Hydrol. Earth Syst. Sci. Discuss., 9, 3087–3127, 2012
www.hydrol-earth-syst-sci-discuss.net/9/3087/2012/
doi:10.5194/hessd-9-3087-2012

Hydrology and Earth System Sciences Discussions

Generating spatial precipitation ensembles: impact of temporal correlation structure

O. Rakovec[1], P. Hazenberg[1], P. J. J. F. Torfs[1], A. H. Weerts[2], and R. Uijlenhoet[1]

[1]Hydrology and Quantitative Water Management Group, Department of Environmental Sciences, Wageningen University, Wageningen, The Netherlands
[2]Deltares, P.O. Box 177, 2600 MH Delft, The Netherlands

Received: 27 February 2012 – Accepted: 5 March 2012 – Published: 12 March 2012

Correspondence to: O. Rakovec (olda.rakovec@wur.nl)

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

Abstract

Sound spatially distributed rainfall fields including a proper spatial and temporal error structure are of key interest for hydrologists to force hydrological models and to identify uncertainties in the simulated and forecasted catchment response. The current paper presents a temporal coherent error identification method based on time-dependent multivariate spatial conditional simulations, which are made further conditional on preceding simulations.

Synthetic and real world experiments are carried out within the hilly region of the Belgian Ardennes. Precipitation fields are simulated for pixels of $10 \times 10\,\text{km}^2$ resolution. Uncertainty analyses in the simulated fields focus on (1) the number of previous simulation hours on which the new simulation is conditioned, (2) the advection speed of the rainfall event, (3) the size of the catchment considered, and (4) the rain gauge density within the catchment. The results for a synthetic experiment show for typical advection speeds of $>20\,\text{km}\,\text{h}^{-1}$, no uncertainty is added in terms of across ensemble spread when conditioned on more than one or two previous simulations. However, for the real world experiment, additional uncertainty can be still added when conditioning on a higher number of previous simulations. This is, because for actual precipitation fields, the dynamics exhibit a larger spatial and temporal variability. Moreover, by thinning the observation network with 50 %, the added uncertainty increases only slightly. Finally, the first order autocorrelation coefficients show clear temporal coherence in the time series of the areal precipitation using the time-dependent multivariate conditional simulations, which was not the case using the time-independent univariate conditional simulations.

1 Introduction

Precipitation is the most dominant input term determining the hydrological response at the catchment scale (Beven, 2001). Historically, spatial precipitation information was obtained by rain gauges measurements based on point scale estimates. However, during the last decades, application of weather radar at larger scales has improved our understanding on the spatial and temporal properties of rainfall even further. Unfortunately, precipitation estimates by weather radar are prone to biases (e.g. Seo et al., 1999; Steiner et al., 1999; Seo and Breidenbach, 2002; Hazenberg et al., 2011). To date, this implies that operational nowcasting/forecasting systems have to make use of rain gauge information to adjust radar precipitation fields for bias (e.g. Schuurmans et al., 2007; Germann et al., 2009; Goudenhoofdt and Delobbe, 2009; Cole and Moore, 2009). As such, rain gauges remain an important tool for the derivation of unbiased spatially distributed rainfall estimates.

To obtain sound spatially distributed rainfall information from rain gauge observations, these devices generally are interpolated to appropriate spatial and temporal resolutions, depending on the hydrological purpose. Among interpolation methods, geostatistical techniques (like kriging) are popular. These methods take into account information about the spatial variation within an area, and provide both a mean rainfall as well as an associated error estimate (Webster and Oliver, 2001; Schuurmans and Bierkens, 2007). These errors in areal rainfall are of key interest for hydrologists because they can be used to estimate uncertainties in catchment response. An evaluation of those errors in a spatially lumped manner has been discussed e.g. by Zawadzki (1973) and Willems (2001). We refer to Villarini et al. (2008) and Ciach and Krajewski (2006) and references cited therein, for analyses of different spatial and temporal sampling errors from a rain gauge perspective.

Nevertheless, kriging is prone to smooth local variability of rainfall. Further away from observation points, high (low) values tend to be underestimated (overestimated) (Goovaerts, 1997). These biases decrease the usability of kriging in applications sensitive to extreme values (Goovaerts, 1997), such as within spatially distributed rainfall-runoff modelling (Bivand et al., 2008).

