Hydrol. Earth Syst. Sci. Discuss., 9, 3527–3579, 2012 www.hydrol-earth-syst-sci-discuss.net/9/3527/2012/ doi:10.5194/hessd-9-3527-2012 © Author(s) 2012. CC Attribution 3.0 License.



This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

# Correcting the radar rainfall forcing of a hydrological model with data assimilation: application to flood forecasting in the Lez Catchment in Southern France

E. Harader<sup>1,2</sup>, V. Borrell Estupina<sup>2</sup>, S. Ricci<sup>1</sup>, M. Coustau<sup>2</sup>, O. Thual<sup>1,3</sup>, A. Piacentini<sup>1</sup>, and C. Bouvier<sup>2</sup>

<sup>1</sup>URA CERFACS-CNRS, URA1875, Toulouse, France <sup>2</sup>HSM, UMR5569 CNRS IRD UM1 UM2, Montpellier, France <sup>3</sup>INPT, CNRS, IMFT, Toulouse, France

Received: 21 February 2012 - Accepted: 5 March 2012 - Published: 15 March 2012

Correspondence to: E. Harader (harader@cerfacs.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.



## Abstract

The present study explores the application of a data assimilation (DA) procedure to correct the radar rainfall inputs of an event-based, distributed, parsimonious hydrological model. A simplified Kalman filter algorithm was built on top of a rainfall-runoff model
<sup>5</sup> in order to assimilate discharge observations at the catchment outlet. The study site is the 114 km<sup>2</sup> Lez Catchment near Montpellier, France. This catchment is subject to heavy orographic rainfall and characterized by a karstic geology, leading to flash flooding events. The hydrological model uses a derived version of the SCS method, combined with a Lag and Route transfer function. Because it depends on geographical
<sup>10</sup> features and cloud structures, the radar rainfall input to the model is particularily uncertain and results in significant errors in the simulated discharges. The DA analysis was applied to estimate a constant correction to each event hyetogram. The analysis was carried out for 19 events, in two different modes: re-analysis and pseudo-forecast. In both cases, it was shown that the reduction of the uncertainty in the rainfall data leads

to a reduction of the error in the simulated discharge. The resulting correction of the radar rainfall data was then compared to the mean field bias (MFB), a corrective coefficient determined using ground rainfall measurements, which are more accurate than radar but have a decreased spatial resolution. It was shown that the radar rainfall corrected using DA leads to improved discharge simulations and Nash criteria compared to the MFB correction.

#### 1 Introduction

25

For flash flood prediction, hydrologists may use tools as rudimentary as rainfall-discharge curves or as refined as complex physical and distributed hydrological models, all with the goal of converting atmospheric and soil conditions into discharge volumes, flood peak amplitudes and arrival times. All of these tools are subject to uncertainties related to their inputs and parameterisations. Rainfall-runoff models are sensitive to rainfall



quantities and their spatial distribution throughout the catchment, as runoff generation depends upon rainfall location. Errors in rainfall estimates have a significant impact on prevision and event reconstruction quality. In studies of flash flood modelling for Romanian catchments between 36 and 167 km<sup>2</sup>, Zoccatelli et al. (2010) demonstrated that

- <sup>5</sup> neglecting rainfall spatial variability resulted in a deterioration of the simulation quality. Roux et al. (2011) showed that the MARINE model (Estupina-Borrell, 2004) was dependent upon the distribution of rainfall data in order to correctly represent the soil saturation dynamics of the 545 km<sup>2</sup> Gardon d'Anduze Catchment in Southern France. Finally, in a study of the 2002 flash flood in the Gard Department of Southern France, Andrew Marine and Marin
- <sup>10</sup> Anquetin et al. (2010) demonstrated that the spatial resolution provided by radar rainfall was necessary for the correct representation of flood dynamics in catchments with areas from 2.5 to 99 km<sup>2</sup>.

The sensitivity of models to rainfall distribution highlights the importance of using a rainfall product with a fine spatial and temporal resolution, such as radar rainfall. However, depending on the characteristics of the model and the storm event, finely re-

- <sup>15</sup> However, depending on the characteristics of the model and the storm event, finely resolved radar rainfall may have a variable effect on hydrological behavior and can result in either improvement or degradation of the simulation quality (Tetzlaff and Uhlenbrook, 2005; Anquetin et al., 2010). As demonstrated by Tetzlaff and Uhlenbrook (2005), radar rainfall often degrades simulation efficiency, except in the case of highly variable con-
- vective events, due to decreased data quality with radar. Using a physically-based, distributed model, Anquetin et al. (2010) showed that simulation error increases with the ratio of the rainfall length scale to the catchment scale, arguing for the necessity of radar rainfall. However, properly calibrated lumped models can outperform distributed models when radar rainfall is relatively uniform over the basin (Smith et al., 2004).
- <sup>25</sup> The use of radar data is often limited by increased uncertainties compared to ground rainfall measurements due to non-linearities in the rainfall-reflectivity relationship, ground clutter and beam blocking (Borga, 2002). In the Cévennes region where the Lez Catchment is located, a hilly terrain complicates the process of separating rainfall and terrain backscatter. Pellarin et al. (2002) demonstrated that selecting the scan used in



mountainous regions based on distance considerations, as done so for the HYDRAM rainfall product used in this study, leads to a lower quality rainfall product compared to using a composite (highest quality scan at any given point) method (Cheze and Helloco, 1999). Additionally, in the Lez Catchment, radar data quality varies by season and

is diminished in winter months due to bright band effects related to predominantly stratiform rainfall (Coustau et al., 2011; Emmanuel et al., 2012; Tabary, 2007). Techniques for removing bright band effects are foreseen in the next generation Météo-France rainfall product (PANTHERE), which will help to improve winter rainfall.

A possible post-treatment correction to radar rainfall is the removal of the mean field <sup>10</sup> bias (MFB) (Wilson and Brandes, 1979), a correction which uses gauge data to eliminate errors due to instrumentation and a non-linear vertical profile reflectivity (VPR). Adjustment of radar rainfall using gauge data has been shown to lead to improved prediction accuracy (Vieux and Bedient, 2004; Cole and Moore, 2008). Building on this approach, Kahl and Nachtnebel (2008) implemented a technique which corrects radar <sup>15</sup> rainfall using a correction plane in which the coefficients are determined by minimizing an objective function.

Identifying a correction to the rainfall data input to hydrological models can also be formulated as an inverse problem (Tarantola, 2005; McLaughlin and Townley, 1996) conveniently solved in the framework of data assimilation. Data assimilation is a math-

- ematical technique inherited from estimation theory, initially developped for meteorological applications, that combines measurements with model simulations in order to improve the numerical prediction of a system. Data assimilation leads to an optimal estimation of the set of parameters, initial or boundary conditions which allow the model to accurately represent the system.
- Data assimilation for the improvement of hydrological event reconstruction or forecast has been already demonstrated as effective (Aubert et al., 2003; Moradkhani et al., 2005; Pauwels et al., 2001; Thirel et al., 2010; Vrugt et al., 2005). Land surface data assimilation can also be applied for the correction of hydrological input variables such as soil moisture or surface brightness temperature (Reichle et al., 2002; Crow, 2003;



Meier et al., 2011). Chumchean et al. (2006) applied a Kalman filter technique (Bouttier and Courtier, 1999) to the calculation of the mean field bias with the use of an autoregressive (AR) function to model the time evolution of the MFB; this technique resulted in a reduction in the error associated with the MFB when given sufficient rain gauge data.

Previous studies have adopted a variety of approaches to implementing the Kalman Filter algorithm. A common factor of these non-variational data assimilation applications is that errors in the model states or parameters are assumed to follow Bayesian statistics and their probability density functions are determined recursively. The Kalman

- Filter explicitly propagates these errors in time, assuming a linear dynamic model. The Extended Kalman Filter (EKF) re-estimates the Jacobian of the dynamic model at each new time step in order to address possible non-linearities in the evolution of the errors (Goegebeur and Pauwels, 2007). In systems where the error covariances are either difficult to define or the computational cost to propagate them is too great, the En-
- <sup>15</sup> semble Kalman Filter (EnKF) estimates the error covariance at each timestep using the ensemble members (Weerts and El Serafy, 2006; Pauwels and De Lannoy, 2009). The use of ensemble methods, however, comes at the cost of greatly increasing the number of model runs. The simplified version of the version of the Kalman Filter used in this study is applied over a single time window covering either the entire flood event or
- the start of the flood event and with a scalar control vector containing a corrective coefficient to the input rainfall. The background error variance is predefined using the MFB and does not evolve in time, thus avoiding the need for an EKF or ensemble approach. These simplifications are adapted to the correction of time-invariant parameters which should not evolve over the course of the rainfall event.
- In this paper, discharge observations are assimilated with a simplified Kalman Filter (KF) built on top of a rainfall-runoff model. The DA analysis is applied to identify a constant correction to each event hyetogram described by radar data. A similar system is described in Ricci et al. (2011) for a hydrodynamics model. The analysis is applied over a time window, during which several observations are available and the rainfall



correction is assumed to remain constant. The state variables, outputs, inputs or parameters to be corrected are gathered into a control vector, which in this case contains only the constant rainfall corrective coefficient. The observation operator mapping the control vector onto the observation space is represented by the integration of the hy-

- drological model. The KF algorithm relies on the hypothesis that the relation between 5 the rainfall corrective parameter and the discharge at the catchment outlet is fairly linear. In our study, an outer loop procedure was applied to account for some of the non-linearities in this relation. The analysis was carried out for 19 heavy rainfall events occuring within the Lez catchment in Southern France between 1997 and 2008. The
- events were of variable intensity and had measured peak flows between approximately  $10-400 \text{ m}^3 \text{ s}^{-1}$  at the watershed outlet. It was shown that the reduction of the uncertainty in the rainfall data leads to a reduction of the error in the simulated discharge and that the radar rainfall corrected using DA leads to improved discharge simulations and Nash criteria, over the MFB correction.
- The paper is outlined as follows: Sect. 2 explores pertinent hydrological features of 15 the Lez Catchment and the structure of the hydrological model. The DA procedure is presented in Sect. 3. The results of the study and the impacts of data assimilation upon the efficiency of the hydrological model are described in Sect. 5. Finally, a summary of the key results obtained and suggestions for future research directions are presented in Sect. 6.

