Modelling radar rainfall residual errors for hydrological applications

Authors: F. Cecinati*, M. A. Rico-Ramirez, D. Han. Department of Civil Engineering, University of Bristol. *francesca.cecinati@bristol.ac.uk

1 - Introduction

Weather radars are a key source of precipitation data in many hydrological applications, thanks to their wide coverage and high spatio-temporal resolution. Nevertheless, the complex nature of the radar systems make them prone to different sources of errors. Although many can be partially corrected, a residual uncertainty is unavoidable in the radar-derived quantitative precipitation estimates. When it comes to hydrological applications, the estimation of uncertainty and assessing its propagation in models is essential. Radar rainfall ensembles are a good methods to model uncertainty in radar rainfall for model applications. This approach consists in estimating the errors and their characteristics by comparing radar rainfall with point ground measurements used as an approximation of true rainfall. Knowing the statistical characteristics of residual errors, a large number of possible, alternative realizations of the rainfall fields are simulated, constituting an ensemble. The error propagation estimation can be accomplished observing the result spread feeding a model with the different ensemble members. Methods to estimate errors and generate ensembles are various in literature, but many are based on the computation of the error covariance matrix (see Germann et al. 2009). The error covariance matrix approach works well with a medium number of point measurements, but is not very robust nor efficient when the number of rain gauges is too large. In addition, it generates error components for the ensemble only in ground measurements points, needing subsequent interpolation, and it consider temporal stationarity.

This work proposes a different approach that trades temporal with spatial stationarity assumption, does not require interpolation and improves robustness and efficiency of the algorithm. The error model is assumed multiplicative. The spatial correlation characteristics of the errors are modelled through a semivariogram, fitted with an exponential function. The error components are generated using a FFT Moving Average generator. In addition, the problem of mean and variance inflation resulting from working in the logarithmic domain is addressed and a simple solution is proposed. Finally, the generated ensembles are used in a case study involving two catchments in England, using PDM models. Comparing the observed flow and the output of the models fed with the ensemble is a good way to assess the quality of the generated ensembles and to observe uncertainty propagation.

2 - Error Model

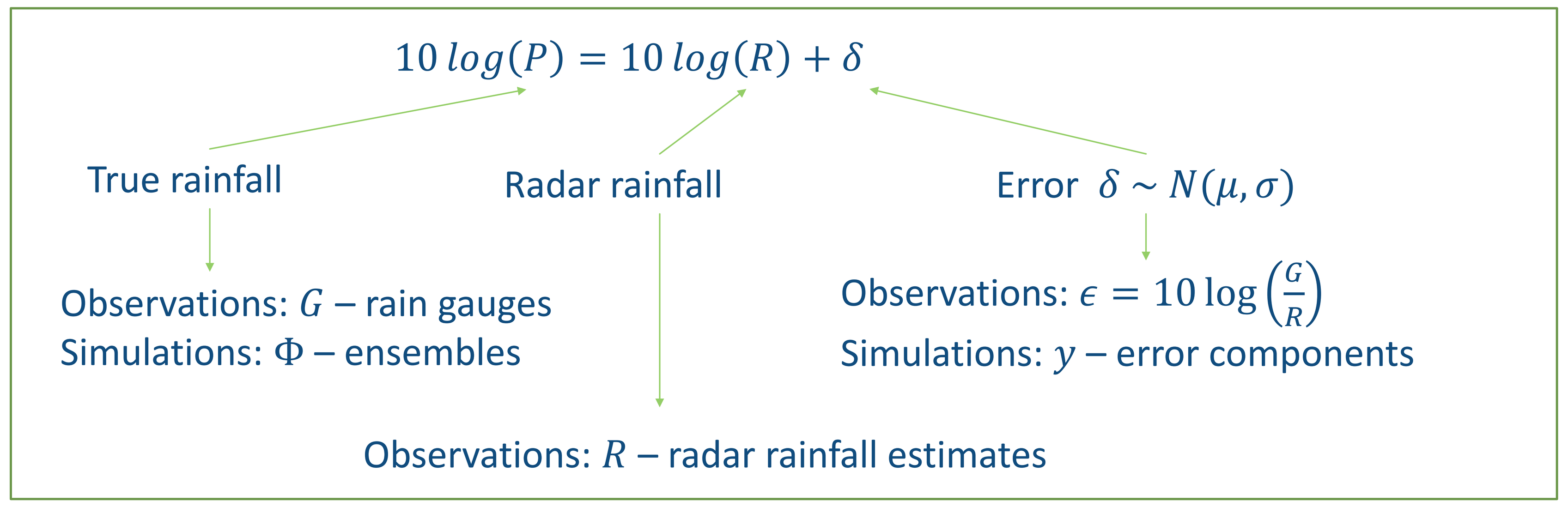


Figure 1 – The scheme above represents the model adopted in this work. Rain gauge data G are assumed as an approximation of true rainfall P and the error component δ can be observed deriving ϵ from the observations of radar R and rain gauge G data. Repproducing the characteristics of ϵ, alternative error fields y can be generated and used, following the model equation, to produce ensemble members Φ.