However, sound spatial and temporal estimates of precipitation and its corresponding uncertainty are of key interest for both scientific and applied hydrological studies

(Liu et al., 2012). The generation of an ensemble, which is a finite and discrete number of spatial realizations over time, is able to realize this goal. A common practice in hydrological data assimilation applications to obtain ensembles is to perturb the interpolated point or spatially distributed estimates by Gaussian white noise with a standard deviation ranging between 15–50 % of the observed precipitation (e.g. Pauwels and De Lannoy, 2006; Weerts and El Serafy, 2006; Clark et al., 2008). Although this approach leads to hydrological model simulations with wide discharge uncertainty bands, the realizations are not very probable from a hydro-meteorological perspective, because of a lack in coherent temporal evolution of each individual precipitation realization.

Sound spatially distributed rainfall fields including a proper spatial error structure can be obtained by conditional simulation. Unlike interpolation this technique provides both the best local estimate, as well as ensures that realizations match the sample statistics and are conditional on neighbouring estimates. In other words, conditional simulations provide proper information about the spatial uncertainty (Goovaerts, 1997). Several hydrological studies have applied conditional simulations at daily (e.g. Schuurmans, 2008; Vischel et al., 2009) and at hourly time steps (Renard et al., 2011).

Unfortunately, conditional simulations do not primarily take the temporal evolution of the spatial field into account (Goovaerts, 1997; Webster and Oliver, 2001; Bivand et al., 2008). Nevertheless, for precipitation the temporal correlation structure can be an important aspect to be considered when generating spatial precipitation ensembles. Theoretically, this can be achieved using spatial conditional simulations which are made conditional on previous simulations. Neglecting this temporal aspect would lead to underestimation of the overall uncertainty in precipitation ensembles.

The objective of this study is to define a plausible precipitation ensemble generator using rain gauges to capture the temporal coherence for each spatial realization at an hourly time step. Our analyses focus on the uncertainty in the simulated fields based on (1) the number of previous simulated hours on which the new simulation is conditioned, (2) the advection speed of the rainfall event, (3) the size of the catchment considered and (4) the rain gauge density within the catchment.

HESSD
9, 3087–3127, 2012

2 Material and methods

2.1 Data

Hourly precipitation data are available from 42 automatic rain gauges situated within the Belgian Ardennes region (Fig. 1). This moderately hilly terrain with maximum elevations of ~700 m a.m.s.l. is predominantly drained by the Meuse River and partly by the Rhine River (Berne et al., 2005; Driessen et al., 2010).

In this paper we focus on the analysis of three representative stratiform winter rainfall events as described and analyzed by Hazenberg et al. (2011): (1) a fast-moving stratiform system (22 October 2002), (2) a large-scale stratiform system (22 December 2002) and (3) a fast-moving frontal stratiform system (1 January 2003). For a further description of these events, the reader is referred to Hazenberg et al. (2011). The main characteristics of these events are given in Table 1. Additionally, the autocorrelation coefficients of the catchment average precipitation of the Upper Ourthe for 30 rainy events with a minimum duration of 13 h are presented in Fig. 2. All 30 events were observed during the winter half year, from 1 October 2002 to 31 March 2003.

2.2 Geostatistical analysis

The variogram is a geostatistical measure of spatial variability in terms of the semivariance over a lag distance h. The experimental omnidirectional semi-variogram, which is generally defined as the variogram, assumes stationarity and isotropy of the predicted variable. It represents half of the variance between paired data values (i.c. measured precipitation) within the same binned lag distance h:

$$\hat{\gamma}(h) = \frac{1}{2\,N_h} \sum_{k=1}^{N_h} \left(z(x_k) - z(x_k + h)\right)^2, \tag{1}$$

where N_h is the number of data observation pairs and $z(x_k)$ and $z(x_k + h)$ are the observations separated by the lag distance h. Because the experimental variogram is

HESSD
9, 3087–3127, 2012

derived only for several discrete lag distances, a parametric variogram model has to be fitted in order to obtain continuous estimates of the semi-variance.

