Journal of Hydrology 548 (2017) 391–405

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Representing radar rainfall uncertainty with ensembles based on a time-variant geostatistical error modelling approach

Francesca Cecinati [a,*], Miguel Angel Rico-Ramirez [a], Gerard B.M. Heuvelink [b], Dawei Han [a]

[a] University of Bristol, Department of Civil Engineering, BS8 1TR Bristol, UK
[b] Wageningen University, Soil Geography and Landscape Group, PO Box 47, 6700 AA Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 19 January 2016
Received in revised form 7 February 2017
Accepted 26 February 2017
Available online 2 March 2017

Keywords:
Radar QPE error model
Time-variant variograms
Radar ensemble
Conditional simulations
Rainfall uncertainty propagation

ABSTRACT

The application of radar quantitative precipitation estimation (QPE) to hydrology and water quality models can be preferred to interpolated rainfall point measurements because of the wide coverage that radars can provide, together with a good spatio-temporal resolutions. Nonetheless, it is often limited by the proneness of radar QPE to a multitude of errors. Although radar errors have been widely studied and techniques have been developed to correct most of them, residual errors are still intrinsic in radar QPE. An estimation of uncertainty of radar QPE and an assessment of uncertainty propagation in modelling applications is important to quantify the relative importance of the uncertainty associated to radar rainfall input in the overall modelling uncertainty. A suitable tool for this purpose is the generation of radar rainfall ensembles. An ensemble is the representation of the rainfall field and its uncertainty through a collection of possible alternative rainfall fields, produced according to the observed errors, their spatial characteristics, and their probability distribution. The errors are derived from a comparison between radar QPE and ground point measurements. The novelty of the proposed ensemble generator is that it is based on a geostatistical approach that assures a fast and robust generation of synthetic error fields, based on the time-variant characteristics of errors. The method is developed to meet the requirement of operational applications to large datasets. The method is applied to a case study in Northern England, using the UK Met Office NIMROD radar composites at 1 km resolution and at 1 h accumulation on an area of 180 km by 180 km.

The errors are estimated using a network of 199 tipping bucket rain gauges from the Environment Agency. 183 of the rain gauges are used for the error modelling, while 16 are kept apart for validation. The validation is done by comparing the radar rainfall ensemble with the values recorded by the validation rain gauges. The validated ensemble is then tested on a hydrological case study, to show the advantage of probabilistic rainfall for uncertainty propagation. The ensemble spread only partially captures the mismatch between the modelled and the observed flow. The residual uncertainty can be attributed to other sources of uncertainty, in particular to model structural uncertainty, parameter identification uncertainty, uncertainty in other inputs, and uncertainty in the observed flow.

1. Introduction

Many hydrological, water quality, and integrated catchment models use rainfall information as primary input. In several applications, weather radars are a precious source of rainfall data, thanks to their distributed nature, the wide coverage, and the high spatial and temporal resolution. Nevertheless, there are several factors that could introduce errors. First of all, radar quantitative precipitation estimation (QPE) relies on a conversion between the measured reflectivity Z in mm^6/m^3 and the physical quantity, the rainfall rate R in mm/h. The relationship is dependent on the rainfall nature, in particular on the drop size distribution (DSD) (Doviak, 1983; Marshall et al., 1947). The adopted Z-R relationships are often calibrated against spatial and temporal average conditions of liquid precipitation, but cannot be tailored to each specific situation and usually fail to correctly estimate extremes or the presence of hail or snow (Austin, 1987; Hasan et al., 2014; Seed et al., 2007). Polarimetric radars can improve the retrieval of the physical quantity R using other polarimetric parameters (Bringi et al., 2011), but often the radar networks are not updated to operationally use polarimetric radars. Other sources of uncertainty are due to the radar beam propagation that can be partially or totally blocked by obstacles (Friedrich et al., 2007; Joss and Lee, 1995;

* Corresponding author.
E-mail address: francesca.cecinati@bristol.ac.uk (F. Cecinati).

http://dx.doi.org/10.1016/j.jhydrol.2017.02.053
0022-1694/

Westrick et al., 1999), can be deviated by anomalous atmospheric conditions (Moszkowicz et al., 1994; Rico-Ramirez and Cluckie, 2008; Steiner and Smith, 2002), can be attenuated due to heavy precipitation (Atlas and Banks, 1951; Delrieu et al., 2000; Meneghini, 1978; Uijlenhoet and Berne, 2008), and may be subject to beam broadening with range, beam overshooting precipitation, and earth curvature effects, that increase the radar beam height and reduce the resolution at longer ranges (Ge et al., 2010; Kitchen and Jackson, 1993). Ground clutter is another source of error, producing disturbing echoes (Hubbert et al., 2009a,b; Islam et al., 2012). The rainfall rate estimates are often subject to variability of the vertical reflectivity profile (VRP) and to phenomena like the bright band effects, due to the higher reflectivity of the layer in which snow melts into rain (Austin and Bernis, 1950; Fabry and Zawadzki, 1995; Kirstetter et al., 2013; Qi et al., 2013; Rico-Ramirez and Cluckie, 2007; Smith, 1986; Zhang and Qi, 2010). Errors are also introduced by the spatial and temporal sampling, in the projection from polar to Cartesian coordinates, and in the averaging operations necessary to obtain the final corrected products (Anagnostou and Krajewski, 1999; Fabry et al., 1994). The list of error sources is long and for an extensive review, the reader is redirected to Villarini and Krajewski (2010) and McKee and Binns (2015). Although many techniques exist to partially correct different types of errors, a residual uncertainty inevitably affects radar QPE. In processed radar products the residual uncertainty is due to a mixed combination of the residual uncorrected errors and the processing errors and approximations. When radar QPE is used for hydrological applications, the estimation of its uncertainty and the assessment of uncertainty propagation in hydrological models is essential (Berne and Krajewski, 2013; Pappenberger and Beven, 2006; Schröter et al., 2011). An effective method to model uncertainty in radar QPE for hydrological model applications is the use of radar ensembles, which can easily be applied to hydrological models to assess residual error propagation in the model output (AghaKouchak et al., 2010a,b; Germann et al., 2009; Villarini et al., 2009). This approach is based on estimating the residual errors in radar QPE as a comparison with reference ground measurements, like those provided by rain gauges, used as an approximation of true rainfall. The observed radar QPE residual errors are then used to build an error model describing the statistical characteristics of the errors; knowing the statistical characterisation of the radar QPE residual errors, a large number of alternative possible realisations of the observed rainfall fields, constituting an ensemble, are synthesised. The uncertainty propagation through models can be estimated by observing the resulting spread after feeding a model with multiple ensemble members.

Several methods for radar ensemble generation are proposed in the literature, of which many are based on the computation of the error covariance matrix (Dai et al., 2014; Germann et al., 2009; Kirstetter et al., 2015; Villarini et al., 2009, 2014). The covariance matrix approach is a powerful and well-tested method that uses the covariance matrix decomposition to condition uncorrelated random normal deviates, in order to simulate alternative error components for the ensemble. A well-formulated example is the REAL generator proposed by Germann et al. (2009). However, it has some limitations when the number of rain gauges is large, because the covariance matrix calculation becomes computationally demanding and the decomposition unstable. In addition, ensemble error components are generated only at ground measurement points, needing subsequent interpolation that alters the spatial structure and introduces significant smoothing problems. Finally, in the calculation of the covariance matrix the spatial non-stationarity of the errors is captured assuming temporal stationarity. In other words, although the covariance approach reproduces the covariances between the errors at each rain gauge location, it assumes temporal stationarity of errors. Radar errors are non-stationary both in space and in time, but with a limited number of observations it is necessary to consider one of the two dimensions stationary in order to have enough observation points to calculate statistics. This paper explores the possibility to model radar errors that are non-stationary in time and stationary in space. The variability in space observed at ground measurement points is partially reproduced using conditional simulations for the error component generation.

This work proposes an ensemble generation approach aiming at reducing the computational load, improving stability, eliminating the need for error component interpolation, and producing time-variant residual error characterisation. This approach allows us to better capture time-dependent characteristics of residual errors, due for example to temporary conditions like the presence of bright band, hail or attenuation. The spatial characterisation of the residual errors is based on the use of variograms fitted with parametric models, which have the advantage of using only a limited number of variogram parameters (i.e. range, sill, and nugget), for full description of the residual errors and of being estimated with shorter time series. In comparison with the covariance matrix approach, the variogram approach constitutes a compromise, by exchanging temporal stationarity of the residual errors with spatial stationarity. In fact, although this method is able to reproduce the variability in error statistics over time, it considers errors stationary in space in the study area. The generation of alternative error components for the ensemble members is accomplished with conditional simulations following the methodology by Delhomme (1979). Error measurements are obtained using quality checked rain gauge data as an approximation of true rainfall. In addition, the problem of mean and variance inflation due to the adoption of a Gaussian error model in the logarithmic domain is addressed and a linear correction is introduced. As a case study, a large area of 180 km by 180 km in the north of England is used. The ensembles generated with the proposed method are validated on an independent set of rain gauges and tested on three different basins of different size using the Probability Distributed hydrological Model (PDM).

2. Datasets and case study

The case study presented in this work is a large portion of northern England, 180 km by 180 km wide. It presents a diversified orography, hillier in the north-west side, and flatter in the south-east, and includes some rural areas as well as some urban ones.

The radar ensemble generator is tested using hourly radar composites derived from the UK Met Office radar network, at 5-min and 1-km resolutions (Met Office, 2003). The product distributed by the UK Met Office has already been processed and corrected for (Harrison et al., 2000):

- Hardware or transmission corruption and noise
- Ground clutter and anomalous propagation
- Blockages
- Attenuation
- VRP adjustment and bright band correction
- Conversion from Z to R
- Conversion from polar to Cartesian projection
- Composition of different radar data

Furthermore, the data are adjusted against rain gauges, using an hourly mean-area adjustment-factor (Harrison et al., 2000). However, the original 5-min resolution radar data had some missing time periods, and therefore the missing scans were interpolated by advecting the rainfall radar field with a nowcasting model, and the 5-min radar data were accumulated at hourly time steps.

The study area is covered by three C-band radars (Hameldon Hill, Ingham, and High Moorsley) as shown in Fig. 1. It must be considered though, that the radar in High Moorsley was installed in 2008, therefore it did not contribute to the radar composites used for the case study, between 2007 and 2008.

The area is also covered by a network of 229 tipping bucket rain gauges from the Environment Agency (EA) with a 0.2 mm resolution. The rain gauge data are provided by the EA at 15-min time steps and then accumulated to one hour. A quality check is performed on the available stations and only 199 of these rain gauges are considered, excluding the ones presenting anomalies (e.g. duplication of time series, prolonged dry spells not agreeing with neighbouring rain gauges, missing data for considerable portions of the time series or for the entire time series, frequent inconsistency of measurements with neighbouring rain gauges and corresponding radar data). The 199 rain gauges are split in two datasets. A subset of 16 rain gauges are separated for validation, with a random selection conditioned to maintain a distance of at least 30 km between the gauges. The remaining 183 are used to model the radar errors. Both datasets are presented in Fig. 1.