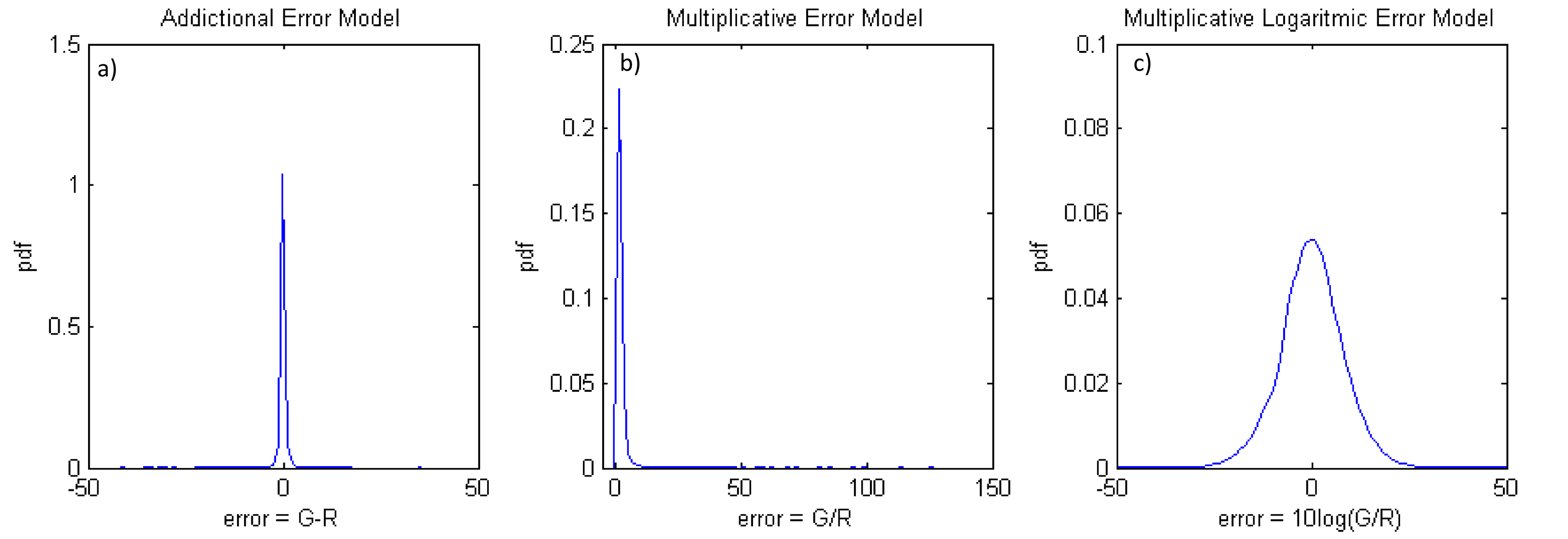


Figure 2 – Using a logarithmic multiplicative model (c), the observed errors have a probability distribution function very close to Gaussian and therefore can be treated as Gaussian and characterized by their mean and variance. This would not happen using simple additive (a) or multiplicative (b) models.

3 - Method

Semivariogram

In this work, the spatial correlation characteristics of the residual errors is modelled through semivariograms. Semivariograms describe the increase of variance that occurs considering elements at increasing distance.

$$\gamma(\epsilon,d) = \frac{1}{2}E\left\{\left(\epsilon-(\epsilon+d)\right)^2\right\}$$

Empirically, it is calculated grouping all the possible element couples in distance bins (here 1km wide). Subsequently, it is fitted with an exponential function:

$$\gamma(d) = (s)\left(1-\exp\left(-\frac{3d}{r}\right)\right)$$

This approach has two advantages:

1. It does not require temporal stationarity as the covariance matrix approach, although it requires a spatial stationarity assumption;
2. It is very light: at each time step the spatial characteristics of the errors can be described by only two parameters: the sill s and the range r.

FFT Moving Average

The FFT Moving Average (FFT-MA) random field generator was introduced by Le Ravalec et al. in 2000.
It generate a random field with a given semivariogram, under the assumption of stationarity in the simulating domain.