The spherical model is a popular variogram model (Berne et al., 2004; Schuurmans et al., 2007; van de Beek et al., 2011; Verworn and Haberlandt, 2011). With only three parameters it is defined as follows:

$$\gamma(h) = \begin{cases} c_o + c_1 \left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right) & \text{if} \quad h \leq a \\ c_o + c_1 & \text{if} \quad h > a. \end{cases} \tag{2}$$

In Eq. (2), the parameter c_0 is the nugget, representing the semi-variance at distance $h=0$. The parameter c_1 is the partial sill, while a represents the range, the distance beyond which the data are not correlated any more.

Figure 3 shows two examples of the experimental variogram as well as the fitted spherical model. Note that the unit of semivariance is in mm instead of mm^2. This is because the quantitative statistical measures, which are employed in this study, are particularly suitable for normally distributed data. However, rainfall by nature does not follow a Gaussian distribution. Therefore, a pragmatic and popular solution to overcome this problem is to transform the rainfall data such that their distribution approaches a Gaussian distribution. As such, rainfall data are square-root transformed (Schuurmans et al., 2007; van de Beek et al., 2011).

Additionally, Eq. (2) can be extended for the time-dependent multivariate case, which relates spatial dependency between two variables z_i and z_j and yields the cross variogram:

$$\hat{\gamma}_{ij}(h) = \frac{1}{2\,N_h}\sum_{k=1}^{N_h}\left[\left(z_i(x_k) - z_i(x_k + h)\right)\left(z_j(x_k) - z_j(x_k + h)\right)\right]. \tag{3}$$

This property is defined as half the expectation of the product of the increments of two variables (Wackernagel, 2003). Obviously, if $z_i = z_j = z$, then Eq. (3) reduces to Eq. (1). Figure 3 shows an example of an experimental cross-variogram and the fitted spherical model. Interpretations of direct and cross-variograms will be discussed in Sect. 3.1.

2.3 Conditional simulation

Conditional simulation is a geostatistical method, which generates multiple realizations that all reasonably match the sample statistics and exactly match the conditioning data (Goovaerts, 1997). As such, conditional simulation is a useful tool to model and quantify spatial uncertainty of a variable such as precipitation (e.g. Schuurmans, 2008; Vischel et al., 2009; Renard et al., 2011). Time-independent univariate conditional simulations depend on the spatial observations by rain gauges for a given simulation hour. However, by performing time-dependent multivariate conditional simulations, it becomes possible to simulate rainfall fields conditional on both previously simulated precipitation fields as well as the rainfall observations by rain gauges. This approach enables to introduce temporal coherence for each simulated grid point over time.

In the current study, the gstat R package (Pebesma, 2004; Rossiter, 2007; R Development Core Team, 2011) was used to simulate conditional precipitation fields. For a comprehensive overview of the theory behind conditional (sequential Gaussian) simulations we refer to Goovaerts (1997). Here, only a brief summary is presented:

Initially, a normal transformation of rainfall data is carried out. Then, the simulation is performed on the transformed dataset according to following steps (Goovaerts, 1997):

1. A random path throughout all the grid nodes is defined, where all nodes are visited only once.

2. At each grid note, a random number is drawn from a Gaussian distribution with parameters equal to the kriging prediction and variance. This number is added to the dataset used to condition the subsequently simulated grid nodes.

3. After an estimate is obtained for all grid nodes, the back-transformation of the simulated normal values to the original rainfall distribution is performed.

HESSD
9, 3087–3127, 2012

By performing these steps, one time-independent univariate ensemble realization for one time step is generated. Other realizations can be obtained using different random paths over the simulation grid domain.

For the purpose of computational stability, we further focus on rainy periods, which are defined as a cluster of consecutive rainy hours, for which each individual hour satisfies a minimal intensity condition. More specifically, the mean of all rain gauge observations should have at least a minimum value of 0.1 mm and the maximum individual observation has to exceed 0.5 mm. Additionally, to prevent computational instability, rain gauges with zero rainfall are set to a small value of 0.05 mm.

Then, time-dependent multivariate conditional simulation are carried out for each rainy period according to the following steps:

1. Initially, N ensemble realizations conditional on the rain gauge data and the variogram model are simulated for the first hour of each rainy period (time-independent univariate conditional simulation).

2. For the following simulation hours, each ensemble realization is conditioned on both (1) direct and (2) cross variogram models, and additionally on a number of previously simulated hours of the corresponding realization. The conditional simulation history will be called simulation memory M in the remainder of this paper.