The generated ensembles are tested using a hydrological model. Inside the 180 km by 180 km considered area, three hydrological basins of different sizes and shapes are selected: the upper part of the Ribble River (446 km^2), the upper part of the Lune (194 km^2) and the Rawthey (219 km^2). A Probability Distributed Model (PDM) (Moore, 2007) is set up to simulate the rainfall-runoff processes in each basin. The PDM is a lumped rainfall-runoff model, characterised by 14 parameters. As model input, basin area, temperature, and evapotranspiration are used, besides precipitation. The hourly temperature is taken from the UK Met Office Integrated Data Archive System (MIDAS) directly (Met Office, 2012). The evapotranspiration is calculated through the Penman-Monteith equation (Allen et al., 1998; Monteith, 1965) using hourly climatic variables (Total solar radiation, Net Solar Radiation, Wet Temperature, Dry Temperature, Wind Speed, Wind Direction, Rainfall, Pressure) from the MIDAS database (Met Office, 2012). Both temperature and evapotranspiration are averaged on the study area and considered constant in space. The flow data for the three basins, used for the comparison, are provided by the UK Centre for Ecology and Hydrology (CEH) through the National River Flow Archive (NRFA, http://nrfa.ceh.ac.uk/).

For the case study, a one-year radar rainfall ensemble is generated, from the 1st October 2007 at 00:00 to the 30th September 2008 at 23:00. The hydrological model calibration is performed on two years, from 1st October 2008 00:00 to 30th September 2010 23:00, in order to have the same seasonal reference; for the calibration, rain gauge data are used from the EA dataset. Fig. 1 shows the study area with the three catchments, the rain gauges, and the radar coverage area.

3. Methods

3.1. Error model

Radar residual errors can be modelled in different ways. Usually, it is recognised that radar residual errors have a bias component and a random component (Ciach et al., 2007). The random component is often modelled as multiplicative (Ciach et al., 2007; Dai et al., 2014; Villarini and Krajewski, 2009), but sometimes also an additive form is used (Kirstetter et al., 2010).

The model adopted in this work is additive in the logarithmic domain, thus it is multiplicative in the original domain:

$$10\log(P) = 10\log(R) + \delta \quad (1)$$

where P is the true rainfall, R is the radar QPE, δ is the residual error that is subsequently modelled to contain a bias correction as well, and the log operation refers to a logarithm with base 10. The model is consistent with previous research, in particular with the model adopted in the REAL method (Germann et al., 2009). The advantage of such a form is that the residual errors have an almost Gaussian probability distribution (Fig. 2), which is characterised only by the mean $\mu(\delta)$, the standard deviation $\sigma(\delta)$, and the spatial covariance. In the phase of error estimation, the true rainfall P is approximated with rain gauge measurements P_G, and the residual errors δ are defined as follows:

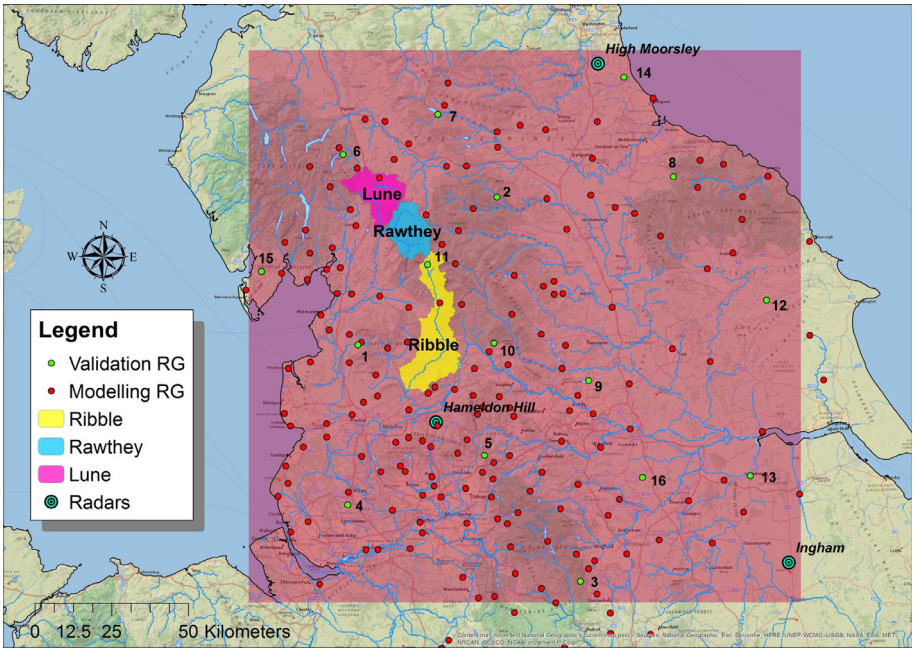

Fig. 1. The figure shows the study area, the radar grid extent, the radar positions, the rain gauges used for modelling, the rain gauges used for validation, and the three study catchments. The validation rain gauges are numbered accordingly to the results in Figs. 6 and 7.

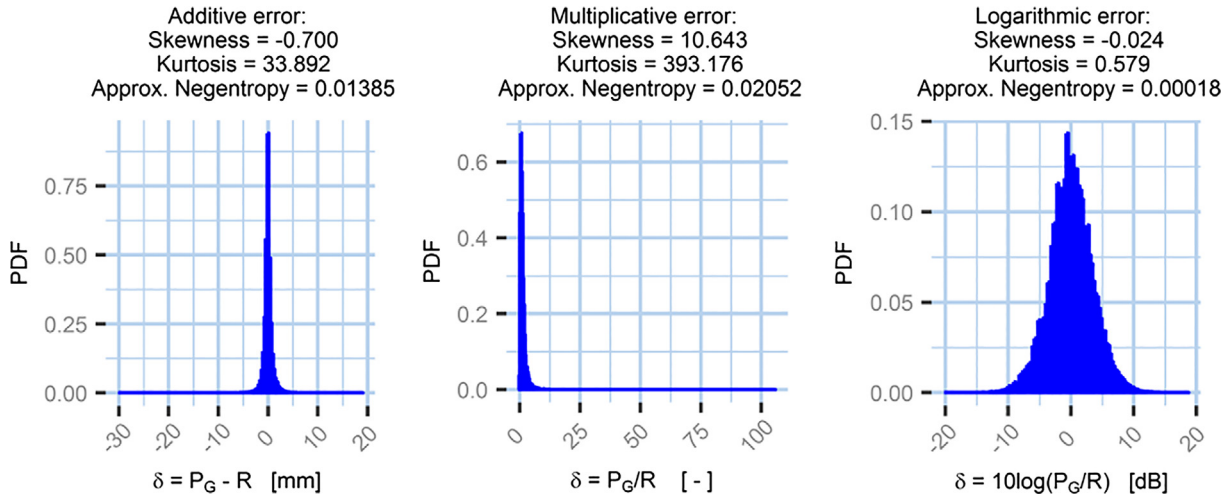

Fig. 2. Empirical probability distributions of radar errors using three different error models. Skewness, kurtosis, and approximation of negentropy are three indicators of a dataset Gaussianity. All of them tend to zero for a Gaussian distribution.

$$\delta = 10\log(P_G) - 10\log(R) \tag{2}$$

$$\mu(\delta) = E\{\delta\} \tag{3}$$

$$\sigma(\delta) = \sqrt{E\{(\delta - \mu(\delta))^2\}} \tag{4}$$

where $E\{\}$ is mathematical expectation, approximated with the mean error.

Fig. 2 also reports the values of Skewness, Kurtosis (Joanes and Gill, 1998), and approximation of negentropy (Hyvärinen and Oja, 2000), as indicators of Gaussianity. All of the indicators should tend to zero for a Gaussian distribution.

The residual errors must also be defined in terms of correlation characteristics. Although the temporal autocorrelation of errors may not be always negligible at hourly time steps (Kirstetter et al., 2010), previously published work showed that in the presented study area it is not significant (see Fig. 7 in Rico-Ramirez et al., 2015), therefore the attention in this work is focused on the spatial correlation structure. Often, the spatial correlation characteristics are depicted with a variance-covariance matrix C, describing the covariance between each pair $\delta(x_i)$ and $\delta(x_j)$ (Germann et al., 2009):

$$C(\delta(x_i), \delta(x_j)) = E\{(\delta(x_i) - \mu_i)(\delta(x_j) - \mu_j)\}$$
$$i, j = 1, \ldots, N \tag{5}$$

where the expected value is in practice calculated on time series. Parameters μ_i and μ_j, short notation for $\mu(\delta(x_i))$ and $\mu(\delta(x_j))$, are the mean of the residual error values $\delta(x_i)$ and $\delta(x_j)$, also calculated from the time series. The variance-covariance matrix may become unstable when the number of measuring points (N) is large. In fact, it must be positive-definite, which an empirical variance-covariance matrix might not be. Moreover, its inversion is computationally demanding for large N and may lead to numerical instabilities when the matrix is near-singular. It is also not suitable for time-variant calculation of error characteristics, because it calculates the expected values on time series, assuming stationarity of the characteristics in time (Le Ravalec et al., 2000). In reality radar errors are neither stationary in time nor space, because they are dependent on the rainfall rate and on temporary conditions like attenuation or bright band phenomena, variability of the Z-R relationship due, for example, to convective storms, drizzle, snow, or hail, and so on.

This work instead represents the spatial correlation characteristics of the residual errors through variograms. Variograms describe the variance as a function of the separation distance (d) (Cressie, 1993):

$$\gamma(d) = \frac{1}{2} E\{(\delta(x) - \delta(x+d))^2\} \tag{6}$$

An empirical variogram is calculated from the observations, binning the observation point distances in regular intervals (in the present case we use 1 km bins). It requires an assumption of spatial intrinsic stationarity of the field (Cressie, 1993). Empirical variograms are then fit with theoretical variogram functions. In the examined case, an exponential function is chosen, which describes the spatial characteristics of the residual errors through three parameters, the range parameter r, the sill s (absolute sill), and the nugget c_0. The exponential form has been selected because it fits well different variograms, empirically more flexibly than the Gaussian or the spherical shape that were tested as well for this case. The fitting is performed with a weighted least square method that uses a weight in the form N_o/d^2, where N_o is the number of available observations per distance bin and d is the distance (Cressie, 1985; Zhang and Eijkeren, 1995). The exponential model used is:

$$\hat{\gamma}(d) = c_0 + (s - c_0)\left(1 - \exp\left(-\frac{3d}{r}\right)\right) \tag{7}$$

Variograms have the advantage of being fast to calculate and easy to store. This allows for the calculation of time-variant residual error characteristics, i.e. at each one hour time step the errors are characterised by calculating a variogram for that time step. Ideally, the shorter the considered time, the better temporary phenomena influencing the error characteristics can be captured. At the same time, both radar and rain gauge time series are characterised by a large number of zeros or missing values, and often a sufficiently large number of observation points for variogram calculation is not available at one hour time steps. Therefore, for each time step, the errors observed at precedent time steps are considered too to calculate a pooled variogram. In this work, windows of 3, 6, 9, and 12 h are tested progressively to meet the stability conditions (defined later in this section) with the smallest possible time window; if the stability conditions are not met even using a 12-h window, average conditions are considered, and a pooled variogram is calculated using all available data, by pairing measurements coincident in time for multiple time steps. This means that the time window considered for variogram calculation varies

from time step to time step. The mean and the standard deviation are calculated in the same time window and are specific for each time step. In the upper part of Fig. 3 the variogram obtained from the average conditions (general variogram) is shown together with three other examples of variograms calculated at specific time steps.