1. The covariance function can be derived from the semivariogram:
$$C(t,d) = \sigma^2 - \gamma(t,d)$$
2. The covariance function can be written as a convolution of a function g and its transpose $\breve{g}$:
$$C = g * \breve{g} \quad \text{where} \quad \breve{g}(x) = g(-x)$$
3. The convolution function is used to generate a Gaussian random field y with mean μ and covariance C as follow:
$$y = \mu + z * g$$
Where z are normal deviates.
4. (Le Ravalec et al. 2000) demonstrate that, in case of stationarity, which we assume in the analyzed time window, the Fourier transform of g is obtained as:
$$G(f) = \sqrt{dx\,S(f)}$$
Where dx is the spatial sampling.
5. Hence, the product $Z \cdot G$ can be calculated and transformed in the space domain in $z * g$, and the error components y are generated with the equation at 3.

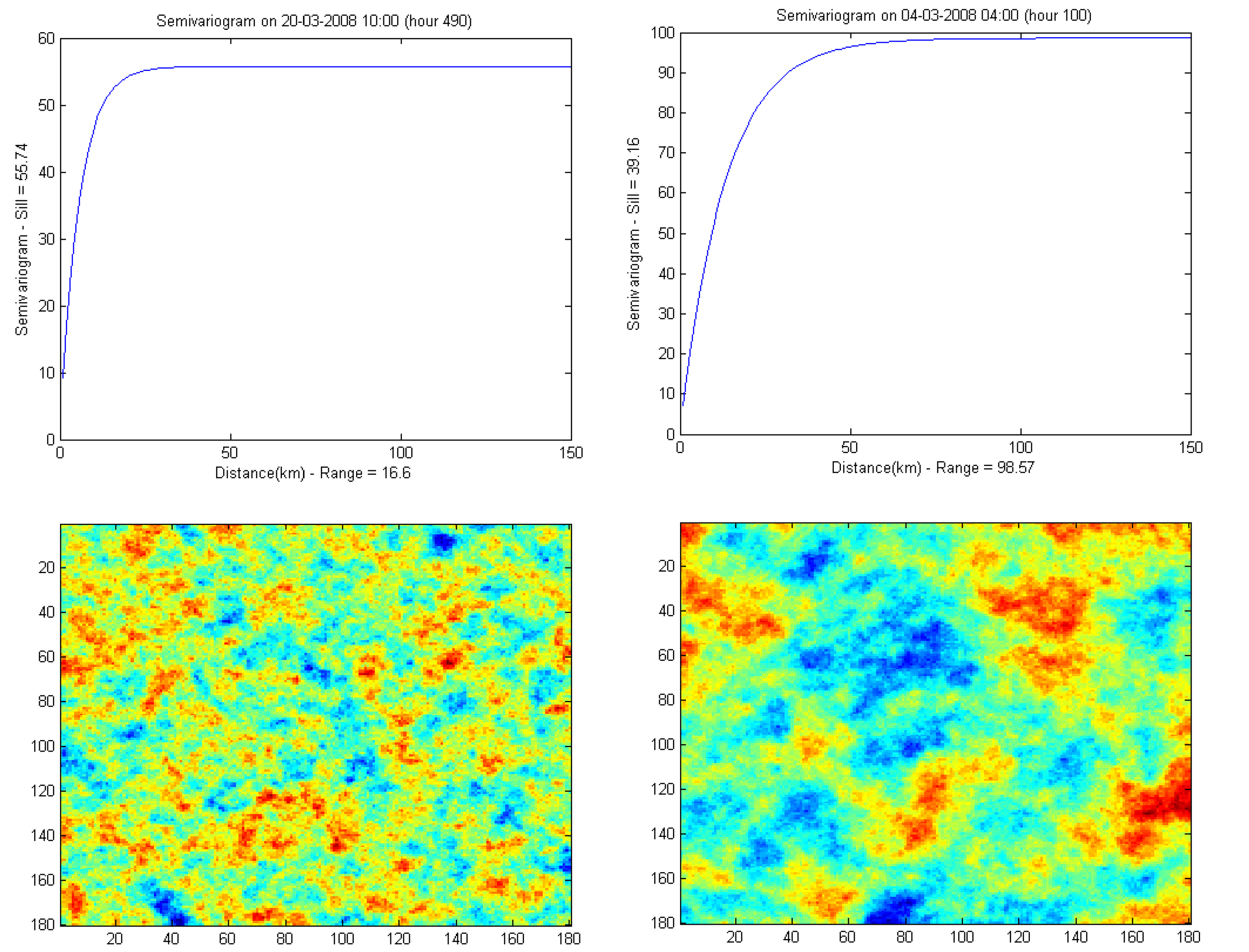


Figure 3 – An example of two observed semivariograms, with different shapes, and the corresponding error components generated with the FFT-MA method. The semivariogram on the left has a smaller range, therefore variations occur in a shorter space, the one on the right has a larger range, which generates a smoother field.

4 - Time Dependent Error Characteristics

Thanks to the use of semivariograms to model the spatial correlation of the residual errors, the assumption of temporal stationarity is no more necessary and the residual error characteristics (mean, standard deviation and semivariogram parameters) can be calculated at each time step using the observations in the 12 hours precedent the analysed time step. 12 hours is selected as a compromise between the temporal and the statistical representativeness of the calculated parameters.