20

25

#### Study site and hydrological model description 2

#### The Lez Catchment 2.1

The Lez Catchment in Southern France (Fig. 1) is a medium sized karstic basin located 15 km north of the town of Montpellier. The catchment is 114 km<sup>2</sup> at Lavalette, where discharge measurements are taken. This portion of the Lez river is fed by several upstream tributaries: the Lirou, Yorgues and Terrieu (Fig. 2). Other tributaries found



downstream of Lavalette, such as the Mosson, the Lironde and the Verdanson were not taken into account in this study. If the Mosson tributary is included, the catchment drains a total of 560 km<sup>2</sup>. This larger catchment includes the town of Montpellier which extends from the gauging station at Lavalette in the North to the Mauguio rain gauge 5 in the South.

The Lez River flows for a total of 26 km before emptying into the Mediterranean Sea. The riparian area of the Lez Hydrosystem has been extensively developed downstream of the town of Castelnau-le-Lez and continuing into the estuary, Etangs-Palavasiens. The undeveloped stretch of the Lez river extends for 10–15 km from the source to Castelnau-le-Lez and has an average slope of 3‰, compared to less than 1‰ in developed areas.

The landscape of the Lez Catchment at Lavalette is defined by plains and hilly garrigue with limestone outcrops and very little urbanisation. The plains are composed of 200 to 800 m thick Valanginian marls<sup>1</sup>, covered by soil usually less than 1 m thick.

<sup>15</sup> Land use ranges between agricultural (vineyards) and forest in the plains, along with undeveloped garrigue; the limestone outcrops have very little soil cover and thin vege-tation.

#### 2.1.1 The Karst system

10

The source of the Lez is a seasonal spring which serves as the main outlet of a 380 km<sup>2</sup>

<sup>20</sup> limestone and dolomite karstic aquifer (shown by the dotted line in Fig. 1) (Avias, 1992). When the spring pool reaches a level of +65 m, it overflows into the Lez River. Several smaller seasonal springs drain the same system; these are discussed in greater detail in Fleury et al. (2009). The Lez Spring is actively pumped to provide water to the city of Montpellier. During the summer, part of the pumped flow is returned to the Lez River in order to maintain a minimum discharge of 1601s<sup>-1</sup>. The aquifer is

<sup>&</sup>lt;sup>1</sup>A mixture of calcium carbonate and clay minerals formed during the Early Cretaceous(Valanginian) period



recharged during winter and fall by infiltration through the limestone outcrops as well as river losses (swallow holes). The recharge cycle appears to be sustainable, with no permanent impact to groundwater resources (Fleury et al., 2009).

Karstic systems are defined by the presence of conduits and fractures in the underlying limestone bedrock, resulting in complex transport networks and variable response times following rainfall events. The main surface processes considered to play a role in runoff generation in the watershed are soil depth and the slope of the hillside. The subsurface processes that contribute to runoff are poorly understood: they may reduce flood intensity by storing water in the epikarst and through deep inflitration (Dörfliger et al., 2008) or they may intensify the flood severity through the contribution of groundwater to peak flow (Kong A Siou et al., 2011). The advanced study and modelling of

water to peak flow (Kong A Siou et al., 2011). The advanced study and modelling of the Lez karstic system is currently the subject of numerous research projects.

### 2.1.2 Climate and rainfall data

The climate of the region is generally dry, with mean annual evapotranspiration (1322 mm at Mauguio for the period 1996 to 2005 – Fig. 1) greater than mean annual rainfall (909 mm at Prades for the period 1992 to 2008)<sup>2</sup>. Most of the yearly rainfall is received in fall and winter in the form of heavy climatic and orographic precipitations. Extreme rainfall events, particularly in fall and late summer periods, are favored in this region due to humidity generated by the warm Mediterranean Sea and a closed cyclone which helps to transport warm, moist air masses to the coast (Nuissier et al., 2008). In September of 2002, rainfall totaled as much as 600–700 mm in certain regions and resulted in destructive floods causing 24 deaths and 1.2 billion euros in economic damages (Boudevillain et al., 2011). The combination of intense autumnal rainfall and small karstic watersheds leads to dangerous flash flood conditions in the

<sup>25</sup> Mediterranean costal region.

<sup>&</sup>lt;sup>2</sup>Mean annual evapotraspiration was calculated using the Penman-Monteith equation; this calculation is not available at the Prades raingauge.



Rainfall in the Lez Catchment is measured by either an S-band radar located in Nîmes at a distance of approximately 65 km from the basin or by a network of 4 rainfall gauges (Prades, Montpellier-ENSAM, Maugio, Saint Martin de Londres – Fig. 1a). Radar data were treated using the HYDRAM algorithm developed by Météo-France for the correction of ground clutters, the vertical profile of reflectivity, and conversion of reflectivity to rainfall using the Marshall-Palmer relationship,

 $Z = 200R^{1.6}$ 

5

(1)

(2)

where Z is the reflectivity in mm<sup>6</sup> m<sup>-3</sup>, R is the radar rainfall intensity in mm h<sup>-1</sup> and 200 and 1.6 are empirical constants derived by the drop size distribution. For the HYDRAM treatment, the same Z-R relationship is used for stratiform and convective 10 rainfall (Tabary, 2007). The Nîmes radar produces scans at three different elevations at 5 min intervals: 2.5° (0–22 km), 1.3° (22–80 km) and 0.6° (distances beyond 80 km). These three different scans are used to produce a composite radar image which describes rainfall for areas at different distances to the radar. The lowest unobstructed scan is selected for a given distance range. For the Lez Catchment, the 1.3° scan was 15 used (Bouilloud et al., 2010) to produce cumulative rainfall depths at a spatial resolution of 1 km<sup>2</sup> and a time step of 5 min. A network of 20 rain gauges within a 50 km range of the catchment provided cumulative rainfall data for adjustments using the MFB (Fig. 1b), a measure of the ratio of radar to rain gauge rainfall during a specified time period (here the length of the flood event): 20

$$\mathsf{MFB} = \frac{\frac{1}{n}\sum_{i}G_{i}}{\frac{1}{n}\sum_{i}R_{i}},$$

25

where  $G_i$  is the rain gauge measurement at location, *i* in mm,  $R_i$  is the radar measurement at the same location in mm and *n* is the number of rain gauges selected. The value of the radar measurement at the gauge location was selected to be the average of the central pixel and its 8 nearest neighbors. The ratio of rain gauge to radar measurements is expected to be greater than 1 for distances between 15 and 80 km from the radar where masking effects play an important role (Cheze and Helloco, 1999).



# 2.1.3 Rainfall events

Table 1 displays the 19 rainfall events measured by HYDRAM-treated radar for the Lez Catchment along with their associated MFB values and peak discharges. In general, events lasted several days and cumulative rainfall was sampled at a time step of 1 h.

The episode MFBs were between 0.87 and 1.80, indicating that radar was never more than 45% away from the "true" rainfall value (assuming absolute confidence in ground measurements) with the exception of November 1997. The very high MFB for this event indicates that either the rain gauges, the radar or both were not functioning properly. With the exception of November 2008, all events have MFB values greater than 1, with an average of 1.39. As mentioned in Sect. 2.1.2, these values are a feature of the distance between the Nîmes radar and the watershed.

Rainfall events were separated into two classes based on their peak discharges: regular events which have a peak discharge greater than  $40 \text{ m}^3 \text{ s}^{-1}$  and very small events which have a peak discharge less than or equal to  $40 \text{ m}^3 \text{ s}^{-1}$ . This classification is used to determine the range of discharges that will be assimilated as discussed in Sect. 5.1.

#### 2.2 The Hydrological model

The hydrological model is event-based, parsimonious and distributed. It operates on independant grid cells, using a SCS-derived runoff production function and a Lag and

Route transfer function. Discharges are calculated using an hourly timestep in order to match the frequency of the observations. The calibration and adaption of this model to the Lez Catchment were undertaken during the doctoral studies of Coustau (2011) and are presented in Coustau et al. (2012).