In order to understand which dimension has the most variability, the coefficient of variation is calculated for both mean and standard deviation. As reported in Table 1, the absolute value of the coefficient of variation calculated over time is slightly higher for the mean, and clearly higher for the standard deviation. This means that assuming stationarity over space introduces a lower error than assuming stationarity over time.

Furthermore, the error components are generated in a conditional way, as will be explained in Section 3.2, so that the observed errors are reproduced and that all other simulated error points are conditioned on the observed ones. Although the mean and variance adjustment presented in Section 3.3 partially alters the reproduction of the observed errors, the geo-statistical approach still contributes to reproduce the spatial variability of errors.

Table 1
Coefficient of variation (unitless) for the mean and the variance, calculated over space and over time.

	CV over space	CV over time
$\mu(\delta)$	8.272	−8.689
$\sigma(\delta)$	0.090	0.336

The approach adopted here uses only one, omnidirectional variogram for each time step, assuming isotropy of residual errors. This hypothesis is not necessarily true and methods to include anisotropy in the variogram calculation exist and could be considered. However, considering the limited time window used for the residual error characteristics estimation in this time-variant application, it would be difficult to have enough measuring points to reliably estimate anisotropy as well.

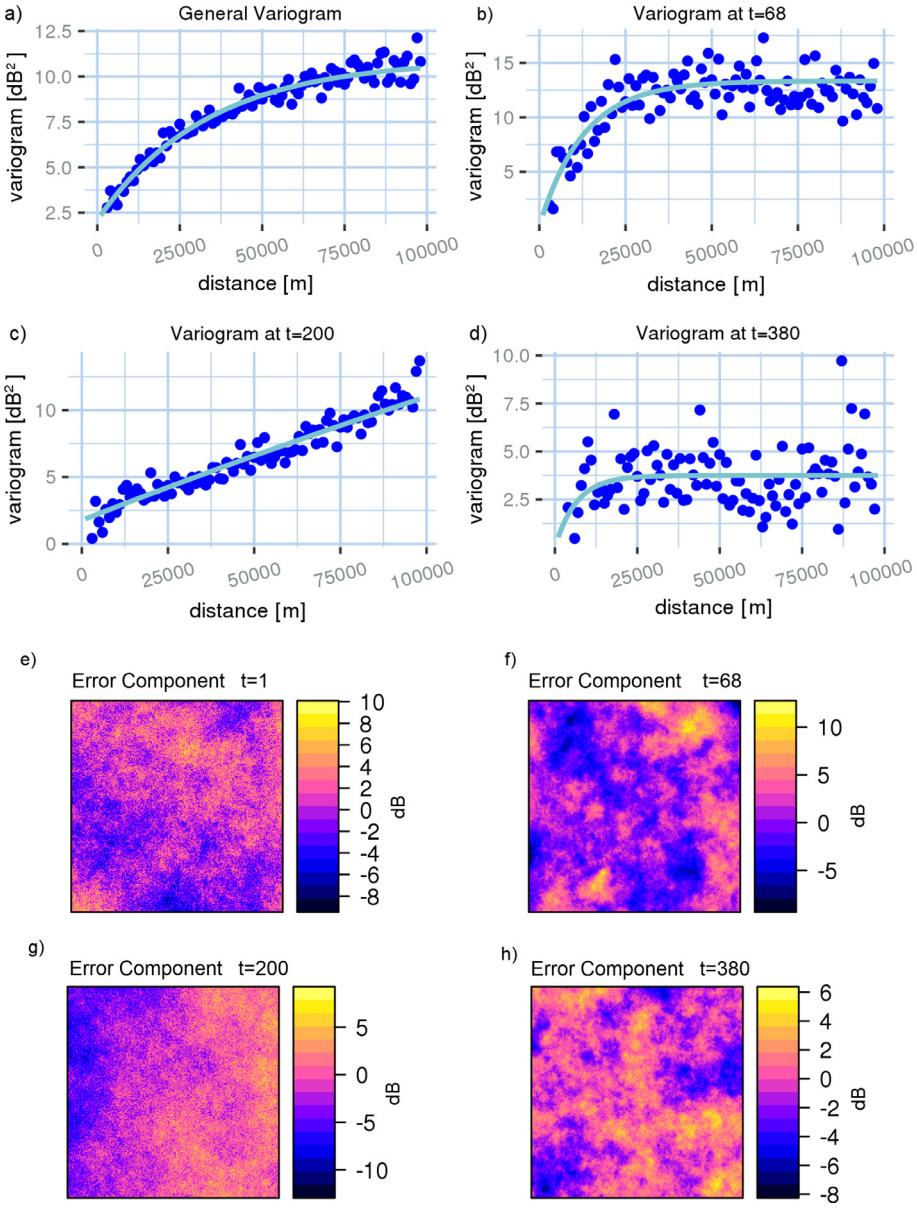

Fig. 3. The general average variogram (a) is compared with three example variograms observed at different time steps (b, c, and d). Examples of simulated error components in the log domain (e, f, g, and h) are produced using respectively the variograms a, b, c, and d.

As mentioned before, rainfall data are characterised by a large number of zeros or missing values that may impede the calculation of the variograms. When this occurs, backup values for mean, standard deviation and variogram sill, range, and nugget are used, calculated on all available data (in this case four years, from 2007 to 2010). In such a situation, the time-variant characteristics of the errors are not captured and the variogram represents the average conditions. However, this does not usually represent a problem, because it happens predominantly in conditions of no rain, very light and sparse rain, or at the beginning of a rainfall event, when not enough data are recorded yet. Although in these cases the error fields may not be accurately reproduced, the presence of zeros and the use of a multiplicative error model allow to produce realistic rainfall fields close to zero as measured. When significant rain events occur, there are soon enough observation points to calculate representative statistics and variograms. In order to establish when a variogram is properly fit, the following rules were set up:

1. There are at least 100 observation pairs to calculate it. Although a rule of thumb is to have at least 30 points per bin (Journel and Huijbregts, 1978), this is not realistic in the presented time-variant variogram fitting and would result in rejecting many variograms that could still provide useful information on the error characteristics. We are not aware of any previous study on the optimal number of observation pairs in a time-variant variogram calculation case, and therefore the number is selected as a compromise between the need for accepting as many variograms as possible, and the stability of the accepted variograms. Therefore, we found that at least 100 observations are sufficient to calculate the specific variogram, and also the mean and standard deviation of the observed errors.
2. The nugget is smaller than the sill (to avoid variograms with inconsistent physical explanation). Having a nugget larger than the sill rarely happens (in the case study around 3% of rejected variograms was rejected because of this rule). Nevertheless, when errors are affected by high levels of noise and not many points are available, the fit could result in a non-realistic variogram with the nugget larger than the sill.

If these conditions are not met in any of the tested time windows, the variogram parameters, the error mean $\mu(\delta)$, and the error standard deviation $\sigma(\delta)$ for the time window considered are substituted with the average ones.

Once the mean, standard deviation, and variogram parameters are established for each time step in which the radar ensembles have to be generated a given number of alternative error fields with the measured characteristics are then produced for each time step. The error components are then added to the original radar field (in the log domain), one by one, to generate the ensemble members.

3.2. Error component and ensemble generation

In order to generate error components with the desired mean, variance and variogram characteristics, conditional simulations are used. The method presented by Delhomme (1979) is selected, due to its calculation speed and its numerical stability, which make it suitable for unsupervised applications to long time series. The method is based on the following steps:

(a) For each time step t, an arbitrary number K of non-conditional simulations $\tilde{\delta}_{NC,i}(t,x)$ are generated, where $i = 1,\ldots,K$. The method used here is the sequential simulation implemented in the *gstat R* package (Pebesma, 2004).
(b) The observed errors at time t are interpolated with kriging, obtaining the interpolated fields $\delta_k(t,x)$.
(c) The values of the non-conditional simulations $\tilde{\delta}_{NC,i}(t,x)$ at observation locations are kriged to obtain the fields $\delta_{k_{NC,i}}(t,x)$.
(d) The conditional simulations $\tilde{\delta}_i(t,x)$ are obtained as follows:

$$\tilde{\delta}_i(t,x) = \delta_k(t,x) - \tilde{\delta}_{NC,i}(t,x) + \delta_{k_{NC,i}}(t,x) \quad (8)$$

Due to the logarithmic formulation, errors cannot be calculated when the rain gauges do not record rainfall. If no rain gauge records rainfall, unconditional simulations are used.

Important features of the generated fields are that they are Gaussian (in the logarithmic domain), are characterised using the observed variogram, and are conditioned on the observed errors. In addition, compared with the fields generated through the REAL method or other methods using the covariance matrix conditioning for spatial correlation modelling, the generated fields are already gridded fields and do not require any interpolation that tends to smooth the spatial features of the error components. In fact, the interpolation uses the kriging mean, i.e. the most probable value for each pixel. Instead, in the methodology applied here, at each pixel is assigned a different possible realisation for each ensemble member, in agreement with the conditional distribution. In this application $K = 100$ error components are generated for each time step, to produce 100 ensemble members. The number is a compromise between the statistical representativeness of the sample and the feasibility of producing them for one-year long hourly time series (Rico-Ramirez et al., 2015).

Following the error model, the i^{th} simulated error field $\tilde{\delta}_i(t,x)$ can be used to produce the i^{th} simulated QPE $\tilde{P}_i(t,x)$ for each time step t:

$$10\log(\tilde{P}_i(t,x)) = 10\log(R(t,x)) + \tilde{\delta}_i(t,x) \quad (9)$$

Since the logarithm of the radar field $R(t,x)$ cannot be calculated when a pixel is zero, pixels that do not record rainfall are not used and the zero values are re-introduced in the ensemble members in a second moment, after the adjustment presented in Section 3.3.

3.3. Variance and mean adjustment

The structure of the model is such that the new error members are Gaussian in the logarithmic domain, but the back-transformation to the final field $\tilde{P}$ gives a different weight to positive and negative deviations, shifting the overall mean toward higher values and increasing the variance. The bias introduced by a logarithmic back-transformation (Erdin et al., 2012) is not discussed by Germann et al. (2009), when the same error model is applied to the REAL ensemble generator. In addition, the new simulated error components are added to the radar field, which already contains errors, inflating the overall variance (Pegram et al., 2011). In order to have rainfall fields consistent with the observed approximation of true rainfall from the rain gauges, the mean and the variance need to be re-adjusted. In this work a linear adjustment is used at each time step t to re-adjust mean and variance of the generated ensemble members, without modifying the spatial characteristics so carefully reproduced:

$$\tilde{P}_{new,i} = \frac{\sigma_G}{\sigma_{\tilde{P}_{orig}}}(\tilde{P}_{orig,i} - m_{\tilde{P}_{orig}}) + m_G \quad (10)$$

where $\tilde{P}_{new,i}$ is the new i^{th} ensemble member after correction, $\tilde{P}_{orig,i}$ is the original i^{th} ensemble member, $\sigma_{\tilde{P}_{orig}}$ is the standard deviation of all original ensemble members across all the rain gauge measuring locations, $m_{\tilde{P}_{orig}}$ is the average of all the original ensem-

ble members at the rain gauge measuring locations, σ_G is the standard deviation of the rain gauge measurements, m_G is the mean of the rain gauge measurements.