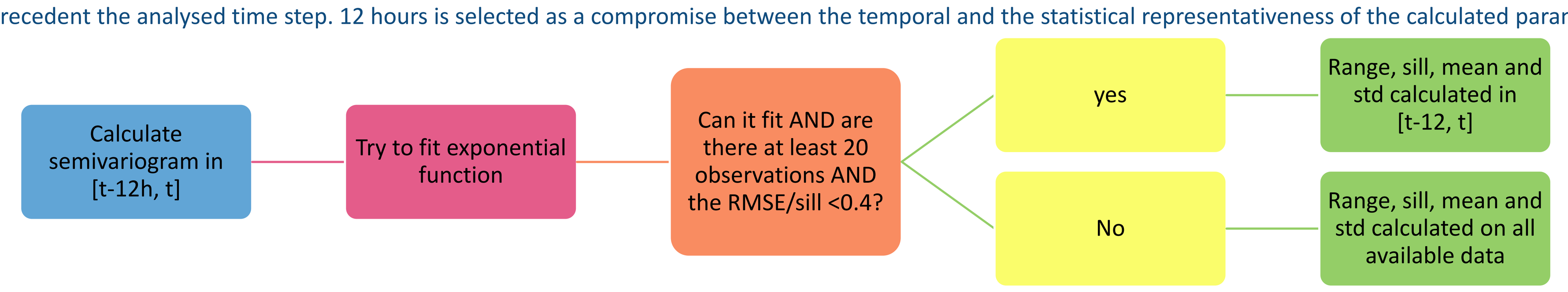


Figure 5 – The scheme represents the method followed to calculate error characteristics in real time, when possible. Usually when it is not possible it is because there is too little rain, therefore using the average statistics does not have a big impact.

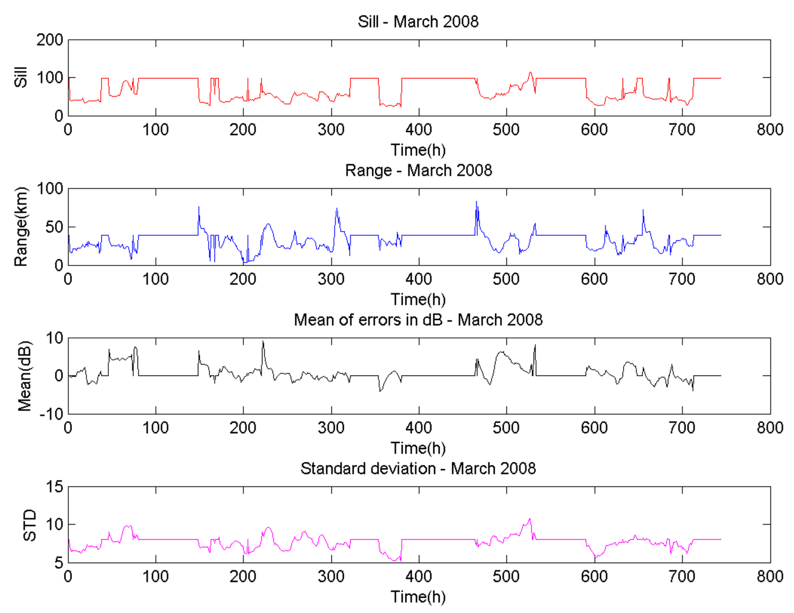


Figure 6 – The residual error characteristics, namely sill, range, mean, and standard deviation, are reported for the month of March 2008 as an example. This example clearly shows how this values vary in time, often with an observable autocorrelation. The average observed values are evident as plateaus when the specific values were not measurable.

5 - Variance and mean correction

The modelled error components are Gaussian. They are recombined with the radar rainfall estimates to obtain the ensemble members following the model:

$$10\log(\Phi) = 10\log(R) + y$$

This operation, done in the logarithmic domain, brings to a distortion when the ensemble members are re-transformed. The obtained ensemble members have on average higher mean and standard deviation, compared to the observations. We opted for a linear correction that preserve the ensemble spatial characteristics:

$$\Phi_{new,i} = \frac{\sigma_G}{\sigma_{\overline{\Phi_{old}}}}\left(\Phi_{old,i} - m_{\overline{\Phi_{old}}}\right) + m_G$$