#### 2.2.1 The runoff production function

The runoff production function is the link between the precipitation falling over the catchment and the discharge emitted to surface waters. Not all rain becomes discharge and processes such as infiltration, evapotranspiration, percolation and intercep-

- tion determine the eventual fate of incident rainfall. The ATHYS software, developed by HydroSciences Montpellier (www.athys-soft.org), permits the use of several runoff production functions, including Green & Ampt, TopModel, Girard, modified SCS and Smith & Parlange. From these options, a derived version of the SCS equations was selected for this study. The SCS model was originally created by engineers at the US
- Soil Conservation Service for use with small agricultural watersheds (less than 8 km<sup>2</sup>). Numerous papers have been published demonstrating the utility of this model for predicting runoff volume (Abon et al., 2011; Sahu et al., 2007; Mishra and Singh, 2004). For predicting the instaneous runoff rate during an event, a derived version of the SCS equation is necessary. The derivation of the SCS function is shown below. The first use of these equations was performed by Gaume et al. (2004). The initial form of the SCS equation is as follows,

$$V(t) = \frac{[P_{\rm b}(t) - I_{\rm a}]^2}{P_{\rm b}(t) - I_{\rm a} + S} A,$$

where *V* is the cumulative runoff volume generated by the watershed at time *t* in  $m^3$ ,  $I_a$  is the initial abstraction in mm,  $P_b$  is the cumulative rainfall depth at time *t* in mm, *S* <sup>20</sup> is the potential storage depth of the watershed at the start of the event (potential maximum retention) in mm and *A* is the catchment area. For this study,  $P_b$  is considered as the level in a cumulative rainfall reservoir. Assuming an initial abstraction equal to 0.2*S* derived from studies of small rural watersheds (USDA Natural Resources Conservation Service, 1986) and dividing by the area of the catchment, Eq. (3) reads,

<sup>25</sup> 
$$P_{\rm e}(t) = \frac{[P_{\rm b}(t) - 0.2S]^2}{P_{\rm b}(t) + 0.8S},$$

Discussion Paper **HESSD** 9, 3527-3579, 2012 Data assimilation for the correction of radar rainfall **Discussion** Paper E. Harader et al. **Title Page** Introduction Abstract Conclusions References **Discussion** Paper Tables **Figures** 14 Back Close **Discussion** Paper Full Screen / Esc **Printer-friendly Version** Interactive Discussion

(3)

(4)

where  $P_{\rm e}(t)$  is the cumulative runoff depth generated by the watershed at time *t* in mm. Taking the time derivative of Eq. (4) leads to the instantaneous runoff rate (or runoff intensity) with units of mm s<sup>-1</sup>, denoted  $i_{\rm e}(t)$ . The time derivative of  $P_{\rm b}(t)$  is the rainfall rate (or rainfall intensity) in mm s<sup>-1</sup>, denoted  $i_{\rm b}(t)$ . The runoff intensity can now be described as a fraction of the rainfall intensity:

$$i_{\rm e}(t) = i_{\rm b}(t) \frac{P_{\rm b}(t) - 0.2S}{P_{\rm b}(t) + 0.8S} \left( 2 - \frac{P_{\rm b}(t) - 0.2S}{P_{\rm b}(t) + 0.8S} \right).$$
(5)

Equation (5) has a form reminiscent of that described in the rational method (Doodge, 1957), which expresses the rate of runoff leaving the watershed outlet as a function of the percentage of area contributing to runoff, C(t), the rainfall rate and the area of the watershed. This method promotes the idea of *saturation excess* (Braud et al., 2010), where the contributing area is the soils which have been saturated and can no longer store water (Dunnian runoff generation). This is in contrast to *infiltration excess* in which the soil's capacity to allow incoming water to infiltrate is overcome and runoff is generated while underlying soils remain unsaturated (Hortonian runoff generation). The equation of the rational method is shown below:

 $Q(t) = C(t)i_{\rm b}(t)A,$ 

5

20

where Q(t) represents the runoff rate in m<sup>3</sup> s<sup>-1</sup>, C(t) is the proportion of the watershed contributing to runoff production,  $i_{b}(t)$  is the rainfall intensity in mm s<sup>-1</sup> and A is the area in m<sup>2</sup>. Although the SCS method does not distinguish between Hortonian and Dunnian processes, thus making the physical meaning of its runoff production different from that of the Rational Method, we can define the proportion of rainfall contributing to runoff, C(t), as follows,

$$C(t) = \begin{cases} \frac{P_{\rm b}(t) - 0.2S}{P_{\rm b}(t) + 0.8S} \left(2 - \frac{P_{\rm b}(t) - 0.2S}{P_{\rm b}(t) + 0.8S}\right) & \text{if } P_{\rm b}(t) > 0.2S \\ 0 & \text{otherwise.} \end{cases}$$

Discussion Paper **HESSD** 9, 3527-3579, 2012 Data assimilation for the correction of radar rainfall Discussion Paper E. Harader et al. **Title Page** Introduction Abstract Conclusions References **Discussion** Paper **Tables Figures** 14 Back Close **Discussion** Paper Full Screen / Esc **Printer-friendly Version** Interactive Discussion

(6)

(7)

The derived SCS method is thus convenient as it allows for the representation of runoff generation as a function of rainfall intensity throughout the event with parameters that can be calibrated using physical measurements. This equation is valid for cumulative rainfalls greater than the initial abstraction. It should also be noted that the runoff coefficient can never be greater than 1, meaning that runoff generation will never exceed rainfall intensity:

 $\lim_{P_{\rm b}(t)\to\infty}C(t)=1.$ 

To represent the ability of the soil to regain part of its absorption potential during pauses in the rainfall, this version of the SCS method allows the soil to drain. The volume of water lost to drainage is a function of two conceptual reservoirs: the cumulative rainfall reservoir, level  $P_{\rm b}(t)$  and the soil reservoir, level stoc(t), shown in Fig. 3. The rate of drainage of the cumulative rainfall reservoir and the soil reservoir is described by:

$$\frac{\mathrm{d}P_{\mathrm{b}}(t)}{\mathrm{d}t} = i_{\mathrm{b}}(t) - ds.P_{\mathrm{b}}(t),$$

15

$$\frac{\mathrm{d}stoc(t)}{\mathrm{d}t} = i_{\mathrm{b}}(t) - i_{\mathrm{e}}(t) - \mathrm{d}s.stoc(t), \tag{10}$$

where ds is the drainage coefficient. The coefficient represents the removal of water through deep infiltration and evapotranspiration during the event and is calculated from the slope of the descending limb of the hydrograph.

The drainage coefficients of the cumulative rainfall reservoir and the soil reservoir were selected to be the same. The water lost to the system by the drainage coefficient is considered to be either lost to deep infiltration or to reemerge as delayed surface runoff,  $i_d(t)$ , calculated by

$$i_{\rm d}(t) = \min(1, \frac{w}{S}).ds.stoc(t),$$

(8)

(9)

(11)

where w is the soil depth coefficient in mm, which is a watershed parameter, calibrated and held constant for all events. The ratio between the capacity of the soil moisture reservoir at the start of the event, S, and the soil depth coefficient determines the fraction of drainage that becomes delayed runoff. As S approaches w (going from high S to low S), the proportion of runoff lost to deep infiltration is diminished and a greater portion of the soil moisture reservoir drainage becomes available as delayed discharge. The drainage coefficient was added by Coustau et al. (2012) in order to adapt the SCS equations to the behavior of karstic watersheds and helps to ensure the proper behaviour of the watershed during the descending limb of the hydrograph by including the participation of subsurface flows.

The total discharge,  $i_t(t)$  is thus:

 $i_{\rm t}(t) = i_{\rm e}(t) + i_{\rm d}(t).$ 

5

10

#### 2.2.2 The transfer function

Supposing that the production function has created runoff at a certain grid location, this runoff must then be transferred to the watershed outlet by what is referred to here as the transfer function. The Lag and Route transfer function is based on a unit hydrograph approach in which the discharge produced by each cell is assumed to follow the form of a Gaussian distribution. In this way, it looks like the impulse solution of the kinematic wave approach. However, in the present case, the form of the hydro-

- graph is assumed and imposed upon the runoff generated by each cell. This runoff is independent and does not interact with that of the other cells. This is in contrast to GIS-based approaches which use the kinematic wave approximation (a simplification of the Saint-Venant equations for shallow water flow) or the Manning equations for open channel flow (Bates and De Roo, 2000). In these two cases, the discharges from
- <sup>25</sup> different cells are allowed to interact and the flow rate will depend upon the depth of the runoff contained within the cell.



(12)

The three parameter Lag and Route transfer function described in Tramblay et al. (2011) was selected for simulating discharges in the Lez Catchment (Fig. 4). The three parameters which describe this function are:  $I_m$  (the length of the flow path from the cell to the outlet – calculated using a method of steepest descent in order to produce drainage paths for each cell),  $V_0$  (the speed of propagation in m s<sup>-1</sup>) and  $K_0$  (a dimensionless coefficient used to calculate the diffusion time).

The first step in the transfer function is the calculation of the propagation time to the outlet,  $T_m$ , which describes the lag between runoff production at time  $t_0$  and the arrival of an associated elementary hydrograph at the watershed outlet:

10 
$$T_{\rm m} = \frac{I_{\rm m}}{V_0}$$
 (13)

From the propagation time, the diffusion time,  $K_m$  can then be calculated. This coefficient represents the velocity distribution of the runoff as it is transferred from the cell to the outlet.  $K_m$  is expressed as

$$K_{\rm m} = K_0 T_{\rm m}. \tag{14}$$

<sup>15</sup> For each grid cell, the diffusion time and propagation time are then used to produce a Gaussian distribution of discharge which represents the elementary hydrograph, q(t) in m<sup>3</sup> s<sup>-1</sup>, produced by a rainfall of intensity  $i_e(t_0)$ :

$$\frac{q(t)}{A} = \begin{cases} 0 \text{ for } t < t_0 + T_m \\ \frac{i_0(t_0)}{K_m} \exp\left(-\frac{t - (t_0 + T_m)}{K_m}\right) \text{ for } t \ge t_0 + T_m, \end{cases}$$
(15)

where A is the area of the grid cell in  $m^2$ .

To measure the quality of the simulations performed by the hydrological model, the Nash-Sutcliff efficiency criterion (IE) was selected (Montanari et al., 2009). This criterion can be expressed as a function of the error between the model discharge at time j ( $Q_{\text{sim},j}$  in m<sup>3</sup> s<sup>-1</sup>) and measured discharge at time j ( $Q_{\text{obs},j}$  in m<sup>3</sup> s<sup>-1</sup>), summed over



20

5

*j*, squared, and normalized by the variance of the measured discharge  $(\sigma_{obs}^2)$ :

$$\mathsf{IE}(\%) = 100 \left( 1 - \frac{\sum_{j} (Q_{\mathsf{sim},j} - Q_{\mathsf{obs},j})^2}{\sigma_{\mathsf{obs}}^2} \right)$$

where *j* varies from 1 to *N*, the total number of observations available for the event. For this study, IE is calculated over the entire length of the rainfall event, regardless of the number of observations assimilated. The window of observations selected for assimilation will be discussed in detail in Sect. 3.