It must be noted that the adjustment is not forcing each ensemble member to reproduce the mean of the rain gauge values. Instead, the adjustment forces the overall ensemble mean to tend to the true value, represented by the rain gauge measurements. This is justified by the definition of ensemble as a representation of the rainfall uncertainty due to the radar, therefore it should convey how much from the true value the radar data can deviate, where the true value is represented by the ensemble mean and the deviations by the single ensemble members. Similarly, the adjustment does not force the ensemble standard deviation at each point, but it corrects the spatial standard deviation of each ensemble member, in order to re-adjust the exponential stretch and avoid unrealistically high intensity values. The adopted solution is an approximation, but it is effective in obtaining possible realistic alternative rainfall fields.

3.4. Rain gauge validation

Validation is performed comparing the ensemble to the observed values by an independent network of rain gauges. Although the true rainfall is not known, rain gauges are used in this work as an approximation of it. Therefore a comparison between the set of 16 gauges kept out of the modelling dataset and the values of the ensembles in the rain gauge measuring locations can provide an assessment of the ensemble quality. The idea is to show that the ensemble encompasses the observations at observed points. However, the comparison should be done bearing in mind that the presented model includes some unavoidable approximations, that there is a significant difference in areal representativeness, being the ensembles at 1 km resolution and the rain gauges point measurements, and that rain gauge measurements contain uncertainty as well.

The rain gauges have a resolution of 0.2 mm. In order to make the comparison fair, ensemble member pixels below 0.2 mm are approximated to either 0.2 mm when the value is above 0.1 mm, or to 0 mm when the value is below 0.1 mm.

The validation is performed with three indicators:

1. The mean ensemble bias is calculated and compared to the mean radar bias.
2. Rank histograms (Hamill, 2001) are used to prove that the ensemble has a correct distribution.
3. A Goodness-Of-Fit estimator (GOF) is defined to check how often the ensemble captures the observed rainfall.

3.5. Hydrological application

The aim of expressing radar rainfall uncertainty through ensembles is to assess the uncertainty propagation in models. For this reason, the application of the generated ensemble to a hydrological model is an interesting test to assess the advantages in using an ensemble (Dai et al., 2015). Besides the rainfall uncertainty, the output of a hydrological model is affected by additional sources of uncertainty, like model structure, parameter uncertainty, other input uncertainty, lumping of parameters and inputs, or model numerical approximations. In addition, the measured flow also has an associated uncertainty. It is expected, therefore, that the radar rainfall uncertainty represented by the ensemble spread only partially explains the discrepancy between the modelled flow and the observed flow, showing the relative importance of radar rainfall uncertainty in the overall model uncertainty.

For this application, three basins have been selected, the upper Lune, the upper Ribble, and the Rawthey, as presented in Section 2 and Fig. 1, and for each a PDM model was set up. The PDM has been developed by the UK Centre for Ecology and Hydrology (CEH) and it is used operationally by the Environment Agency in the National Flood Forecasting System. It is a very flexible lumped rainfall-runoff model, easily applicable to different catchments, and does not require a detailed description of the catchment characteristics (Moore, 2007). The choice of a simple lumped model avoids the introduction of excessive additional uncertainty due to a large set of parameters and input data. Nevertheless, this approach requires the averaging of each ensemble member over the catchment areas. In order to observe the averaging effect on the target area, the three chosen basins are selected to have different size. PDM requires 14 parameters, of which one, the exponent for the groundwater storage momentum equation, can be assumed fixed according to Smith (1977). Therefore 13 parameters need calibration. The PDM offers three possibilities for the recharge function: standard, demand-based and based on a splitting coefficient. Here the standard solution is adopted (Moore, 2007). The three PDM models are calibrated with the Environment Agency rain gauge data, from 1st October 2008 00:00 to 30th September 2010 23:00 (two years). For the calibration optimisation, both a Monte Carlo (MC) approach (Metropolis and Ulam, 1949) with 10,000 samples from the parameter space, where all the parameters are given a uniform distribution in plausible intervals, and a shuffled complex evolution method, developed at The University of Arizona (SCE-UA) (Duan et al., 1992, 1994) were tested. The SCE-UA outperformed the MC, in terms of Nash Sutcliffe Efficiency (NSE) coefficient, in terms of result stability, and in terms of speed, therefore it was selected for calibration (also according to van Griensven and Meixner (2007)). In particular, although the NSE improves by only 0.01–0.03 [–] using the SCE-UA, the algorithm takes around one third of the time necessary to complete compared to the MC calibration and always converges on the same parameter estimate, while the MC returns a different estimate each time.

4. Results

4.1. Error components

Error components are generated with conditional simulations for each time step. Table 2 shows how many times the general variogram was used, as opposed to variograms calculated in time windows of 3, 6, 9 or 12 h, according to the methodology described in Section 3.1. The numbers show that more than half of the times, the variogram cannot be calculated. This figure is consistent with the fact that, according to the rain gauge measurements, 22.4% of the hours during the study period are completely dry and that in 54.9% of the times an average rainfall lower than 0.02 mm/h is recorded. Another important outcome is that the 3-h window is sufficient to calculate a variogram in around 65% of cases and that increasing the time window helps only in a limited number of cases.

Fig. 3 shows four different examples of error components (e, f, g, and h) generated using, respectively, the general variogram (a), i.e. the one calculated with all available time series, and variograms specific for three different time steps (b, c, and d). The use of specific variograms has an evident effect on the error components. As can be observed, the range, which is large in (c), medium in (a) and (b), and low in (c), controls the granularity of the spatial variation, while the nugget (present in a and c, but not in b and d) controls the speckle. The use of an exponential variogram function allows us to achieve a fit without supervision in most of the cases when sufficient observations are available. This model can sometimes derive range and sill parameters unrealistically high if the empirical variogram lays on a straight line like in (c), because the

Table 2
Percentage of adoption of the general conditions or specific time windows in the variogram, mean and standard deviation calculation.

	General variogram	3-h window	6-h window	9-h window	12-h window
Total	52.1%	31.0%	8.0%	5.0%	3.9%
Relative	–	64.8%	16.9%	10.3%	8.0%

physical range is out of the observation distances and the trend is extrapolated. Nevertheless, similar situations do not affect the error component generation because the variogram model still reproduces the observed characteristics in the observed and modelled domain.

4.2. Ensembles

Fig. 4 compares a radar rainfall field, a kriging interpolated rain gauge rainfall field, an ensemble member before mean and variance adjustment, and after adjustment. All maps refer to the same time step (01-12-2008 10:00). The rain gauge position and values are superimposed. The ensemble member before adjustment shows anomalously high values (e.g. in the top left corner in Fig. 4c). The re-adjustment corrects the bias of the whole ensemble, without forcing each member to have the rain gauge specific mean and variance.

A sample of resulting ensemble members, after adjustment, is presented in Fig. 5. As can be observed, the ensemble members are more speckled than the corresponding rainfall field. This is as expected, because ensemble members have the spatial variability observed through rain gauge – radar comparison. If the differences present a high spatial variability at short distances, corresponding to a short range and/or to a nugget effect in the corresponding variogram, this effect appears in the ensemble, but not in the radar image, that has a smoother behaviour. Although we do not have a sufficient instrumentation density to observe the short scale variability of rainfall, the granularity can be a sign of high spatial variability of errors, as observed by the available data. The average of all the ensemble members will tend to a bias correction of the radar. For this reason, also the intermittency (rain-no rain) features may be slightly different than the ones in the radar image, and remarkably better than the non-corrected version, where the exponential stretch results in larger areas of no rain.

4.3. Rain gauge validation

The validation method is based on the comparison between the ensemble and measurements from an independent set of 16 tipping bucket rain gauges that were kept out from the modelling phase. The rainfall rate values of the ensemble members are compared with the rain gauge measurements at rain gauge positions for the 16 gauges in the validation set. As an example, the comparison between the ensemble spread and the rainfall rate measured by the rain gauges is reported in Fig. 6 for an example event in September 2008.

The variability expressed by the ensemble captures a large part of the differences between the radar and rain gauge measurements. Rainfall intensity peaks in particular can have a large variability, as compared to deterministic predictions, due to the

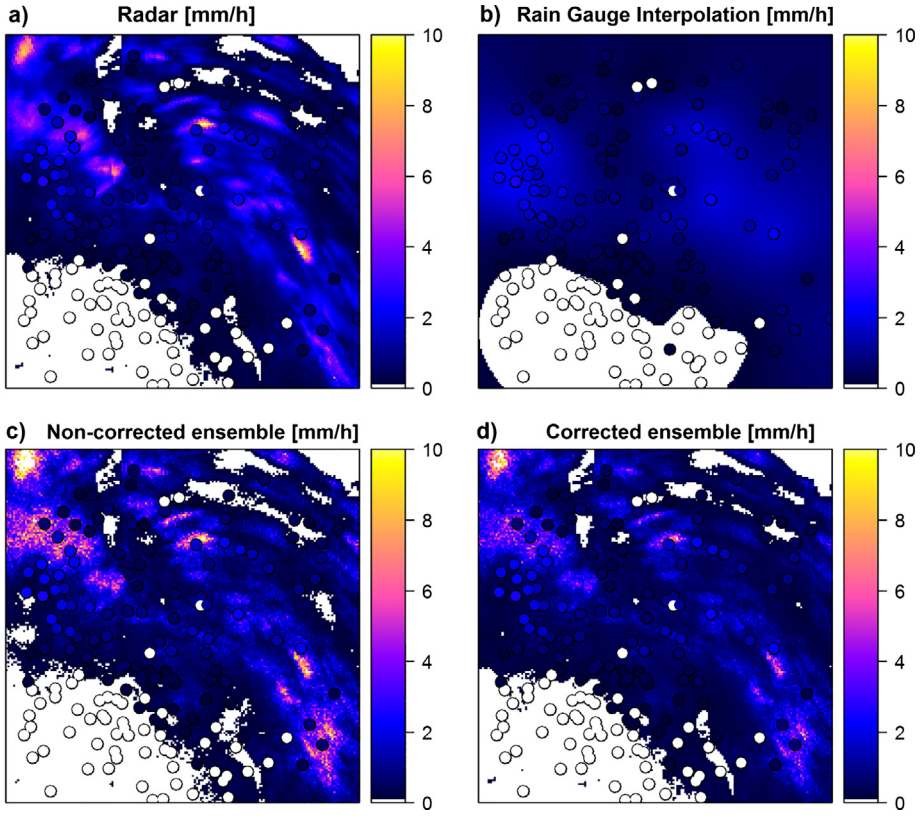

Fig. 4. Radar rainfall, rain gauge interpolation, and example ensemble members before and after correction at time 2008-01-01 10:00. Rain gauge measurements and positions are superimposed.