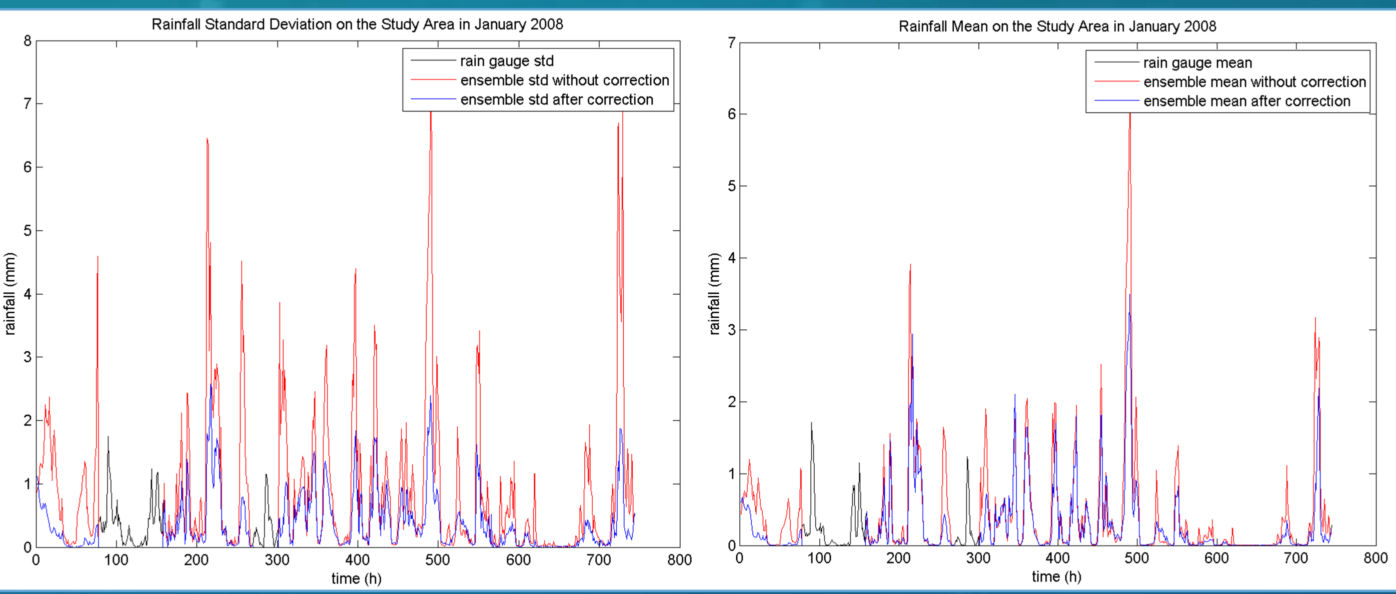


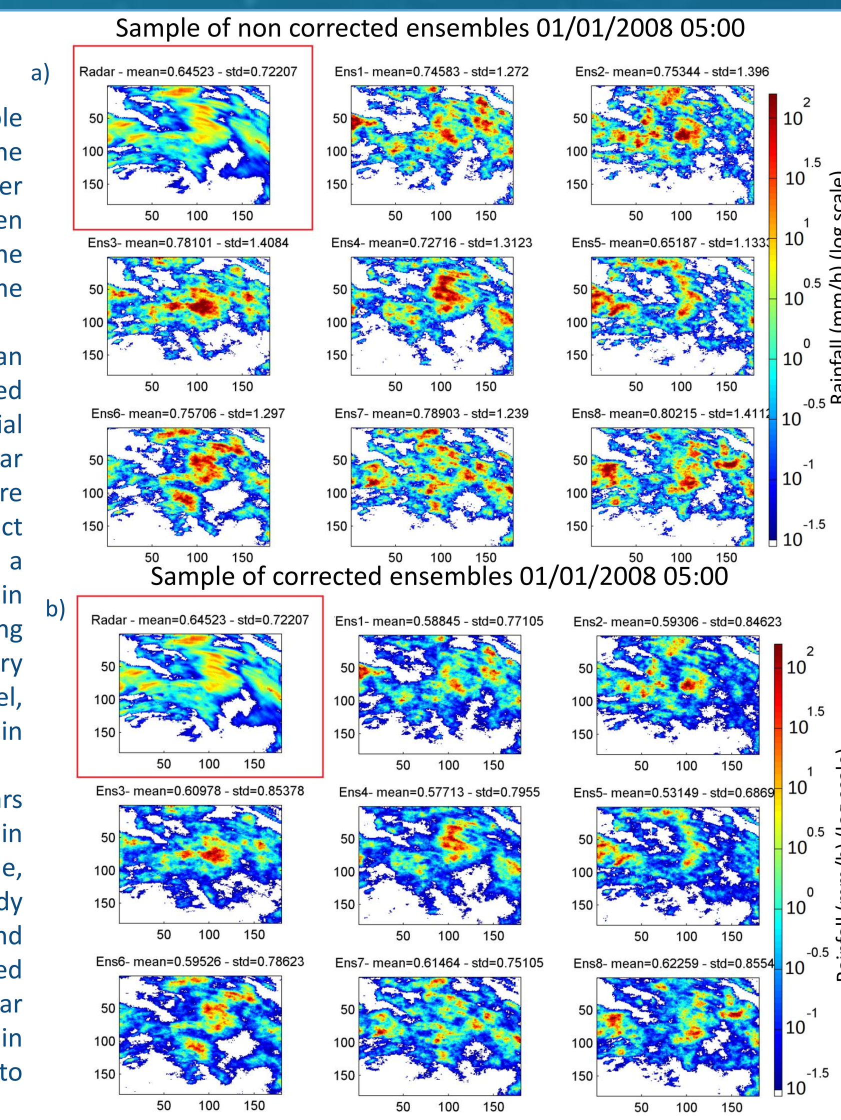
Figure 4 – The mean and STD of the ensembles before correction are on average higher than the observations. After correction the average of the ensembles coincides with the observations (In some intervals the ensembles could not be generated because of missing radar data).

