clark, M. Dixon, N. Roberts, A. Fitch, R. Forbes, C. Halliwell, Characteristics of high resolution versions of the Met. Office Unified Model for forecasting convection over the UK, Monthly Weather Review In Press.

[31] L. LeCam, A stochastic description of precipitation, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 3 (1961) 165–186.

[32] L. Li, W. Schmid, J. Joss, Nowcasting of Motion and Growth of Precipitation with Radar over a Complex Orography, Journal of Applied Meteorology 34 (6) (1995) 1286–1300.

[33] C. Lin, S. Vasic, I. Zawadzki, B. Turner, Precipitation forecast based on Numerical Weather Prediction models and radar nowcasts, Proceedings of ERAD 201 (205).

[34] S. Lovejoy, B. Mandelbrot, Fractal Properties of Rain, and a Fractal Model, Tellus 37.

[35] S. Lovejoy, D. Schertzer, Multifractals and rain, in: New Uncertainty Concepts in Hydrology and Hydrological Modeling, Cambridge Press, 1995.

[36] D. MacKay, Information Theory, Inference, and Learning Algorithms, 2003.

[37] C. Marzban, Variograms as a verification tool, submitted to Mon. Wea. Rev. (2007).

[38] R. Mathurin, B. Rottembourg, A combinatorial approach for rain cell tracking, in: Twelfth International Conference and Workshops on Applied Geologic Remote Sensing, Denver, USA, 1997.

[39] P. Meischner, Weather radar: principles and advanced applications, Berlin; New York: Springer, 2004.

[40] R. Moore, V. Bell, D. Jones, Forecasting for flood warning, Comptes Rendus Geoscience 337 (2005) 203–217.

[41] C. Mueller, T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, J. Yee, NCAR Auto-Nowcast System, Weather and Forecasting 18 (4) (2003) 545–561.

[42] I. Nabney, NETLAB: Algorithms for Pattern Recognition, Springer-Verlag, 2001.

[43] P. Northrop, A clustered spatial-temporal model of rainfall, Proceedings of the Royal Society of London A454 (1997) 1875–1888.

[44] C. Onof, H. Wheater, Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model, Journal of hydrology(Amsterdam) 149 (1-4) (1993) 67–95.

[45] C. Pierce, E. Ebert, A. Seed, M. Sleigh, C. Collier, N. Fox, N. Donaldson, J. Wilson, R. Roberts, K. Mueller, The nowcasting of precipitation during Sydney 2000: an appraisal of the QPF algorithms, Weather And Forecasting 19 (2004) 7–21.

[46] C. Pierce, P. Hardaker, C. Collier, C. Hagget, Gandolf: a system for generating automated nowcasts of convective precipitation, Meteorological Applications 7 (4) (2000) 341–360.

[47] R. Rinehart, E. Garvey, Three-dimensional storm motion detection by conventional weather radar, Nature 273 (5660) (1978) 287–289.

[48] I. Rodriguez-Iturbe, D. Cox, V. Isham, Some Models for Rainfall Based on Stochastic Point Processes, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences (1934-1990) 410 (1839) (1987) 269–288.

[49] A. Seed, A Dynamic and Spatial Scaling Approach to Advection Forecasting, Journal of Applied Meteorology 42 (3) (2003) 381–388.

[50] J. Smith, W. Krajewski, Statistical modeling of space-time rainfall using radar and rain gage observations, Water Resources Research 23 (10) (1987) 1893–1900.

[51] J. Smithers, G. Pegram, R. Schulze, Design rainfall estimation in South Africa using Bartlett–Lewis rectangular pulse rainfall models, Journal of Hydrology 258 (1-4) (2002) 83–99.

[52] Y. Tessier, S. Lovejoy, D. Schertzer, Universal Multifractals: Theory and Observations for Rain and Clouds, Journal of Applied Meteorology 32 (2) (1993) 223–250.

[53] M. Tipping, Sparse Bayesian learning and the relevance vector machine, Journal of Machine Learning Research 1 (2001) 211–244.

[54] UK Meteorological Office, British Atmospheric Data Centre, Rain radar Products (NIMROD), Internet (2003).
URL http://badc.nerc.ac.uk/data/nimrod

[55] D. Veneziano, P. Furcolo, V. Iacobellis, Imperfect scaling of time and space–time rainfall, Journal of Hydrology 322 (1-4) (2006) 105–119.

[56] B. Vieux, J. Vieux, Statistical evaluation of a radar rainfall system for sewer system management, Atmospheric Research 77 (1-4) (2005) 322–336.

[57] H. S. Wheater, V. S. Isham, D. R. Cox, R. E. Chandler, A. Kakou, P. K. Northrop, L. Oh, C. Onof, I. Rodriguez-Iturbe, Spatial temporal rainfall fields: modelling and statistical aspects, Hydrology and Earth System Sciences 4 (4) (2000) 581–601.

[58] C. Wikle, A kernel-based spectral model for non-gaussian spatio temporal processes, Statistical Modelling 2 (4) (2002) 299–314.

[59] J. Wilson, A. Crook, C. Mueller, J. Sun, M. Dixon, Nowcasting thunderstorms: A status report, Bulletin of the American

Meteorological Society 79 (10) (1998) 2079–2099.

[60] K. Xu, C. Wikle, N. Fox, A Kernel-Based Spatio-Temporal Dynamical Model for Nowcasting Weather Radar Reflectivities, Journal of the American Statistical Association 100 (472) (2005) 1133–1144.