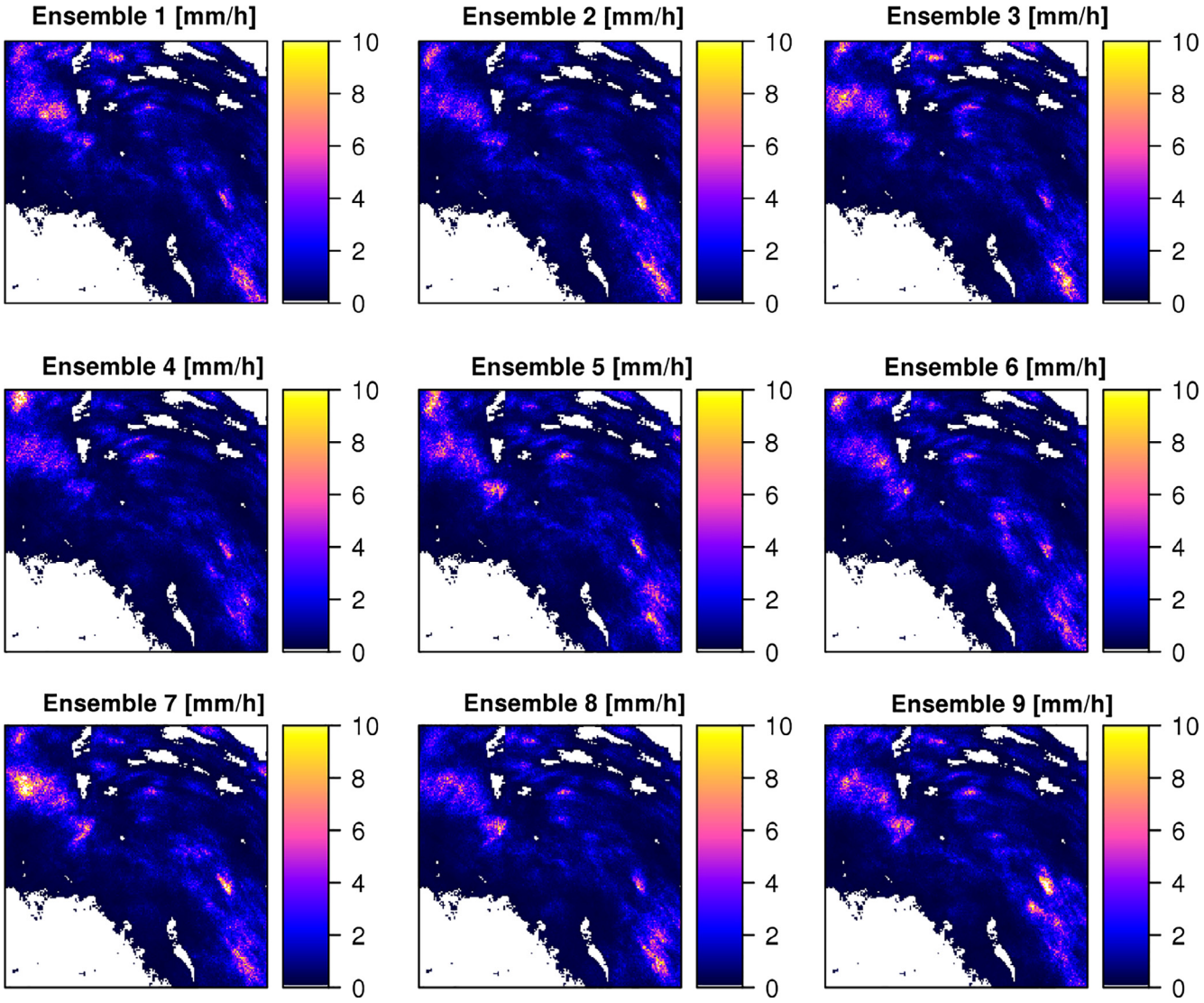

Fig. 5. Nine ensemble members from the same date and time of Fig. 4 (2008-01-01 10:00) are compared.

multiplicative nature of errors. For peaks, the ensemble spread is wider, because the errors are proportional to the rain rate. The ensemble is therefore able to predict the observed values where it would have been underestimated (e.g. rain gauge 2) or overestimated (e.g. rain gauge 12) by the deterministic values.

Looking at the discrepancies between the ensemble and the measured values, it must be considered that the ensemble only represents radar sources of uncertainty and does not include the uncertainty due to the rain gauge measurements or to the difference in areal representativeness. This is particularly evident where the discrepancies are often explainable by typical rain gauge errors: the tipping bucket rain gauges have a non-continuous behaviour, showing steps in the values or tipping delays (e.g. rain gauge 8), especially for low rainfall rates.

Besides external uncertainty factors, some mismatch is unavoidable with the presented method, because of the error characterisation and the geostatistical modelling, which are based on a limited number of observations, compared to the simulation points, and some unavoidable approximations and assumptions.

To obtain a quantitative measure of the match between the radar ensemble and the rain gauge measurements, we use three estimators:

(1) The mean bias of the ensemble mean is compared to the mean bias of the radar. The results are reported in Table 3.
(2) Rank histograms are produced for the 16 rain gauges. A rank histogram shows how often the observation falls in each of the ensemble quantiles.
(3) A goodness of fit estimator (*GOF*) is calculated as follow:

$$GOF = 1 - \frac{q_1 + q_{100}}{n_{tot}} \tag{11}$$

where n_{tot} is the total number of time steps, q_1 is the number of time steps in which the observation falls in the first percentile and q_{100} is the number of time steps in which the observation falls in the last percentile.

Table 3 reports the bias between the mean of the ensemble and the rain gauge measurements, compared to the bias between the radar rainfall estimates and the rain gauge values. The bias is sensibly reduced using the ensemble mean.

The rank histograms are reported in Fig. 7 for all the 16 validation rain gauges. Some rank histograms are mostly flat, which is a sign of correct ensemble distribution. In some others, the last bar of the histogram is higher, showing that the ensemble tends to underestimate the rainfall rate as recorded by the corresponding

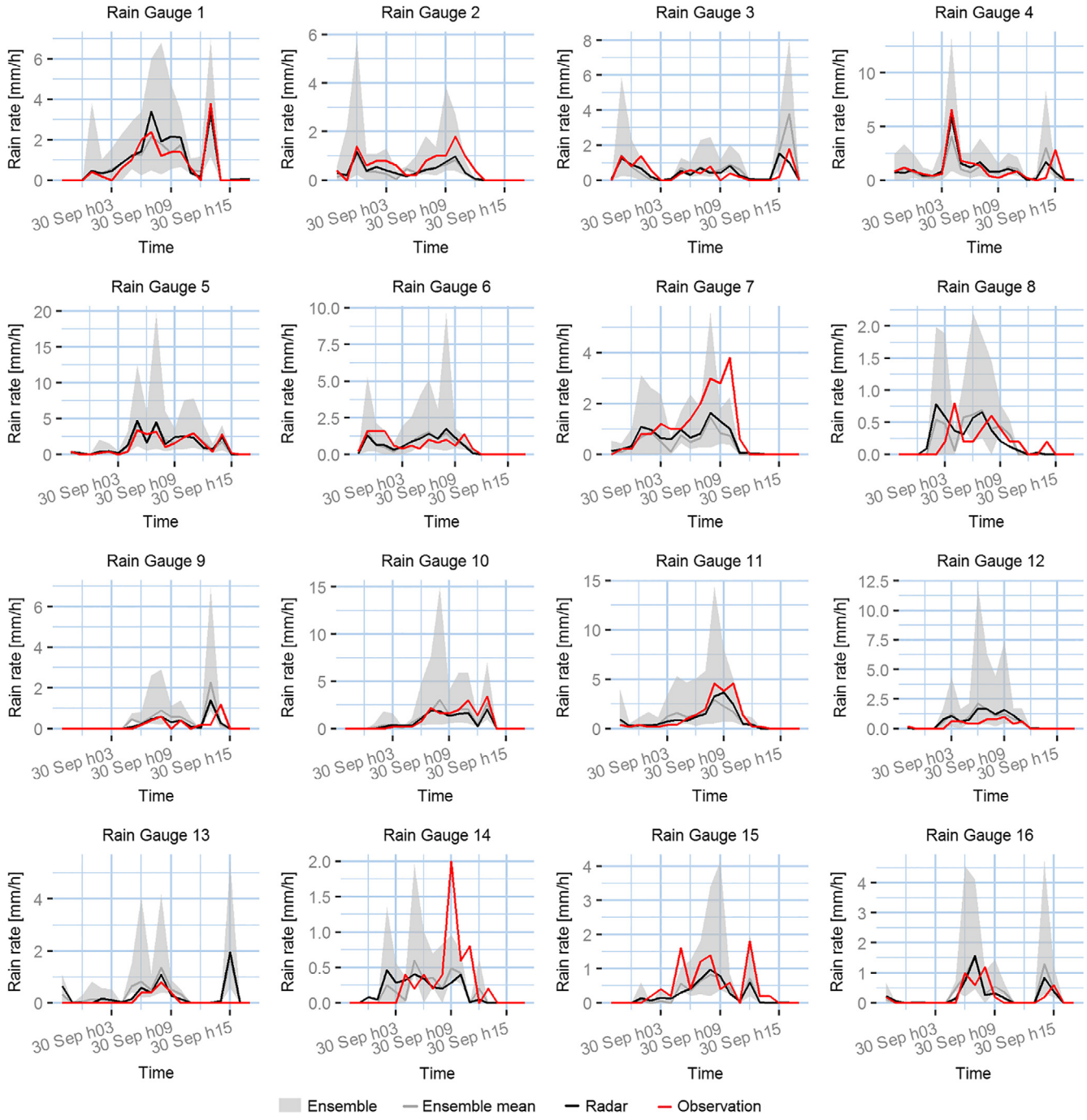

Fig. 6. The radar rainfall ensemble is compared to the rain gauge measurements, and to the radar measurements, for an example event in September 2008.

Table 3
Mean ensemble bias as opposed to mean radar bias for the 16 validation rain gauges [mm/h].

Rain gauge	Ensemble bias	Radar bias	Rain gauge	Ensemble bias	Radar bias
1	−0.022	−0.172	9	0.013	−0.102
2	−0.015	−0.151	10	0.070	−0.141
3	0.002	−0.107	11	0.017	−0.222
4	−0.015	−0.130	12	0.019	−0.102
5	0.010	−0.166	13	<0.001	−0.073
6	−0.101	−0.248	14	<0.001	−0.086
7	−0.005	−0.152	15	−0.071	−0.153
8	−0.026	−0.104	16	0.018	−0.080