6 - Results

In order to test the method, an ensemble of 100 members was produced in a one year interval and analysed. The number was selected as a compromise between the statistical representativeness and the feasibility of producing it in a real-time scenario.

Figure 7 shows the benefits of the mean and variance correction. The produced ensembles maintain a similar spatial structure compared to the radar acquisition, although they show more granularity. This is due to both the fact that the radar usually presents a smoother behaviour compared to rain gauge observations, due to averaging operations, and to the spatial stationary and isotropy used in the error model, that may not be accurate in certain situations.

Nevertheless, the overall result appears good and captures the uncertainty in radar rainfall estimates. As an example, the rainfall rates on the three study catchments are reported in Figure 8 and compared to the rainfall rates calculated interpolating rain gauge data. It is clear that, although both datasets contain uncertainty, the ensembles are able to account for it.



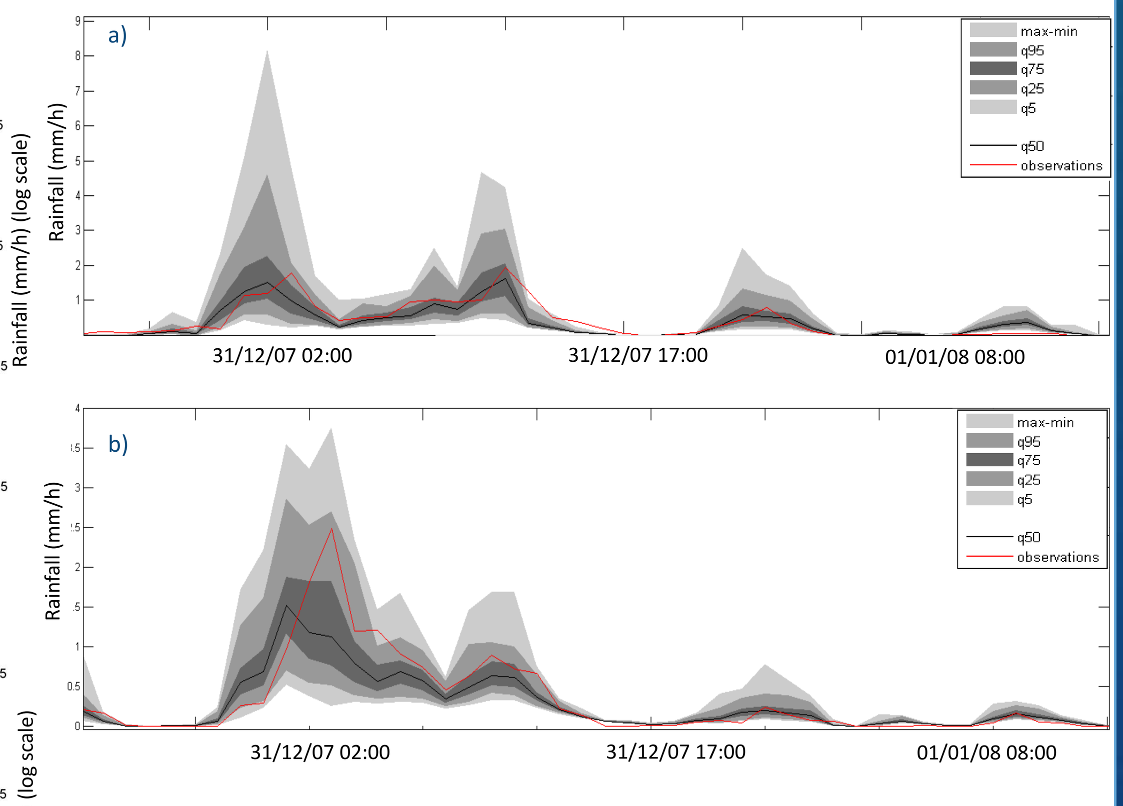


Figure 7 – Example of ensemble members compared with the original radar (in the red square) before (a) and after (b) correction. The event refer to the 1st January 2008 at 05:00.