A second measure of quality is the normalized difference in peak flow between the simulation ( $Q_{sim,peak}$ ) and the observations ( $Q_{obs,peak}$ ), PH:

$$\mathsf{PH} = \frac{Q_{\mathsf{sim},\mathsf{peak}} - Q_{\mathsf{obs},\mathsf{peak}}}{Q_{\mathsf{obs},\mathsf{peak}}}.$$
(17)

#### 10 2.2.3 Sensitivity of the model to rainfall inputs and parameterisation

Rainfall plays a key role in the estimation of discharges using hydrological models. The model used in this study is sensitive to the quantity and intensity of rainfall and this sensitivity varies depending on previous conditions. As the soil reservoir becomes saturated, a greater proportion of incident rainfall runs off and is emitted as discharge.

<sup>15</sup> To illustrate this phenomena, a linear multiplier of the rainfall intensity, denoted  $\alpha$ , was introduced into the model:

$$i_{\mathsf{b}}(t) = \alpha \; i_{\mathsf{b}}^{\star}(t),$$

where  $i_{b}^{*}$  is the observed radar rainfall and  $i_{b}$  is the rainfall used by the model. Figure 5 displays the discharge as a function of  $\alpha$  at 3 h before the flood peak. This timestep was selected because it demonstrates saturated behaviour for larger values of  $\alpha$  and non-saturated behaviour for small  $\alpha$ . The nonlinear relationship between flowrate and  $\alpha$  is due to (i) a nonlinear runoff production function which depends on soil saturation and



(16)

(18)

(ii) the differential equations describing soil and rainfall reservoir drainage. Because of the strong influence of the rainfall input upon model results,  $\alpha$  was chosen as the target of the DA procedure.

In the work of Coustau et al. (2012), the model was tested for its sensitivity to the parameterisation selected. *S* (the potential storage depth at the start of the event) and  $V_0$  (the velocity of transfer) were identified as key parameters. *S* was shown to have an important effect on the amplitude of the flood peak, while  $V_0$  influenced the arrival time. The calibration of *S* and  $V_0$  for this study will be discussed in Sect. 4.

#### 3 Data assimilation methods

25

J

- <sup>10</sup> The conceptual hydrological model used in this study can be represented as a nonlinear operator,  $\mathcal{H}$ , as in Fig. 6. This operator translates rainfall input data  $i_{\rm b}$  into discharge data  $Q_{\rm sim}$ , using model parameters (such as *S* and  $V_0$ ) to solve ordinary differential equations for the state variables, *stoc* and  $P_{\rm b}$ . Each of the parameters, inputs, state variables and outputs of this model is uncertain and could potentially be <sup>15</sup> corrected with a DA algorithm using observed data. The output from the background
- simulation (background discharge before assimilation) is shown as  $Q_{sim,b}$ , the analysis simulation (analysis discharge after data assimilation) is shown as  $Q_{sim,a}$ , and the observations used to produce the analysis are  $Q_{obs}$ . In this study, the assimilation algorithm was applied to correct the radar rainfall,  $i_b$  input to the model using discharge

<sup>20</sup> observations. The correction, shown in Fig. 6, is a multiplicative coefficient of the rainfall intensity, denoted  $\alpha$ , presented in Sect. 2.2.3. This coefficient is contained within the control vector,  $\mathbf{x} = (\alpha)$ .

Assuming that the errors in the model parameters and the observations follow a Gaussian distribution, the optimal value of the control vector is the analysis,  $x^{a}$ , which minimizes the cost function *J*:

$$\mathbf{Y}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^{\mathrm{b}})^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^{\mathrm{b}}) + (\mathbf{y}^{\mathrm{o}} - \mathcal{H}(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y}^{\mathrm{o}} - \mathcal{H}(\mathbf{x})).$$
(19)



The cost function, *J* expresses the difference between the control vector, *x* and its a priori value,  $x^{b}$  (the background), and the difference between the control vector in the observation space and the observation vector,  $y^{o}$ , weighted respectively by the background and observation error covariance matrices, **B** and **R**. The background control vector is selected as  $x_{b} = (1)$  (no change to the input rainfall). The observation vector gathers together the observed discharges during the flood event.

Both  $x^{b}$  and  $y^{o}$  are considered to deviate from the "true" state of the system,  $x^{t}$ . The difference between each of  $x^{b}$ ,  $x^{a}$  and  $y^{o}$  and the true state,  $x^{t}$  are the background, analysis and observation errors denoted  $\epsilon_{b}$ ,  $\epsilon_{a}$  and  $\epsilon_{o}$ . These errors are assumed to be unbiased and uncorrelated.

In Eq. (19), the observation operator  $\mathcal{H}$  is non-linear as it represents the integration of the hydrological model for a given radar rainfall data set. As a consequence, J is non-quadratic and an incremental approach is used to approximate  $x^a$  such that  $\nabla J(x^a) = 0$ . The increment,  $\delta x$ , is defined as:

15  $\delta \mathbf{X} = \mathbf{X} - \mathbf{X}^{\mathsf{b}}$ .

5

(20)

Discussion Paper

**Discussion** Paper

Discussion Paper

Discussion Paper

(24)

The Jacobian matrix **H** of the observation operator  $\mathcal{H}$  is expressed using the Taylor expansion computed around a reference vector  $\mathbf{x}^{ref}$ , initially chosen as  $\mathbf{x}^{b}$ :

$$\mathcal{H}(\mathbf{x}^{\mathrm{b}} + \delta \mathbf{x}) \approx \mathcal{H}(\mathbf{x}^{\mathrm{b}}) + \frac{\partial \mathcal{H}}{\partial \mathbf{x}}|_{\mathbf{x}^{\mathrm{b}}}(\delta \mathbf{x}).$$
 (21)

Using Eq. (21), Eq. (19) reads:

$$J_{\rm inc}(\delta \mathbf{x}) = \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + (\mathbf{y}^{\rm o} - \mathcal{H}(\mathbf{x}^{\rm b}) - \mathbf{H} \delta \mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y}^{\rm o} - \mathcal{H}(\mathbf{x}^{\rm b}) - \mathbf{H} \delta \mathbf{x}).$$
(22)

Taking the gradient of  $J_{inc}$  leads to,

$$\nabla J_{\text{inc}}(\delta \mathbf{x}) = \mathbf{B}^{-1} \delta \mathbf{x} - \mathbf{H}^T \mathbf{R}^{-1} (\underbrace{\mathbf{y}^{\circ} - \mathcal{H}(\mathbf{x}^{\circ})}_{d} - \mathbf{H} \delta \mathbf{x}),$$
(23)

where *d* is the innovation vector.  $J_{inc}$  is at a minumum when  $\nabla J_{inc}$  is null. By setting Eq. (23) equal to 0 and solving for  $\delta x$ , the following solution is generated:

 $\delta x^{a} = \mathbf{K} d$ 

HESSD

9, 3527-3579, 2012

Using Eq. (20),  $\delta x^{a}$  is solved for  $x^{a}$ :

 $\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{b}} + \mathbf{K}\boldsymbol{d},$ 

where  $x^{a}$  is the Kalman Filter analysis and **K** is the gain matrix described by,

 $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}.$ 

<sup>5</sup> The use of the Kalman Filter analysis equations relies on the hypothesis that  $\mathcal{H}(\mathbf{x})$  is linear on  $[\mathbf{x}^{a}, \mathbf{x}^{ref}]$ , where  $\mathbf{x}_{a}$  is the minimum of  $J_{inc}$ , the quadratic approximation of J. To compensate for non-linearities in  $\mathcal{H}(\mathbf{x})$ , the outer loop process used in this study allows for the recalculation of the linear tangent, **H** at the location of the analysis of the previous iteration( $\mathbf{x}^{a}$ ) in order to create a new quadratic approximation of J, as shown in Fig. 7.

The formulation of the Kalman Filter derived above presents several key simplifications when compared to other work. The control vector, often used to correct model states, contains a single constant parameter that is applied to the model inputs, ( $\alpha$ ). If applicable in this case, the state vector would include ( $stoc_{i,j}(t_k)$ ,  $P_{b,i,j}(t_k)$ ), where *i* and *i* represent the x and y coordinates of each point and *t* is comparison to be a state.

and *j* represent the x and y coordinates of each point and  $t_k$  is some time step, k. Since the control vector contains only  $\alpha$  and is assumed constant in time, its background error covariance matrix is not propagated in time. The analysis is then determined over the entire assimilation window.

The background error is assumed to follow a Gaussian model and is described by its variance as the control vector is a scalar. The background error variance is set to 40 % of  $x_b$  squared. The value of 40 % was determined by taking the average of  $|1 - MFB_j|$ , thus relating **B** to the average rainfall correction.

The observation errors are supposed uncorrelated, making **R** a diagonal matrix. The observation error variance,  $\sigma_{obs}^2$ , is chosen to be inversely proportional to the discharge value, thus favoring larger discharge observations. A proportionality coefficient,  $\beta_{obs}$  was introduced in order to control the observation error variance depending on the



(25)

(26)

assimilation window selected (reanalysis or forecast mode).

$$\sigma_{\text{obs},i}^{2} = \max\left((\frac{\beta_{\text{obs}}}{Q_{i}})^{2}, 0.01\right) \text{ for } i = t_{i}: t_{\text{f}}.$$
 (27)

**R** has a lower bound of  $0.01 \text{ m}^6 \text{ s}^{-2}$  and no upper bound. As the errors coming from each source of information are not precisely known, different values of the proportionality coefficient were considered as described in Sect. 5.1.