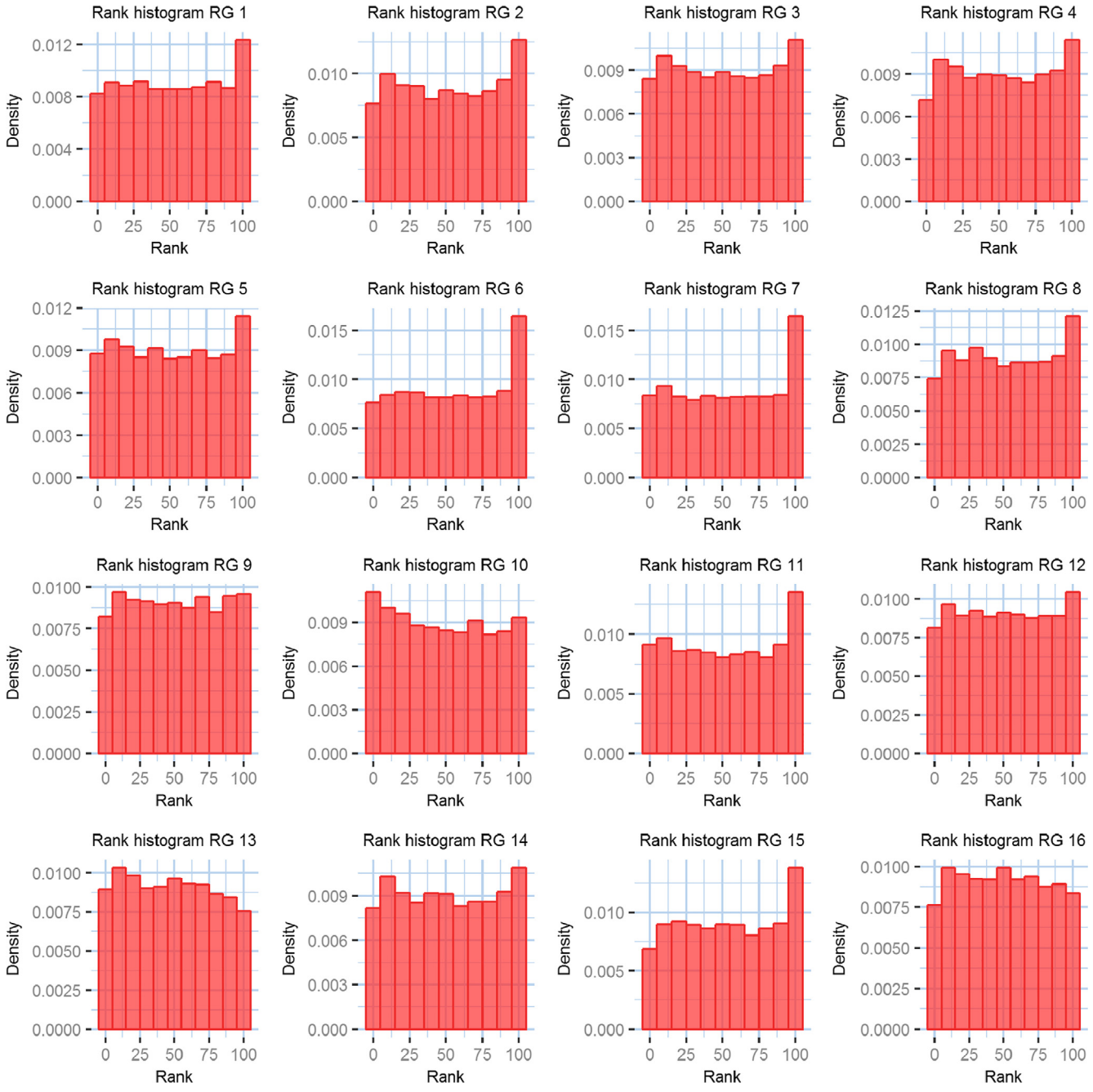

Fig. 7. Rank histograms are reported for the ensemble at validation rain gauge locations. The rank histograms show in which quantile of the ensemble the observation falls. A well balanced ensemble has a flat rank histogram.

rain gauges in some situations. Nevertheless, the value of the last bar is never doubled compared to the other bars, which means that, with eleven bins, one out of eleven measurement points is incorrectly underestimated. In most of the histograms, the value is even lower. This mismatch is in line with the mismatch observed in Fig. 6 and can be due to rain gauge errors, to differences in areal representativeness, or to model approximations.

Table 4 shows the goodness of fit estimator values for the 16 rain gauges, which represent how often the ensemble covers the observed rain rate.

Considering the mentioned residual sources of uncertainty, the agreement of the ensemble with the rain gauges is positive.

Table 4
Goodness of fit statistic for the 16 validation rain gauges, as explained in Eq. (11).

Rain gauge	GOF [–]	Rain gauge	GOF [–]
1	0.894	9	0.917
2	0.897	10	0.891
3	0.906	11	0.876
4	0.913	12	0.910
5	0.898	13	0.930
6	0.859	14	0.902
7	0.851	15	0.893
8	0.900	16	0.935

Table 5
NSE coefficients for the calibration of the three hydrological PDM models.

	Lune	Ribble	Rawthey
SCE-UA	0.92	0.89	0.92

4.4. Hydrological applications

The application of the generated radar rainfall ensembles to hydrologic models is used here to explain the advantage in using an ensemble as probabilistic input rather than a deterministic rainfall information. The radar ensembles are used with three PDM hydrological models calibrated for three basins in the study area, the upper Lune, the upper Ribble, and the Rawthey, as shown in Fig. 1. The generated ensemble covers the whole 180 km by 180 km area shown in Fig. 1 as a red square, and a time span of one year, from 1st October 2007 00:00 to 30th September 2008 23:00. An ensemble of 100 one-year time series on each basin is extracted averaging the distributed rainfall of each ensemble member $\tilde{P}_i(t,x)$ over the basin area. Each time series is used as a rainfall input for the PDM model. The model therefore returns an ensemble of predicted flows for each basin that are compared with the observed flow.

4.4.1. Calibration

The PDM models are calibrated with two years of rain gauge data (1st October 2008 00:00 to 30th September 2010 23:00). The calibration is done on 13 parameters. The optimisation is performed using the SCE-UA algorithm. The Nash Sutcliffe Efficiency (NSE) coefficients for the calibration are reported in Table 5. The obtained sets of parameters are reported in Table 6, with a brief description (Moore, 2007).

4.4.2. Results

The comparison between the ensemble and the observed flow for an example period in February 2008 is reported for each basin in Fig. 8.

To obtain a quantitative measure of the match between the output flow ensemble and the observed flows, the goodness of fit estimator as described in Eq. (11) is calculated for the flow as well and reported in Table 7.

It must be considered that in hydrological applications of radar rainfall, beyond the uncertainty in the radar measurements represented by the ensemble, there are numerous additional sources of uncertainty. Firstly, a model is an approximation of real processes and will never be able to perfectly reproduce reality, especially in the case of simple models like the PDM. In addition, the PDM requires other inputs, such as evapotranspiration, which have their

Table 6
Parameters obtained through the calibration of the three catchment PDM models with the SCE-UA.

		Lune	Ribble	Rawthey
f_c	Rainfall factor [ ]	1.08	0.94	1.08
c_{min}	Min store capacity [mm]	34.9	32.7	34.4
c_{max}	Max store capacity [mm]	104.7	50.0	50.0
b	Exponent of Pareto distribution for store capacity [ ]	1.37	1.35	0.83
b_e	Exponent in actual evaporation function [ ]	4.99	1.02	1.87
k_1	Time constant first linear reservoir [h]	2.47	2.37	2.09
k_2	Time constant second reservoir [h]	1.46	11.16	1.46
k_b	Baseflow time constant [h mm^{m-1}]	5.0	37. 9	5.0
k_g	Groundwater recharge time constant [h mm^{bg-1}]	89,400	194,800	237,100
S_t	Soil tension storage capacity [mm]	21.4	26. 7	17.1
b_g	Exponent of recharge function [ ]	3.04	4.16	3.71
q_c	Constant flow returns/abstraction [m^3 s^{-1}]	1.59	4.20	3.16
t_d	Time delay [h]	1.29	0.69	1.06

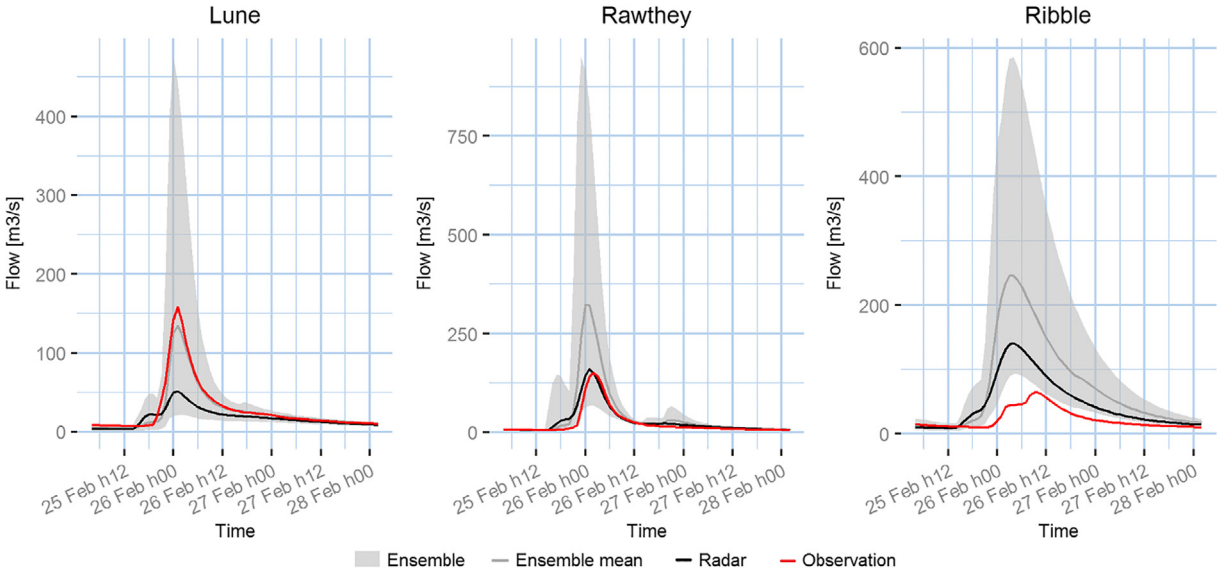

Fig. 8. The flow ensembles obtained using the rainfall ensemble as input for the PDM models of the three study catchments are compared to the actual measurement and to the radar prediction for the three study catchment during an example event in February 2008.

Table 7
Goodness of fit statistic for the flow in the three study catchments, as in Eq. (11).

	Ribble	Lune	Rawthey
GOF [–]	0.35	0.17	0.27

own associated uncertainties due to the data collection and to the averaging on the study area. The calibration results in high NSE scores, but still includes an uncertainty in the estimation of the necessary 13 parameters for the PDM, which may be variable over time. Finally, the observed flow used as reference has an associated uncertainty too.

Considering this, the quantitative goodness of fit reported in Table 7 is consequently considerably lower than the goodness of fit for the rain gauge validation.

The goodness of fit estimators are highest for the Ribble, which is the largest basin, and lowest for the Lune, which is the smallest.

The use of a radar rainfall ensemble can be used as a diagnostic tool. Since the ensemble proved to be well calibrated in the validation phase in Section 4.3, while the results of the flow GOF in Table 7 are poor, it is clear that the rainfall uncertainty is not the dominant source of uncertainty in the presented model. Despite the good calibration results, the uncertainty due to the other inputs, to the model structure, to the parameter identification, or to other errors or approximations are still determinant in the case study.

The ensemble can be used to identify in which cases the mismatch between the radar measurement and the observation is due to rainfall uncertainty, like in the panel of Fig. 8 relative to the Lune catchment, where the ensemble clearly improves the radar measurement and captures the observed flow, or when the mismatch is due to other modelling problems, like in the panel relative to Ribble, were the mismatch is not captured by the ensemble.

5. Considerations and conclusions

This work presents an innovative method to generate radar QPE ensembles. Compared to other methods based on the use of the residual error covariance matrix, the presented method introduces some advantages. Firstly, the use of variograms allows for a faster and more flexible calculation of spatial correlation of errors, at the point that a time-variant error characterisation is possible; the time-variant application allows us to capture temporary phenomena that may affect the nature of errors, generating ensembles that are specific for the simulated time step. The ability to calculate a time-variant variogram shows consistency with the characteristics of the precipitation in time: the necessity of using a backup general variogram occurs during dry periods or for very low rain rates, when only a limited number of rain gauges are recording rainfall. During significant rain events the applied method is able to calculate a specific variogram, using a 3-h window in almost 65% of cases and increasing the window to 6, 9, or 12 h only in the residual cases. The development of a fast and robust automatic method to characterise errors is done to allow possible operational and real-time application. For this reason the variogram modelling is done using only backward time windows, rather than centred ones.