Figure 8 – Rainfall rate for the same rainfall event in the two study catchments: Lune (a) and Ribble (b). The corrected ensemble quantiles are compared with the rain gauge rainfall rate in red.

7 - Hydrological Applications

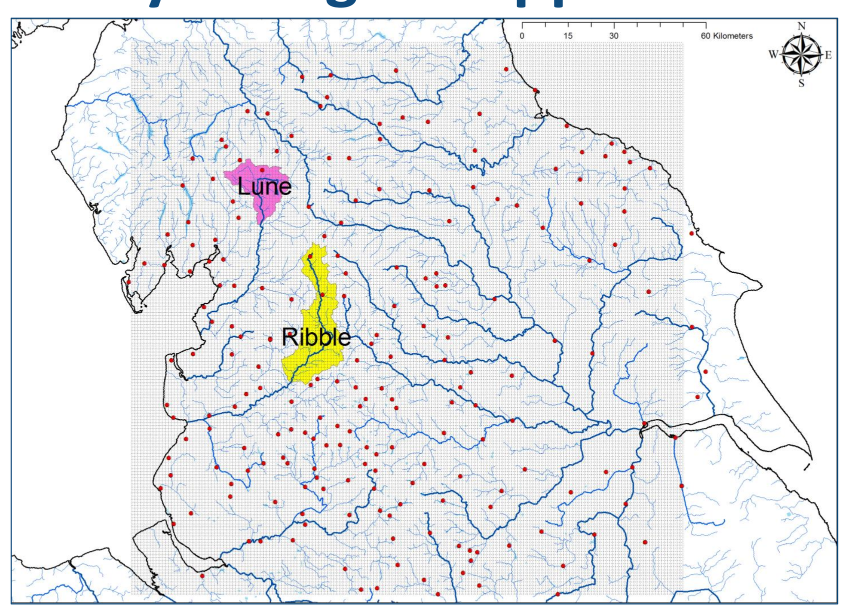


Figure 9 – The two selected catchments are represented with the radar grid and the rain gauges (as red dots), together with the river network.

The direct comparison of the ensemble rainfall rate or rainfall accumulation with the rain gauge or radar rainfall data does not provide a good estimate of the quality of the generated ensemble, because they have been used to condition the ensemble itself. To validate the ensemble, a hydrological case study is used. An area of 180 km x 180 km in England has been selected to generate an ensemble of 100 members for a one-year interval between October 2007 and September 2008. The radar data used is the 1 km NIMROD radar composite from the MetOffice, accumulated at hourly time steps. 203 rain gauges from the Environmental Agency were available. In the area, two catchments were selected, namely the upper part of the Lune and the upper part of the Ribble. For these catchments, flow data are available from the CEH and the necessary meteorological data for a model application were taken from the MIDAS dataset. For the three catchments, the Probability Distributed Model (PDM) was set up and calibrated with only rain gauges data for a two-year interval, between October 2008 and September 2010.