The data assimilation algorithm described above is applied in 2 modes: reanalysis and forecast. In reanalysis, all valid observations during the rainfall event are assimilated and in forecast, the assimilation window ends 3 h before the flood peak. Beyond this time, the model is integrated to provide a forecast of the event peak.

#### **4** Initialisation and calibration of the model

5

The hydrological model contains several types of parameters: watershed constants which can be calibrated using batch calibration procedures, mathematical properties of the equations and one initial condition of the watershed, S, which must be calibrated separately for each event. The watershed constants and the initial condition are calibrated by selecting the value which maximises the IE of the simulated discharge for a given event. In the case of S, the calibration process is finished at this stage, while batch parameters are averaged over all events. It should be noted that while the language "initialisation" is used here, S is a parameter in this data assimilation system and not a model state, thus it does not evolve during the event.

#### 20 4.1 Calibration of model parameters

During the calibration process, a mixture of ground rainfall events and high quality (early autumn) radar rainfall events was used in order to minimize the error associated with the rainfall in the model parameters. The 2 batch-calibrated model parameters:  $V_0$  and w were calculated using past rainfall events from 1994–2008 for which



either ground rainfall or high quality radar rainfall data and discharges were available. The drainage coefficient, ds, is a mathematical property of the hydrological model equal to the coefficient of the exponential recession limb of the hydrograph. This is a constant for all episodes. Since the diffusion time,  $K_m$ , is a function of both  $V_0$  and  $\kappa_0$ , many values of these two parameters can result in the same velocity distribution

 $_{5}$   $K_{0}$ , many values of these two parameters can result in the same velocity distribution at the watershed outlet. To avoid problems of equifinality,  $K_{0}$  was set as a fixed value before calibrating  $V_{0}$ . The parameters  $V_{0}$ , w, ds and  $K_{0}$  were set as  $1.3 \text{ m s}^{-1}$ , 101 mm,  $0.28 \text{ d}^{-1}$  and 0.3 (dimensionless) respectively for all events.

# 4.2 Initialisation of S

- <sup>10</sup> The soil moisture deficit at the start of the event, represented by the parameter, *S*, must be calculated at the beginning of each event. In reanalysis mode, a posteriori *S* values, hereafter referred to as  $S_{cal}$ , were calibrated for each episode of the 19 radar rainfall episodes using an objective function which maximises the IE value for simulations forced with the MFB corrected radar rainfall in order to minimise the error in the
- <sup>15</sup> parameterisation. In forecast mode, the event hydrograph is not known. As a consequence, *S* must be estimated at the start of the event using known indicators of the soil hydric state at this time. For example, piezometric readings could be used to estimate the hydric state of the watershed in the morning if heavy rain is predicted for the evening. In this study, a calibration curve relating *S* to indicators of the catchment wet-<sup>20</sup> ness state is used to estimate a priori *S* values for each episode from measurements
- of aquifer piezometry or soil moisture indicators derived from surface models (Coustau et al., 2012). These estimated S values are referred to as  $S_{reg}$ .

Using the historical record of discharge and rainfall from 1997–2008, calibration curves for *S* were developed using 3 catchment wetness state indicators: Hu2 (%), the piezometer located at Bois Saint Mathieu (m) and the piezometer located at Claret (m). These two piezometers were selected for the quality of their relation to the hy-



dric state of the watershed. The Hu2 indicator is modeled by Météo-France (Quintana

Seguí, 2008) and estimates the % soil saturation at the root horizon. The measurements for each event are taken as the value of the indicator at 06:00 a.m. the day of the event. Hu2 data are available for 18 of the 19 events and piezometer data are available for 14 of the 19 events.

For each indicator, a regression of slope *M* and y-intercept *b* was formed using the indicator as the independant variable and  $S_{cal}$  as the dependant variable as shown in Table 2.  $R^2$  is the coefficient of determination for the linear regression between  $S_{cal}$  and the physical indicators. To validate each regression, split sample tests were performed. Each regression was performed using only the first half of the data available to construct a "historical period"; the  $S_{reg}$  values calculated using the validation regression were then compared with the  $S_{reg}$  values calculated using the regression for the entire record. The average and standard deviation of the % difference between these two  $S_{reg}$  values are presented in Table 2. The average % difference between  $S_{reg}$  for the validation curve and  $S_{cal}$  during the validation period was 0.22, 0.21 and 0.16 for Bois Saint Mathieu, Claret, and Hu2 respectively. When considering the difference between  $S_{reg}$  for the validation curve, the piezometer at Claret was the most robust indicator during validation. The validation curve  $S_{reg}$  values for Hu2 were closest to  $S_{cal}$  during the validation period.

The  $S_{reg}$  values calculated using the different regressions are shown in Table 3. The starred entries in this table represent events for which either indicator data was not available or data assimilation was not preformed (October 2008). An analysis of the impact of errors in the parameterisation is presented in Sect. 5.3.1.

#### 5 Application of data assimilation for the correction of radar rainfall

This section first explores the method of data assimilation applied in 2 modes: reanalysis and pseudo-forecast. In Sect. 5.2, the results of the reanalysis mode are discussed, followed by a presentation of the pseudo-forecast mode results in Sect. 5.3.



As mentioned in Sect. 3, the assimilation window ends 3h before the flood peak in pseudo-forecast mode. This choice of assimilation window is intended to demonstrate the possible performance of the algorithm in a real-time forecasting environment, while acknowledging that the arrival time would not be known in this case. A complementary approach would be a sliding window, reflecting increased knowledge of the event as time goes on.

#### 5.1 Illustration of the data assimilation procedure

5

For the 19 radar rainfall events, the range of assimilated discharges is 15–300 m<sup>3</sup> s<sup>-1</sup> for normal episodes and 2–40 m<sup>3</sup> s<sup>-1</sup> for very small episodes (peak discharge less
than or equal to 40 m<sup>3</sup> s<sup>-1</sup>). Very large discharges are unrealiable due to the use of a rating curve to calculate the river stage-discharge relationship beyond 300 m<sup>3</sup> s<sup>-1</sup>. Small discharges are eliminated in order to better represent the flood behavior of the watershed. For each calculation of the analysis control vector in both analysis and forecast modes, 5 iterations of the outer loop method were used.

Episodes with notable double peaks (September 2002, October 2002, December 2002, September 2005 and October 2008) were separated into single peaks prior to assimilation due to the inability of the hydrological model to properly represent multiple peaks in succession. Data assimilation was applied to all episodes in both forecast and reanalysis mode, with the exclusion of October 2008. The rising limb of this event takes place
 over a period of time less than three hours long, thus no discharge measurements are assimilated in forecast mode.

To illustrate the procedure in the two different modes, the episode of November 2008 was selected. In renalysis mode, the soil moisture parameter,  $S_{cal}$  is 142 mm.  $\beta_{obs}$  is chosen to be  $0.25 \text{ m}^6 \text{ s}^{-2}$  in order to reflect an almost complete confidence in the observations. As shown in Fig. 8, the IE is improved from -0.52 to 0.72 following assimilation. Observations are in blue, the background simulation in pink and the analysis simulation in green. The hyetogram is on the inverted y-axis: initial rainfall is in dark



blue and the corrected rainfall is in light blue with each bar the width of a 1 h timestep. This color scheme is conserved throughout the paper. In this case,  $\alpha = 0.70$  for the analysis, meaning that the optimal state of the rainfall is less than that predicted by the uncorrected radar data. This reduced rainfall then results in an analysis hydrograph that is less than the background hydrograph.

5

10

In forecast mode, observations are assimilated from the start of the event until 3 h before the peak discharge. This process is illustrated in Fig. 9 for the forecast mode of November 2008; the first analysis and the final result of the outer loop are shown. For this demonstration, S and  $\beta_{obs}$  were kept the same as those for the reanalysis mode. The first iterate of the outer loop (black) overestimates the correction to the reinfall for a majority of the period. This guarantimation is corrected with

- rainfall for a majority of the assimilation period. This overestimation is corrected with subsequent iterates. The first iterate has the best IE with 0.71 which is nearly equal to that of the reanalysis mode. However, this is not the optimal state for the assimilation period (up to 3 h before peak flow). Following new estimations of the Jacobian matrix,
- <sup>15</sup> **H**, at the analysis location, the final IE after all iterations of the outer loop is 0.62. The final  $\alpha$  was 0.61, suggesting that the algorithm underestimates the rainfall in forecast mode. The analysis hydrograph (green curve) is still improved over the background hydrograph (pink curve), as it reduces the amount of rainfall; however, the reduction is overestimated when only the start of the episode is assimilated.
- It should be noted that several modifications to the assimilation procedure are necessary to optimally assimilate data in the forecast mode. First of all, an a priori estimation of S ( $S_{reg}$ ), as presented in Sect. 4, is required. Secondly, the observation error covariance must be adjusted to reflect representativeness errors due to the size of the assimilation window (only the start of the event is known). To account for increased errors in the observations,  $\beta_{obs}$  should be increased in the forecast mode. Trials using different values of  $\beta_{obs}$  are presented in Sect. 5.3.1.