The adoption of a logarithmic multiplicative error model allows for Gaussian modelling of the errors that makes the error characterisation and the alternative error field generation easier. As a drawback, it introduces some bias in the back-transformation which is here adjusted with a simple linear model. It could be interesting to derive alternative methods to obtain Gaussian errors with a stable back-transformation. Another significant advantage of the presented method is that using conditional simulations allows us to generate spatially-correlated Gaussian realisations of the random fields. Therefore, no interpolation of the simulated error components is needed. The use of interpolation has a smoothing effect due to the use of the kriging mean, rather than the full probabilistic kriging outcome. This can partially cover the problem of mean and variance inflation, but also modifies the spatial characteristics of the errors. The errors are generated with conditional simulations, where sufficient error observations were available, drawing correlated realisation for the conditional distribution of the whole grid. Conditioning the simulation to the observations allows us to partially reproduce the spatial variability of the error statistical properties that is neglected using an omnidirectional variogram.

In this application, the problem of mean and variance inflation due to the use of a logarithmic model is addressed with a linear readjustment that corrects the absolute values of the ensemble members without affecting their spatial characteristics. Using the overall mean and standard deviation of the ensemble for adjustment, the adjusting method forces the whole ensemble to have the observed statistical characteristics, but does not coerce the single ensemble members.

The generated ensemble represented in Figs. 4 and 5 show a correct rainfall distribution, preserving the intermittency and the overall mean and variance observed by the rain gauges. The validation showed an improvement in the bias and a correct ensemble distribution in the rank variograms reported in Fig. 7.

The residual mismatch in the validation and the higher bars at the positive extreme of some of the rank variograms can be explained with other sources of uncertainty, due for example to the rain gauge errors, to the different areal representativeness of ensemble and rain gauges, and to the geostatistical approximations and assumptions unavoidable in the model.

The hydrological application is an important test, because the use of ensembles is particularly important in model application: the nature of radar QPE ensembles makes them very suitable to Monte Carlo-based rainfall input uncertainty propagation assessment. The use of ensembles as compared to deterministic prediction of the flow can improve the understanding of the relative rainfall uncertainty and partially captures the variability of the discrepancy between modelled and observed flows. The radar QPE uncertainty, modelled through the ensemble, is only one source of uncertainty. Other sources of uncertainty, like the simplified model structure, numerical approximations, parameter estimation errors, or uncertainty in the other inputs and in the measured flow, contribute to the overall discrepancy between simulated and observed flow. The PDM hydrological model, has a simple lumped structure, therefore many approximations and simplifications are made. The use of a more complex model could reduce the structural uncertainty, but would require more parameters and more inputs with associated uncertainty and would make the calibration more difficult. The use of radar rainfall ensembles can explain which part of the discrepancy between the modelled flow and the observed one is due to rainfall uncertainty, and which part is due to other uncertainty sources.

In terms of possible improvements of the work, it could be interesting to address the problem of spatial stationarity of the errors already in the geostatistical modelling. In this work, in fact, the temporal variability of the error characteristics is represented in trade of an assumption of spatial stationarity of the radar residual errors. Although error temporal variability is shown to be more significant than the spatial one in terms of error mean and standard deviation, spatial stationarity is not a realistic assumption, because radar residual errors characteristics are actually varying in space, for example proportionally to the rainfall rate, increasing with the distance from the radars, or with the presence of reliefs and obstacles. The problem of error characteristic spatial variability could be partially addressed firstly introducing anisotropy in the variogram approach. Changing more drastically the methodology, it could be

possible to develop methods able to integrate external factors in the error spatial characterisation and in the simulation, like the distance from the radar, elevation information, and so on. Another interesting approach that could be applied is the modelling of the "dry drift", i.e. the fact that average areal rainfall intensity varies with the distance to the surrounding dry areas (Schleiss et al., 2014). The use of conditional simulations helps in reproducing the observed errors, and therefore their spatial characterisation.

The presented case study is a wide area where around 200 rain gauges are available. A sufficient number of rain gauges is necessary for the application of the method, as well as a sufficient density, in order to reliably measure the radar error characteristics during rainfall events. The ensemble generator is therefore designed for sufficiently large and well instrumented areas, and cannot be used for particularly small study areas, or data-scarce environments.

Another important issue that is debated in the literature is the opportunity to use rain gauges as an approximation of true rainfall, given that ground measurements are subject to uncertainty as well, and that they are representative of points, while radar estimations are areal (Bringi et al., 2011; Ciach and Krajewski, 1999; Kitchen and Blackall, 1992; Peleg et al., 2013). This approximation is still considered better than the radar error estimation made with other methodologies, for example analysing and modelling each source of error independently. Nevertheless, it could be interesting to integrate an estimation of rain gauge uncertainty in the radar error models.

Finally, this work addressed the modelling of the residual errors of radar QPE, but does not improve the deterministic precipitation estimates. Methods to merge radar data and rain gauges could potentially improve the radar QPE. A possible future development could be to adapt the presented methodology to merged products, keeping in mind that the error estimation and modelling needs to be approached differently, because in this case rain gauges cannot be used as reference anymore.

Acknowledgements

This work has been completed as part of the Marie Curie Initial Training Network QUICS. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under Grant agreement No 607000. M.A. Rico-Ramirez also acknowledges the support of the Engineering and Physical Sciences Research Council (EPSRC) via Grant EP/I012222/1. The authors would like to thank the UK Met Office (meteorological data sets requested at: Met Office 2003, 2012), the Environment Agency (rain gauge data sets requested at http://environment-agency.gov.uk/), and the Centre for Ecology and Hydrology & the National River Flow Archive (river flow data sets requested at http://nrfa.ceh.ac.uk/) for providing the various data sets to develop this study. We would like to thank Dr N. Nanding and Dr Jia Liu who assisted in the hydrological model set up, and H. Badger for the python code support. We would like to thank the anonymous reviewers for their helpful and constructive comments that helped to improve this manuscript.

References

AghaKouchak, A., Bárdossy, A., Habib, E., 2010a. Copula-based uncertainty modelling: Application to multisensor precipitation estimates. Hydrol. Process. 24, 2111–2124. http://dx.doi.org/10.1002/hyp.7632.

AghaKouchak, A., Habib, E., Bárdossy, A., 2010b. Modeling radar rainfall estimation uncertainties: random error model. J. Hydrol. Eng. 15, 265–274. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000185.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements – FAO Irrigation and drainage paper 56.

Anagnostou, E., Krajewski, W.F., 1999. Uncertainty quantification of mean-areal radar-rainfall estimates. J. Atmos. Ocean. Technol., 206–215

Atlas, D., Banks, H.C., 1951. The interpretation of microwave reflections from rainfall. J. Meteorol. 8, 271–282. http://dx.doi.org/10.1175/1520-0469(1951)008<0271:tiomrf>2.0.co;2.

Austin, P.M., 1987. Relation between measured radar reflectivity and surface rainfall. Mon. Weather Rev. 115, 1053–1070. http://dx.doi.org/10.1175/1520-0493(1987) 115<1053:RBMRRA>2.0.CO;2.

Austin, P.M., Bernis, A.C., 1950. A quantitative study of the "bright band" in radar precipitation echoes. J. Meteorol. 7, 145–151. http://dx.doi.org/10.1175/1520-0469(1950) 007<0145:AQSOTB>2.0.CO;2.

Berne, A., Krajewski, W.F., 2013. Radar for hydrology: unfulfilled promise or unrecognized potential? Adv. Water Resour. 51, 357–366. http://dx.doi.org/10.1016/j.advwatres.2012.05.005.

Bringi, V.N., Rico-Ramirez, M., Thurai, M., 2011. Rainfall estimation with an operational polarimetric C-Band radar in the united kingdom: comparison with a gauge network and error analysis. J. Hydrometeorol. 12, 935–954. http://dx.doi.org/10.1175/JHM-D-10-05013.1.

Ciach, G.J., Krajewski, W.F., 1999. On the estimation of radar rainfall error variance. Adv. Water Resour. 22, 585–595. http://dx.doi.org/10.1016/S0309-1708(98)00043-8.

Ciach, G.J., Krajewski, W.F., Villarini, G., 2007. Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. J. Hydrometeorol. 8, 1325–1347. http://dx.doi.org/10.1175/2007JHM814.1.

Cressie, N., 1985. Fitting variogram models by weighted least squares. J. Int. Assoc. Math. Geol. 17, 563–586. http://dx.doi.org/10.1007/BF01032109.

Cressie, N.A.C., 1993. Statistics for Spatial Data.

Dai, Q., Han, D., Rico-Ramirez, M., Srivastava, P.K., 2014. Multivariate distributed ensemble generator: a new scheme for ensemble radar precipitation estimation over temperate maritime climate. J. Hydrol. 511, 17–27. http://dx.doi.org/10.1016/j.jhydrol.2014.01.016.

Dai, Q., Han, D., Zhuo, L., Huang, J., Islam, T., Srivastava, P.K., 2015. Impact of complexity of radar rainfall uncertainty model on flow simulation. Atmos. Res. 161–162, 93–101. http://dx.doi.org/10.1016/j.atmosres.2015.04.002.

Delhomme, J.P., 1979. Spatial variability and uncertainty in groundwater flow parameters – a geostatistical approach. Water Resour. Res. 15, 269–280.

Delrieu, G., Andrieu, H., Creutin, J.D., 2000. Quantification of path-integrated attenuation for X- and C-Band weather radar systems operating in mediterranean heavy rainfall. J. Appl. Meteorol. 39, 840–850. http://dx.doi.org/10.1175/1520-0450(2000) 039<0840:QOPIAF>2.0.CO;2.

Doviak, R.J., 1983. A survey of radar rain measurement techniques. J. Clim. Appl. Meteorol. http://dx.doi.org/10.1175/1520-0450(1983) 022<0832:ASORRM>2.0.CO;2.

Duan, Q., Sorooshian, S., Gupta, H.V., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031. http://dx.doi.org/10.1029/91WR02985.

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158, 265–284. http://dx.doi.org/10.1016/0022-1694(94)90057-4.

Erdin, R., Frei, C., Künsch, H.R., 2012. Data transformation and uncertainty in geostatistical combination of radar and rain gauges. J. Hydrometeorol. 13, 1332–1346. http://dx.doi.org/10.1175/JHM-D-11-096.1.

Fabry, F., Zawadzki, I., 1995. Long-term radar observations of the melting layer of precipitation and their interpretation. J. Atmos. Sci. 52, 838–851. http://dx.doi.org/10.1175/1520-0469(1995) 052<0838:LTROOT>2.0.CO;2.

Fabry, F., Bellon, A., Duncan, M.R., Austin, G.L., 1994. High resolution rainfall measurements by radar for very small basins: the sampling problem reexamined. J. Hydrol. 161, 415–428. http://dx.doi.org/10.1016/0022-1694(94)90138-4.

Friedrich, K., Germann, U., Gourley, J.J., Tabary, P., 2007. Effects or radar beam shielding on rainfall estimation for the polarimetric C-band radar. J. Atmos. Ocean. Technol. 24, 1839–1859. http://dx.doi.org/10.1175/JTECH2085.1.