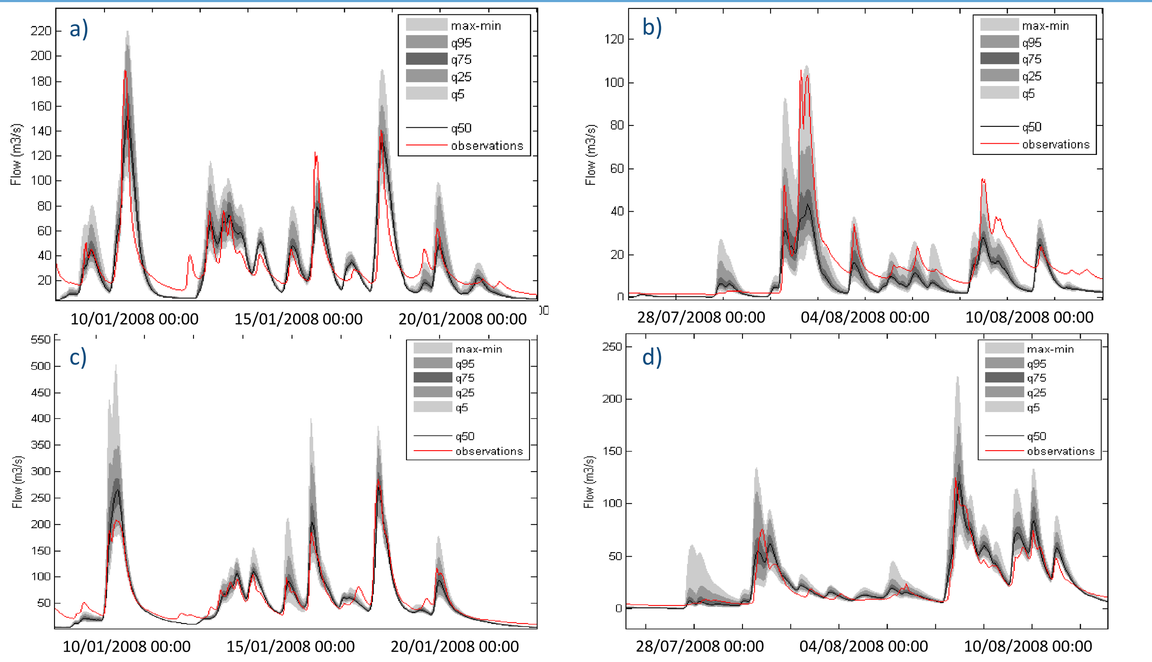


Figure 10 – The output of the PDM model for the Lune (a,b) and Ribble (c,d) basin, using the ensemble are compared with flow observations for a winter event (6 January 2008 – 27 January 2008) (a,c) and for a summer event (24 July 2008 – 14 August 2008) (b,d).

8 - Conclusions and Future Work

The method presented here proved to be a good way of producing radar ensembles. Compared with commonly used covariance matrix approaches, it is more robust and faster when applied to big datasets, and solves some issues like the error component interpolation and the variance and mean inflation. The quality of the results has been tested in two test catchments, both comparing the rainfall rates with the ones derived by rain gauge interpolation, and using a hydrologic model to compare the output flow. While the use of a hydrologic model provides an independent comparison, it also involves other forms of uncertainty (other datasets, model approximations, flow measurement uncertainty, scales and averaging, etc…) that can produce an underestimation of the ensemble quality. One of the next improvements to implement, is to compare the produced ensemble with independent rain gauges, or, in absence of additional data, implementing a cross-validation.

The modelling of temporal variability of residual errors is definitely a positive improvement, but there are some aspects that can be enhanced:

- The semivariogram fitting can be improved, including a selection of the best semivariogram model and of a nugget effect.
- A relaxation of the spatial stationarity assumption can be implemented accounting for anisotropy.
- The rules to accept or reject a semivariogram can be improved to better maintain all meaningful semivariograms.
- A Bayesian inference method can be implemented to derive the best semivariogram when sufficient data is not available.

Another ground for improvement can be the consideration of error temporal correlation structures, which has been neglected in this study because the autocorrelation of residual errors is on average negligible at hourly steps. Nevertheless, exactly as the spatial correlation structure, it can vary in time and be significant in certain situations.

Although there is still room for improvement, the developed method to generate radar rainfall ensembles is already a valuable instrument and could be used in a real-time scenario or in a big data one.

9 - Acknowledgments

This work has been completed as part of the Marie Curie Initial Training Network QUICS. QUICS is supported by the European Commission under the Seventh Framework Programme FP7 with Grant agreement no.: 607000. The authors would like to thank the UK Met Office and the Environment Agency which provided the radar rainfall data, the rain gauge data, and the river flow data to develop this study. A special thank is also for Prof. Gerard Heuvelink from the University of Wageningen who provided a precious feedback; and N Nanding and Jia Liu, from the University of Bristol, who assisted in the hidrological model set up.

10 - References

AghaKouchak, a., Habib, E. & Bárdossy, a., 2010. Modeling Radar Rainfall Estimation Uncertainties: Random Error Model. *Journal of Hydrologic Engineering*, 15(April), pp.265–274.
Bohling, G., 2005. Introduction to Geostatistics and Variogram Analysis.
Ciach, G.J., Krajewski, W.F. & Villarini, G., 2007. Product-Error-Driven Uncertainty Model for Probabilistic Quantitative Precipitation Estimation with NEXRAD Data. *Journal of Hydrometeorology*, 8(6), pp.1325–1347.
Germann, U. et al., 2009. REAL – Ensemble radar precipitation estimation for hydrology in mountainous region. *Q. J. R. Meteorol. Soc*, 135(February), pp.445–456.
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. *Mathematical Geology*, 32(6), pp.701–723.
Moore, R. J.. 2007. The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, 11, pp. 483-499.
Pappenberger, F. & Beven, K.J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Water Resources Research*, 42(December 2005), pp.1–8.
Pegram, G., Llort, X. & Sempere-Torres, D., 2011. Radar rainfall: Separating signal and noise fields to generate meaningful ensembles. *Atmospheric Research*, 100(2-3), pp.226–236.
Rico-Ramirez, M. a., Liguori, S. & Schellart, a. N. a., 2015. Quantifying radar-rainfall uncertainties in urban drainage flow modelling. *Journal of Hydrology*, 528, pp.17–28.
Villarini, G. & Krajewski, W.F., 2010. Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surveys in Geophysics*, 31, pp.107–129.