#### 5.2 Reanalysis mode results

### 5.2.1 Impact of the rainfall correction

For the reanalysis mode, results are compared to the background state simulations and then to simulations with the MFB-corrected radar rainfall. Figure 10 presents the change in IE for the reanalysis compared with the background state for 19 episodes with 7 additional peaks due to separation of multi-peak episodes. The IE values for simulations using uncorrected radar rainfall (the background simulation) are poor and in most cases are not of sufficient quality to reproduce the flood event. The quality of the simulation is greatly improved following data assimilation and IE values are between 0.5 and 1 for a majority of episodes. The only degraded episode is that of December 2003; this deterioration is related to the 300 m<sup>3</sup> s<sup>-1</sup> upper assimilation limit described in Sect. 5.1 and is discussed in greater detail in Sect. 5.2.3.

#### 5.2.2 Comparison of data assimilation to the MFB correction

A linear regression was performed between the MFB and  $\alpha$  as shown in Fig. 11. The <sup>15</sup> two quantities are expected to be related as they both represent corrections to the same rainfall. If errors due to other sources are minimized (parameterisations, measurement errors for the gauges and discharge), the two corrective factors should tend towards the same value. The two quantities are well correlated with a  $R^2$  equal to 0.77. The slope, however, is 1.12, which suggests a systematic underestimation of rainfall by the <sup>20</sup> MFB correction if  $\alpha$  is considered to be the optimal state.

The difference between the simulated discharges resulting from the rainfall corrected by the DA procedure and the MFB correction is presented in Fig. 12. The change in PH was calculated as  $PH_{MFB} - PH_{\alpha}$ ; positive results are thus increases in the positive y-axis. 78% of episodes showed an improved IE and 81% of episodes showed an improved PH compared to the MFB correction. The average improvement in IE was +0.23 versus -0.20 for PH (improvements are negative for PH which has an optimal



value of 0). When deteriorations in the IE occured, they have the tendancy to be small, (-0.01 to -0.06). Deteriorations in the PH have a much larger range (+0.02 to 0.21). In most cases,  $\alpha$  provides improved results over the MFB correction. However, some of the improvement in the simulations with  $\alpha$  when considering double peaks may be due to an increased time resolution. The MFB was calculated using rainfall over the entire event, whereas the events were seperated into single peaks when using  $\alpha$ . The MFB is also calculated over a much larger spatial extent than that of the physical basin, leading perhaps to representativeness errors.

# 5.2.3 Limitations

- <sup>10</sup> The quality of the December 2003 simulation (Fig. 13a) was degraded following data assimilation when compared to the background state. This is the result of a nonmonotonic error in the discharge during the episode, as seen in Fig. 13b. Positive errors in the rising and descending limbs of the hydrograph result in an analysis state with a reduced rainfall. However, the sign of the error in the region near the peak is negative
- <sup>15</sup> and this part of the hydrograph is not well-represented. To counteract this problem, the upper limit of assimilated observations can be increased to include more observations at the hydrograph peak. The inclusion of these points increases the number of negative errors taken into account by the algorithm and results in an analysis which decreases rainfall less than when discharge observations are limited to less than 300 m<sup>3</sup> s<sup>-1</sup>.

# 20 5.3 Forecast mode results

25

# 5.3.1 Sources of uncertainty

In forecast mode, the efficacity of the DA algorithm is affected by both a lack of information about the event (representativeness errors) and a poor parameterisation compared to the a posteriori *S* values ( $S_{cal}$ ). In this study, representativeness errors refer to the fact that the start of the event may not be indicative of what comes later. For example,



a gently sloped rising limb may then be followed by a large and quickly-rising peak flow. In this case, the algorithm would miss the peak region if it was to match observations at the start of the event as closely as possible. With regards to the parameterisation,  $S_{cal}$  requires a fully described event to be calculated, thus in forcast mode, *S* must be estimated by physical indicators of catchment wetness state at the start of an event using the regression curves described in Sect. 4.2. To select the highest quality regression, the assimilation results for 3 different catchment wetness state indicators will be considered in Sect. 5.3.2.

To compare the effects of the two sources of uncertainty discussed above, IE and <sup>10</sup> PH values were compared for simulations using analyses calculated in forecast mode with (1) parameterisation using  $S_{Hu2}$  ( $\beta_{obs} = 0.25 \text{ m}^6 \text{ s}^{-2}$ ) and (2) different values of the **R** matrix ( $\beta_{obs} = 0.25 \text{ m}^6 \text{ s}^{-2}$ ,  $25 \text{ m}^6 \text{ s}^{-2}$  and  $250 \text{ m}^6 \text{ s}^{-2}$ ) and  $S_{cal}$ . The range of  $\beta_{obs}$  values helps to estimate the uncertainty coming from the observations, while the comparison of the data assimilation results using  $S_{Hu2}$  and  $S_{cal}$  gives an idea of the <sup>15</sup> uncertainty resulting from the parameterisation.

Figure 14a presents a box plot of the change in IE for the 4 cases and Fig. 14b presents the results for the PH. The error in the parameterisation affects the median, as seen by the decreased median for the simulations using  $S_{Hu2}$ , while the representativeness error affects the spread of the results. The quality of improvements possible in forecast mode is thus affected by the physical parameter selected to calculate  $S_{reg}$ 

<sup>20</sup> In forecast mode is thus affected by the physical parameter selected to calcula and the range of those improvements is controlled by  $\beta_{obs}$ .

#### 5.3.2 Results for 3 different soil moisture parameterizations

Figure 15 presents boxplots of the improvements to IE and PH values for the three different *S* parameterizations.  $\beta_{obs}$  is selected as 250 m<sup>6</sup> s<sup>-2</sup> in order to limit the amount of confidence placed in the observations when the event is incompletely described (the assimilation window covers only the start of the event). This value of  $\beta_{obs}$  still results in a positive average improvement in PH and IE, but it has a tighter distribution (Fig. 14) which results in smaller deteriorations for cases in which data assimilation deteriorates



the simulation quality. Examination of the boxplot reveals that Bois Saint Mathieu and Claret both have improved median IE improvement compared to Hu2. The spreads of Bois Saint Mathieu and Claret improvements are similar. The medians of each of the three catchment wetness state indicators are similar for Claret, Bois Saint Mathieu

- <sup>5</sup> and Hu2, though Claret has the narrowest spread (but also several negative outliers). The IE was improved by an average of 0.23, 0.31 and 0.16 for Bois Saint Mathieu, Claret and Hu2 respectively, compared to 0.40 for  $S_{cal}$ . The PH was improved by an average of 0.07, 0.04 and 0.07 for Bois Saint Mathieu, Claret and Hu2 respectively, compared to 0.14 for  $S_{cal}$ . The IE results are more positive than the PH results since
- IE takes into account the assimilation and forecast periods. In addition, it should be noted that the DA algorithm seeks to reduce the distance between the observed and simulated hydrographs as a whole and not simply at the peak region, thus it is not expected that DA will always improve peak criteria. For the IE criteria, 67 %, 71 % and 67 % of episodes were improved by DA using the Claret, Bois Saint Mathieu and Hu2 parameterisations respectively. For the PH criteria, 67 %, 62 % and 64 % of episodes were improved by DA using the Claret Bois Saint Mathieu and Hu2 parameterisations.
- were improved by DA using the Claret, Bois Saint Mathieu and Hu2 parameterisations respectively. Regressions were performed between the  $\alpha$  values for each indicator and the MFB.

 $R^2$  values are presented in Table 4. The variables were less well correlated than for the renalysis mode due representativeness and parameterisation errors. Claret had an improved MFB- $\alpha$  correalation compared to the other two soil moisture indicators.

#### 6 Summary and conclusions

25

A simplified Kalman Filter algorithm was implemented on top of a distributed, eventbased parsimonious rainfall-runoff model. Discharges observed at the catchment outlet were assimilated in order to correct input radar rainfall using a multiplicative coefficient ( $\alpha$ ) held constant during a given event. In reanalysis mode, the DA algorithm is capable of finding an optimal control vector that produces simulations improved over those



produced by the MFB for most episodes, given an appropriate parameterization. These corrections are well correlated with MFB values.

The second part of this study focused on the correction of rainfall inputs in a pseudoforecast mode. For both the IE and the PH criteria, over 60% of episodes were im-<sup>5</sup> proved following data assimilation. Average improvement in the IE was notable, while that of the PH was near 0. These results were subject to representativeness and parameterisation errors which diminished the efficacity of data assimilation. A sliding assimilation window or an autoregressive update function may be necessary to improve the analysis guality in the forecast mode. The use of a sliding window to calculate

 $\alpha$ , with comparisons made to MFB values calculated with the same temporal resolution would be particularily interesting. Using a distributed  $\alpha$  value is another possible approach, given the sensitivity of radar measurements to distance. From a prevision standpoint, testing modelled future rainfall with this algorithm is essential for judging its utility for operational flood forecasting. At the present time, modelled rainfall is not available at a suitable temporal resolution for this region.

Futher research is also necessary to adapt this technique for other types of models and floods. This case relates to a conceptual model used for flash flooding events, but physically-based models may prove to be more robust in forecast environments when sufficient data is available on the watershed. Floods based on phenomena which take place at a longer timescale may also lead to different results.

In spite of certain limitations of this assimilation system in the forecast mode, it shows great promise for the correction of radar rainfall in the context of event reanalysis. For basins that have available radar rainfall, but scarce or inaccurate ground rainfall measurements, discharge measurements can serve as a replacement for the MFB

<sup>25</sup> correction using an appropriate hydrological model and assimilation procedure.

20





The publication of this article is financed by CNRS-INSU.