Ge, G., Gao, J., Brewster, K., Xue, M., 2010. Impacts of beam broadening and earth curvature on storm-scale 3D variational data assimilation of radial velocity with two doppler radars. J. Atmos. Ocean. Technol. 27, 617–636. http://dx.doi.org/10.1175/2009jtecha1359.1.

Germann, U., Berenguer, M., Sempere-Torres, D., Zappa, M., 2009. REAL – ensemble radar precipitation estimation for hydrology in mountainous region. Q. J. R. Meteorol. Soc 135, 445–456. http://dx.doi.org/10.1002/qj.375.

Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Mon. Weather Rev. 129, 550–560. http://dx.doi.org/10.1175/1520-0493(2001) 129<0550:IORHFV>2.0.CO;2.

Harrison, D.L., Driscoll, S.J., Kitchen, M., 2000. Improving precipitation estimates from weather radar using quality control and correction techniques. Meteorol. Appl. 7, 135–144. http://dx.doi.org/10.1017/S1350482700001468.

Hasan, M.M., Sharma, A., Johnson, F., Mariethoz, G., Seed, A., 2014. Correcting bias in radar Z-R relationships due to uncertainty in point rain gauge networks. J. Hydrol. 519, 1668–1676. http://dx.doi.org/10.1016/j.jhydrol.2014.09.060.

Hubbert, J.C., Dixon, M., Ellis, S.M., 2009. Weather radar ground clutter. Part II: real-time identification and filtering. J. Atmos. Ocean. Technol. 26, 1181–1197. http://dx.doi.org/10.1175/2009JTECHA1160.1.

Hubbert, J.C., Dixon, M., Ellis, S.M., Meymaris, G., 2009. Weather radar ground clutter. Part I: identification, modeling, and simulation. J. Atmos. Ocean. Technol. 26, 1165–1180. http://dx.doi.org/10.1175/2009JTECHA1159.1.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Networks 13, 411–430.

Islam, T., Rico-Ramirez, M.a., Han, D., Srivastava, P.K., 2012. Artificial intelligence techniques for clutter identification with polarimetric radar signatures. Atmos. Res. 109–110, 95–113. http://dx.doi.org/10.1016/j.atmosres.2012.02.007.

Joanes, D.N., Gill, C.A., 1998. Comparing measures of sample skewness and kurtosis. J. R. Stat. Soc. Ser. D (The Stat.) 47, 183–189. http://dx.doi.org/10.1111/1467-9884.00122.

Joss, J., Lee, R., 1995. The application of radar-gauge comparisons to operational precipitation profile corrections. J. Appl. Meteorol. 34, 2612–2630. http://dx.doi.org/10.1175/1520-0450(1995) 034<2612:TAORCT>2.0.CO;2.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London.

Kirstetter, P.-E., Delrieu, G., Boudevillain, B., Obled, C., 2010. Toward an error model for radar quantitative precipitation estimation in the Cévennes-Vivarais region, France. J. Hydrol. 394, 28–41. http://dx.doi.org/10.1016/j.jhydrol.2010.01.009.

Kirstetter, P.E., Andrieu, H., Boudevillain, B., Delrieu, G., 2013. A physically based identification of vertical profiles of reflectivity from volume scan radar data. J. Appl. Meteorol. Climatol. 52, 1645–1663. http://dx.doi.org/10.1175/JAMC-D-12-0228.1.

Kirstetter, P.-E., Gourley, J.J., Hong, Y., Zhang, J., Moazamigoodarzi, S., Langston, C., Arthur, A., 2015. Probabilistic precipitation rate estimates with ground-based radar networks 1–56.

Kitchen, M., Blackall, R.M., 1992. Representativeness errors in comparisons between radar and gauge measurements of rainfall. J. Hydrol. 134, 13–33. http://dx.doi.org/10.1016/0022-1694(92)90026-R.

Kitchen, M., Jackson, P.M., 1993. Weather radar performance at long range – simulated and observed. J. Appl. Meteorol. 32, 975–985.

Le Ravalec, M., Noetinger, B., Hu, L.Y., 2000. The FFT moving average (FFT-MA) generator: an efficient numerical method for generating and conditioning Gaussian simulations. Math. Geol. 32, 701–723. http://dx.doi.org/10.1023/A:1007542406333.

Marshall, J.S., Langille, R.C., Palmer, W.M.K., 1947. Measurement of rainfall by radar. J. Meteorol. 4, 186–192. http://dx.doi.org/10.1175/1520-0469(1947) 004<0186:MORBR>2.0.CO;2.

McKee, J.L., Binns, A.D., 2015. A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology. Can. Water Resour. J./Rev. Can. des ressources hydriques 1784, 1–18. http://dx.doi.org/10.1080/07011784.2015.1064786.

Meneghini, R., 1978. Rain-rate estimates for an attenuating radar. Radio Sci. 13, 459–470.

Met Office, 2003. 1 km Resolution UK Composite Rainfall Data from the Met Office Nimrod System [WWW Document]. NCAS British Atmos. Data Cent. URL http://catalogue.ceda.ac.uk/uuid/27dd6ffba67f667a18c62de5c3456350.

Met Office, 2012. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current) [WWW Document]. NCAS Br. Atmos. Data Cent. URL . http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0.

Metropolis, N., Ulam, S., 1949. The Monte Carlo Method. J. Am. Stat. Assoc. http://dx.doi.org/10.1080/01621459.1949.10483310.

Monteith, J.L., 1965. Evaporation and environment. Symp. Soc. Exp. Biol. 19, 205–234.

Moore, R.J., 2007. The PDM rainfall-runoff model. Hydrol. Earth Syst. Sci. Discuss. 11, 483–499.

Moszkowicz, S., Ciach, G.J., Krajewski, W.F., 1994. Statistical detection of anomalous propagation in radar reflectivity patterns. J. Atmos. Ocean. Technol. 11, 1026–1034. http://dx.doi.org/10.1175/1520-0426(1994) 011<1026:SDOAPI>2.0.CO;2.

Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: or seven reasons not to use uncertainty analysis. Water Resour. Res. 42, 1–8. http://dx.doi.org/10.1029/2005WR004820.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691. http://dx.doi.org/10.1016/j.cageo.2004.03.012.

Pegram, G., Llort, X., Sempere-Torres, D., 2011. Radar rainfall: separating signal and noise fields to generate meaningful ensembles. Atmos. Res. 100, 226–236. http://dx.doi.org/10.1016/j.atmosres.2010.11.018.

Peleg, N., Ben-Asher, M., Morin, E., 2013. Radar subpixel-scale rainfall variability and uncertainty: lessons learned from observations of a dense rain-gauge network. Hydrol. Earth Syst. Sci. 17, 2195–2208. http://dx.doi.org/10.5194/hess-17-2195-2013.

Qi, Y., Zhang, J., Zhang, P., Cao, Q., 2013. VPR correction of bright band effects in radar QPEs using polarimetric radar observations. J. Geophys. Res. Atmos. 118, 3627–3633. http://dx.doi.org/10.1002/jgrd.503642013.

Rico-Ramirez, M.A., Cluckie, I.D., 2008. Classification of ground clutter and anomalous propagation using dual-polarization weather radar. IEEE Trans. Geosci. Remote Sens. 46, 1892–1904.

Rico-Ramirez, M.a., Cluckie, I.D., 2007. Bright-band detection from radar vertical reflectivity profiles. Int. J. Remote Sens. 28, 4013–4025. http://dx.doi.org/10.1080/01431160601047797.

Rico-Ramirez, M.a., Liguori, S., Schellart, a.N.a., 2015. Quantifying radar-rainfall uncertainties in urban drainage flow modelling. J. Hydrol. 528, 17–28. http://dx.doi.org/10.1016/j.jhydrol.2015.05.057.

Schleiss, M., Chamoun, S., Berne, A., 2014. Nonstationarity in intermittent rainfall: the "dry drift". J. Hydrometeorol. 15, 1189–1204. http://dx.doi.org/10.1175/JHM-D-13-095.1.

Schröter, K., Llort, X., Velasco-Forero, C., Ostrowski, M., Sempere-Torres, D., 2011. Implications of radar rainfall estimates uncertainty on distributed hydrological model predictions. Atmos. Res. 100, 237–245. http://dx.doi.org/10.1016/j.atmosres.2010.08.014.

Seed, A.W., Nicol, J.C., Austin, G.L., Stow, C.D., Bradley, S.G., 2007. The impact of radar and raingauge sampling errors when calibrating a weather radar. Meteorol. Appl. 3, 43–52. http://dx.doi.org/10.1002/met.5060030105.

Smith, J.M., 1977. Mathematical Modelling and Digital Simulation for Engineers and Scientists. John Wiley & Sons Inc, New York, NY, USA.

Smith, C.J., 1986. The reduction of errors caused by bright bands in quantitative rainfall measurements made using radar. J. Atmos. Ocean. Technol. 3, 129–141. http://dx.doi.org/10.1175/1520-0426(1986) 003<0129:TROECB>2.0.CO;2.

Steiner, M., Smith, J.a., 2002. Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data. J. Atmos. Ocean. Technol. 19, 673–686. http://dx.doi.org/10.1175/1520-0426(2002) 019<0673:UOTDRS>2.0.CO;2.

Uijlenhoet, R., Berne, A., 2008. Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction. Hydrol. Earth Syst. Sci. 12, 587–601. http://dx.doi.org/10.5194/hess-12-587-2008.

van Griensven, A., Meixner, T., 2007. A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models. J. Hydroinf. 9, 277. http://dx.doi.org/10.2166/hydro.2007.104.

Villarini, G., Krajewski, W.F., 2009. Empirically based modelling of radar-rainfall uncertainties for a C-band radar at different time-scales. Q. J. R. Meteorol. Soc. 1438, 1424–1438. http://dx.doi.org/10.1002/qj.

Villarini, G., Krajewski, W.F., 2010. Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. Surv. Geophys. 31, 107–129. http://dx.doi.org/10.1007/s10712-009-9079-x.

Villarini, G., Krajewski, W.F., Ciach, G.J., Zimmerman, D.L., 2009. Product-error-driven generator of probable rainfall conditioned on WSR-88D precipitation estimates. Water Resour. Res. 45, 1–11. http://dx.doi.org/10.1029/2008WR006946.

Villarini, G., Seo, B.C., Serinaldi, F., Krajewski, W.F., 2014. Spatial and temporal modeling of radar rainfall uncertainties. Atmos. Res. 135–136, 91–101. http://dx.doi.org/10.1016/j.atmosres.2013.09.007.

Westrick, K.J., Mass, C.F., Colle, B.a., 1999. The Limitations of the WSR-88D Radar Network for Quantitative Precipitation Measurement over the Coastal Western United States. Bull. Am. Meteorol. Soc. 80, 2289–2298. http://dx.doi.org/10.1175/1520-0477(1999) 080<2289:TLOTWR>2.0.CO;2.

Zhang, X.F., Eijkeren, J.C.H.V.A.N., 1995. On the weighted least-squares method for fitting a semivariogram model. Comput. Geosci. 21, 605–608.

Zhang, J., Qi, Y., 2010. A real-time algorithm for the correction of brightband effects in radar-derived QPE. J. Hydrometeorol. 11, 1157–1171. http://dx.doi.org/10.1175/2010JHM1201.1.