#### References

Abon, C. C., David, C. P. C., and Pellejera, N. E. B.: Reconstructing the Tropical Storm Ket-

- sana flood event in Marikina River, Philippines, Hydrol. Earth Syst. Sci., 15, 1283–1289, doi:10.5194/hess-15-1283-2011, 2011. 3537
  - Anquetin, S., Braud, I., Vannier, O., Viallet, P., Boudevillain, B., Creutin, J.-D., and Manus, C.: Sensitivity of the hydrological response to the variability of rainfall fields and soils for the Gard 2002 flash-flood event, J. Hydrol., 394, 134–147, 2010. 3529
- <sup>10</sup> Aubert, D., Loumagne, C., and Oudin, L.: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model, J. Hydrol., 280, 145–161, 2003. 3530
  - Avias, J.: Karstic aquifer of mediterranean type, geological controls: Lez spring (North Montpellieran karsts) example, IAH, Hydrogeology of selected karst regions, 13, 89–113, 1992. 3533
- <sup>15</sup> Bates, P. and De Roo, A.: A simple raster-based model for flood inundation simulation, J. Hydrol., 236, 54–77, 2000. 3540
  - Borga, M.: Accuracy of radar rainfall estimates for streamflow simulation, J. Hydrol., 267, 26– 39, 2002. 3529

Boudevillain, B., Delrieu, G., Galabertier, B., Bonnifait, L., Bouilloud, L., Kirstetter, P.-E.,

- <sup>20</sup> and Mosini, M.-L.: The Cévennes-Vivarais Mediterranean Hydrometeorological Observatory database, Water Resour. Res., 47, W07701, doi:10.1029/2010WR010353, 2011. 3534
  - Bouilloud, L., Delrieu, G., Boudevillain, B., and Kirstetter, P.-E.: Radar rainfall estimation in the context of post-event analysis of flash-flood events, J. Hydrol., 394, 17–27, 2010. 3535
- Bouttier, F. and Courtier, P.: Data assimilation concepts and methods, in: ECMWF Training Course, 1999. 3531



Bouvier, C. and Delclaux, F.: ATHYS : a hydrological environment for spatial modelling and coupling with a GIS, in: Proceedings of the HydroGIS96 conference held in Vienna, Austria, IAHS publication no. 235, 1996. 3567, 3568

Braud, I., Roux, H., Anquetin, S., Maubourguet, M.-M., Manus, C., Viallet, P., and Dartus, D.:

- <sup>5</sup> The use of distributed hydrological models for the Gard 2002 flash flood event: Analysis of associated hydrological processes, J. Hydrol., 394, 162–181, 2010. 3538
  - Cheze, J.-L. and Helloco, F.: The HYDRAM Project of Météo-France: a quality-based approach for quantitative radar measurements, in: 29th Conference on Radar Meteorology, Montréal, Québec (Canada), 1999. 3530, 3535
- <sup>10</sup> Chumchean, S., Seed, A., and Sharma, A.: Correcting of real-time radar rainfall bias using a Kalman filtering approach, J. Hydrol., 317, 123–137, 2006. 3531
  - Cole, S. and Moore, R.: Hydrological modelling using raingauge- and radar-based estimators of areal rainfall, J. Hydrol., 358, 159–181, 2008. 3530

Coustau, M.: Contribution à la prévision des crues sur le bassin du Lez: modélisation de la

- relation pluie-débit en zone karstique et impact de l'assimilation de données, Ph.D. thesis, Université de Montpellier II, 2011. 3536, 3566
  - Coustau, M., Estupina-Borrell, V., and Bouvier, C.: Improvement of rainfall-runoff modelling with distributed radar rainfall data: a case study in the Lez French Mediterranean Catchment, in: Proceedings of the Weather Radar and Hydrology symposium, held in Exeter, UK. IAHS Publ. 351–55, 2011. 3530
  - Coustau, M., Bouvier, C., Estupina-Borrell, V., and Jourde, H.: Flood modelling with a distributed event-base parsimonious rainfall-runoff model: case of the Lez river catchment, Nat. Hazards Earth Syst. Sci., accepted, 2012. 3536, 3540, 3543, 3547

20

Crow, W.: Correcting Land Surface Model Predictions for the Impact of Temporally Sparse

- Rainfall Rate Measurements Using an Ensemble Kalman Filter and Surface Brightness Temperature Observations, J. Hydrometeorol., 4, 960–973, 2003. 3530
  - Doodge, J.: The rational method for estimating flood peaks, Engineering, 184, 311–313, 374– 377, 1957. 3538

Dörfliger, N., Jourde, H., Ladouche, B., Fleury, P., Lachassagne, P., Conroux, Y., Pistre, S., and

<sup>30</sup> Vestier, A.: Active water management resources of karstic water catchement: the example of Le Lez spring (Montpellier, Southern France), in: Proceedings of the World Water Congress held in Montpellier, Southern France., 2008. 3534

Emmanuel, I., Andrieu, H., and Tabary, P.: Evaluation of the new French operational



weather radar product for the field of urban hydrology, Atmos. Res., 103, 20–32, doi:10.1016/j.atmosres.2011.06.018, 2012. 3530

- Estupina-Borrell, V.: Vers une modélisation hydrologique adaptée à la prévision opérationelle des crues éclair. Application sur de petits bassins versants du Sud de la France., Ph.D. thesis, Insitut National Polytechnique de Toulouse, 2004. 3529
- thesis, Insitut National Polytechnique de Toulouse, 2004. 3529 Fleury, P., Ladouche, B., Conroux, Y., Jourde, H., and Dörfliger, N.: Modelling the hydrological functions of a karst aquifer under active water management, J. Hydrol., 365, 235–243, 2009. 3533, 3534

Gaume, E., Livet, M., Desbordes, M., and Villeneuve, J.-P.: Hydrological analysis of the river

- Aude, France, flash flood on 12 and 13 November 1999, J. Hydrol., 286, 135–154, 2004. 3537
  - Goegebeur, M. and Pauwels, V.: Improvement of the PEST parameter estimation algorithm through Extended Kalman Filtering, J. Hydrol., 337, 436–451, 2007. 3531
  - Kahl, B. and Nachtnebel, H.: Online updating procedures for a real-time hydrological
- forecasting system, IOP Conf. Series: Earth and Environmental Science, 4, 012001, doi:10.1088/1755-1307/4/1/012001, 2008. 3530
  - Kong A Siou, L., Johannet, A., Borrell, V., and Pistre, S.: Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France), J. Hydrol., 403, 367–380, 2011. 3534
- McLaughlin, D. and Townley, L.: A reassessment of the groundwater inverse problem, Water Resour. Res., 32, 1131–1161, 1996. 3530
  - Meier, P., Frömelt, A., and Kinzelbach, W.: Hydrological real-time modelling in the Zambezi river basin using satellite-based soil moisture and rainfall data, Hydrol. Earth Syst. Sci., 15, 999–1008, doi:10.5194/hess-15-999-2011, 2011. 3531
- <sup>25</sup> Mishra, S. and Singh, V.: Validity and extension of the SCS-CN method for computing infiltration and rainfall-excess rates, Hydrol. Process., 18, 3323–3345, 2004. 3537
  - Montanari, M., Hostache, R., Matgen, P., Schumann, G., Pfister, L., and Hoffmann, L.: Calibration and sequential updating of a coupled hydrologic-hydraulic model using remote sensingderived water stages, Hydrol. Earth Syst. Sci., 13, 367–380, doi:10.5194/hess-13-367-2009, 2000. 2541
- 30 2009. 3541
  - Moradkhani, H., Sorooshian, S., Gupta, H., and Houser, P.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter, Adv. Water Resour., 28, 135–147, 2005. 3530



- Nuissier, O., Ducrocq, V., Ricard, D., Lebeaupin, C., and Anquetin, S.: A numerical study of three catastrophic precipitating events over southern France. I: Numerical framework and synoptic ingredients, Q. J. Roy. Meteorol. Soc., 134, 111–130, 2008. 3534
- Pauwels, V. and De Lannoy, G.: Ensemble-based assimilation of discharge into rainfall-runoff
- <sup>5</sup> models: A comparison of approaches to mapping observational information to state space, Water Resour. Res., 45, 1–17, 2009. 3531
  - Pauwels, V., Hoeben, R., Verhoest, N., and Troch, F. D.: The importance of spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation, J. Hydrol., 251, 88–102, 2001. 3530
- <sup>10</sup> Reichle, R., McLaughlin, D., and Entekhabi, D.: Hydrologic Data Assimilation with the Ensemble Kalman Filter, Mon. Weather Rev., 130, 103–114, 2002. 3530
  - Ricci, S., Piacentini, A., Thual, O., Le Pape, E., and Jonville, G.: Correction of upstream flow and hydraulic state with data assimilation in the context of flood forecasting, Hydrol. Earth Syst. Sci., 15, 3555–3575, doi:10.5194/hess-15-3555-2011, 2011. 3531
- <sup>15</sup> Roux, H., Labat, D., Garambois, P.-A., Maubourguet, M.-M., Chorda, J., and Dartus, D.: A physically-based parsimonious hydrological model for flash floods in Mediterranean catchments, Nat. Hazards Earth Syst. Sci., 11, 2567–2582, doi:10.5194/nhess-11-2567-2011, 2011. 3529

Sahu, R. K., Mishra, S., Eldho, T., and Jain, M.: Theoretical Justification of SCS Method for Runoff Estimation, Hydrol. Process., 21, 2872–2881, 2007. 3537

Smith, M., Koren, V. I., Zhang, Z., Reed, S., Pan, J.-J., and Moreda, F.: Runoff response to spatial variability in precipitation: an analysis of observed data, J. Hydrol., 298, 267–286, 2004. 3529

Tabary, P.: The New French Operational Radar Rainfall Product. Part I: Methodology, Weather

<sup>25</sup> Forecasting, 22, 393–408, 2007. 3530, 3535

20

- Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics, 2005. 3530
- Tetzlaff, D. and Uhlenbrook, S.: Significance of spatial variability in precipitation for processoriented modelling: results from two nested catchments using radar and ground station data,
- <sup>30</sup> Hydrol. Earth Syst. Sci., 9, 29–41, doi:10.5194/hess-9-29-2005, 2005. 3529
  - Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., and Habets, F.: A past discharges assimilation system for ensemble streamflow forecasts over France – Part 1: Description and validation of the assimilation system, Hydrol. Earth Syst. Sci., 14, 1623–1637,



doi:10.5194/hess-14-1623-2010, 2010. 3530

- Tramblay, Y., Bouvier, C., Ayral, P.-A., and Marchandise, A.: Impact of rainfall spatial distribution on rainfall-runoff modelling efficiency and initial soil moisture conditions estimation, Nat. Hazards Earth Syst. Sci., 11, 157–170, doi:10.5194/nhess-11-157-2011, 2011. 3541
- <sup>5</sup> USDA Natural Resources Conservation Service: Urban Hydrology for Small Watersheds, TR-55, US Dept. of Agriculture, 1986. 3537

Vieux, B. and Bedient, P.: Assessing urban hydrologic prediction accuracy through event reconstruction, J. Hydrol., 299, 217–236, 2004. 3530

Vrugt, J., Diks, C., Gupta, H., Bouten, W., and Verstraten, J.: Improved treatment of uncertainty

- in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water Resour. Res., 41, 1–17, 2005. 3530
  - Weerts, A. and El Serafy, G.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, 1–17, 2006. 3531
- <sup>15</sup> Wilson, J. W. and Brandes, E.: Radar Measurement of Rainfall A Summary, B. Am. Meteorol. Soc., 60, 1048–1060, 1979. 3530
  - Zoccatelli, D., Borga, M., Zanon, F., Antonescu, B., and Stancalie, G.: Which rainfall spatial information for flash flood response modelling? A numerical investigation based on data from the Carpathian range, Romania, J. Hydrol., 394, 148–161, 2010. 3529



#### Table 1. Rainfall events.

Event date	MFB	$Q_{\text{peak}}$ (m <sup>3</sup> s <sup>-1</sup> )
3 November 1997	4.66	14
16 December 1997	1.74	122
11 November 1999	1.09	43
28 September 2000	1.79	51
23 December 2000	1.50	48
16 January 2001	1.53	93
8 September 2002	1.80	103
8 October 2002	1.74	43
9 December 2002	1.69	376
22 September 2003	1.27	91
15 November 2003	1.58	64
21 November 2003	1.35	95
29 November 2003	1.05	424
5 September 2005	1.29	467
27 January 2006	1.24	52
23 September 2006	1.43	23
1 May 2007	1.01	9
19 October 2008	1.07	109
1 November 2008	0.87	31



Discussion Pa	<b>HESSD</b> 9, 3527–3579, 2012		
per   Discussio	nilation for ection of rainfall der et al.		
n Pape	Title Page		
Ϋ́,	Abstract	Introduction	
_	Conclusions	References	
iscussi	Tables	Figures	
on P	14	►I	
aper	•		
	Back	Close	
Discus	Full Scre	en / Esc	
sion	Printer-friendly Version		
Paper	Interactive Discussion		
	$\sim$	BY	

**Table 2.** The S – catchment wetness state indicator relationship.

Indicator	no. points	М	b	$R^2$	% change	σ
Hu2	21	-8.84 mm	732.00 mm	0.69	0.065	0.055
Bois Saint Mathieu	12	$-5.15 \mathrm{mm}\mathrm{m}^{-1}$	547.57 mm	0.77	0.10	0.14
Claret	12	$-2.98{\rm mmm^{-1}}$	426.79 mm	0.71	0.038	0.037

M is the slope of the linear regression between S and the wetness state indicator and b is the y-intercept for this regression.

	Event date	$S_{ m hu2}~( m mm)$	S <sub>Bois Saint Mathieu</sub> (mm)	$\mathcal{S}_{ ext{Claret}}$ (mm)
	3 November 1997	251	****	****
1	6 December 1997	184	****	****
1	1 November 1999	196	****	****
28	September 2000	220	293	248
2	3 December 2000	197	****	****
	16 January 2001	107	134	125
8	September 2002	211	209	202
	8 October 2002	165	177	213
9	9 December 2002	119	136	153
22	2 September 2003	273	291	294
1	5 November 2003	119	128	139
2	1 November 2003	74	59	80
2	9 November 2003	64	55	80
5	5 September 2005	302	282	288
	27 January 2006	139	136	168
23	3 September 2006	188	181	197
	1 May 2007	216	177	210
	19 October 2008	****	****	****
	1 November 2008	179	155	182

**Table 3.**  $S_{req}$  estimated using physical indicators.



**Table 4.**  $\alpha$ -MFB regression for catchment wetness state indicators.

Indicator	М	b	$R^2$
Hu2	0.73	0.16	0.40
Bois Saint Mathieu	0.77	0.07	0.36
Claret	0.77	0.14	0.47





**Fig. 1.** Visualisation of the Lez Catchment and its monitoring network: **(a)** map of the Lez Catchment and the rain gauges used for the measurement of ground rainfall; **(b)** map of the radar and raingauge network surrounding the Lez Catchment.









Cumulative rainfall reservoir



Fig. 3. Schematic of the ATHYS runoff production function (Bouvier and Delclaux, 1996).





**Fig. 4.** Schematic of the Lag and Route runoff transfer function used in ATHYS (Bouvier and Delclaux, 1996).





**Fig. 5.** Discharge as a function of  $\alpha$  at 3 h before the flood peak.





**Fig. 6.** Schematic representation of the DA algorithm: inputs (blue), model parameters (orange), state variables (purple) and outputs ( $Q_{sim,b}$  – maroon). Inputs, parameters, state variables or model outputs can be corrected by DA using observations ( $Q_{obs}$  – red) and the outputs of the background run ( $Q_{sim,b}$ ) in order to produce the analysis run ( $Q_{obs,a}$  – green).





**Fig. 7.** The outer loop process. The x-axis represents the value of the control vector and the y-axis is the misfit cost (cost function). The red curve represents the non-quadratic true value of the cost function, while the dotted curves represent successive iterations of the outer loop, each with a new estimate of the Jacobian of  $\mathcal{H}$  in the vicinity of the previous analysis.





**Fig. 8.** Reanalysis mode for November 2008. The horizontal dashed line is the lower assimilation threshold (2 m<sup>3</sup>s<sup>-1</sup>).  $\beta_{obs} = 0.25 \text{ m}^6 \text{ s}^{-2}$  and S = 142 mm. Observations are in blue, the background simulation in pink and the analysis simulation in green. Assimilated observations are marked with blue crosses.





**Fig. 9.** Forecast mode for November 2008:  $\beta_{obs} = 0.25 \text{ m}^6 \text{ s}^{-2}$  and S = 142 mm. The black vertical line represents the end of the assimilation period and the start of the pseudo-prediction mode (3h before the flood peak). The horizontal dashed line is the lower assimilation threshold  $(2 \text{ m}^3 \text{ s}^{-1})$ . Observations are in blue, the background simulation in pink, the first iterate of the external loop in black and the analysis simulation in green. Assimilated observations are marked with blue crosses. All simulations have 5 iterates of the external loop, however, the algorithm converges after the second iterate in this case, so only the first iterate and the final analysis are shown.





**Fig. 10.** Comparison of background IE values with IE values following data assimilation (analysis). The x-axis contains the episode label in the format mYYpp, where m is the first letter of the month (j is January and m is May), YY is the year and pp is the peak number for the 2nd and greater peaks.





**Fig. 11.** Regression of  $\alpha$  versus MFB. y = x is drawn in red and the regression in blue.





**Fig. 12.** Improvements in simulation quality indicators for the 19 rainfall runoff episodes. The dark blue bars represent  $IE_{\alpha} - IE_{MFB}$ . The light blue bars are the difference in the normalized peak flow criteria,  $PH_{MFB} - PH_{\alpha}$ .





**Fig. 13.** Reanalysis mode December 2003: (a) discharges for December 2003: observations are in blue, the background simulation in pink and the analysis simulation in green; (b) the error in the simulated discharge,  $Q_{\text{background}} - Q_{\text{observations}}$  (red).





**Fig. 14.** Boxplots of simulation performance: **(a)**  $IE_{analysis} - IE_{background}$ ; **(b)**  $PH_{background} - PH_{analysis}$ . The simulations shown are: Hu2025 ( $S = S_{hu2}$ ;  $\beta_{obs} = 0.25 \text{ m}^6 \text{ s}^{-2}$ ), Opt025 ( $S = S_{cal}$ ;  $\beta_{obs} = 0.25 \text{ m}^6 \text{ s}^{-2}$ ), Opt25 ( $S = S_{cal}$ ;  $\beta_{obs} = 25 \text{ m}^6 \text{ s}^{-2}$ ) and Opt250 ( $S = S_{cal}$ ;  $\beta_{obs} = 250 \text{ m}^6 \text{ s}^{-2}$ ).





**Fig. 15.** Boxplots of simulation performance:  $6 \text{ IE}_{analysis} - \text{IE}_{background}$ ;  $6 \text{ PH}_{background} - \text{PH}_{analysis}$ . The simulations shown are: bsm ( $S = S_{\text{BoisSaint Mathieu}}$ ;  $\beta = 250 \text{ m}^6 \text{ s}^{-2}$ ), claret ( $S = S_{\text{Claret}}$ ;  $\beta = 250 \text{ m}^6 \text{ s}^{-2}$ ), and hu2 ( $S = S_{\text{Hu2}}$ ;  $\beta = 250 \text{ m}^6 \text{ s}^{-2}$ ).