2.4 Mathematical notation

Throughout this paper the following notations are used. A time series of rainfall realizations at the n-th to be simulated pixel is defined using the following matrix:

$$\mathbf{R}_{j,t}^{n,m} = \begin{pmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,T} \\ R_{2,1} & R_{2,2} & \cdots & R_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ R_{J,1} & R_{J,2} & \cdots & R_{J,T} \end{pmatrix} \tag{4}$$

HESSD

9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.

where m is the simulation memory scenario, j is an index of the ensemble realization and t stands for time. J is the ensemble size and T is the duration of the rainy period.

In general, two types of approaches can be used to disentangle the uncertain variation. The first way of looking at one precipitation realization is to evaluate and quantify its corresponding statistics for the whole rainfall event. This approach is widely employed by catchment hydrologists, who are interested in the overall uncertainty over the precipitation event (bold line in Fig. 4). The second way quantifies the uncertainty across the ensemble for each individual time step (□ in Fig. 4) and is more of an interest for hydrologists dealing with flood forecasting.

Having defined the **R** matrix, we can derive the first two central moments of **R** from two different perspectives: (1) event-based (along the time axis having index T, Eqs. 5 and 6) and (2) across-ensemble (having index J, Eqs. 7 and 8):

$$\hat{\mu}_{J,t}[\mathbf{R}] = \frac{1}{J}\sum_{j=1}^{J} R_{j,t} \tag{5}$$

$$\hat{\sigma}^2_{J,t}[\mathbf{R}] = \frac{1}{J}\sum_{j=1}^{J} \left(R_{j,t} - \hat{\mu}_{J,t}[\mathbf{R}]\right)^2 \tag{6}$$

$$\hat{\mu}_{T,j}[\mathbf{R}] = \frac{1}{T}\sum_{t=1}^{T} R_{j,t} \tag{7}$$

$$\hat{\sigma}^2_{T,j}[\mathbf{R}] = \frac{1}{T}\sum_{t=1}^{T} \left(R_{j,t} - \hat{\mu}_{T,j}[\mathbf{R}]\right)^2. \tag{8}$$

The uncertainty within the precipitation ensemble can be expressed by the coefficient of variation (CV), the ratio of the standard deviation of the data set to its mean, which

HESSD
9, 3087–3127, 2012

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

represents a normalized dispersion and enables comparison between the generated ensembles for different scenarios (e.g. Wackernagel, 2003):

$$\mathrm{CV} = \frac{\hat{\sigma}}{\hat{\mu}}. \tag{9}$$

2.5 Experimental setup

2.5.1 Interpretation of direct and cross variograms

To understand how the direct and cross-variograms reflect the spatial variability between different precipitation fields, two examples of experimental and modelled variograms are analyzed for the case of a synthetic spherical rainfall cell moving over a $145 \times 145\,\mathrm{km}^2$ grid with $1 \times 1\,\mathrm{km}^2$ resolution (Fig. 5). As part of this case, the effect of rain gauge density on the experimental and modelled variograms is addressed and includes sampling either from all 21 025 grid pixels (a dense synthetic observation network) or using only the 42 rain gauge pixels (actual real world network). As such, both the impact of using a sparse gauge network and temporal correlation can be identified.

2.5.2 Conditional simulations: synthetic experiment

Next in this study, simulations are carried out for a number of synthetic experiments (scenarios), which enable one to obtain better understanding of individual contributions of uncertainty in the synthetic simulated fields. The synthetic experiments encompass four aspects:

1. Time: eight types of simulations with, given the time lag, simulation memories of 0–7 h.

2. Advection speed: five synthetic spherical rainfall cells with an area of about $4100\,\mathrm{km}^2$ (72.5 km diameter), an intensity of $2\,\mathrm{mm\,h}^{-1}$ and moving at different advection speeds of 6, 8, 11, 17, $25\,\mathrm{km\,h}^{-1}$ over the simulation domain (Fig. 5).

HESSD
9, 3087–3127, 2012

The duration of the rainfall events is 18, 12, 9, 6 and 4 h, respectively. As such, the dimensions of these synthetic rainfall cells are similar to those observed within the region.

3. Area: six synthetic nested sub-catchments over which the analysis is carried out (Fig. 6a).

4. Observation density: three types of rain gauge densities (Fig. 6b): (1) the actual observation network consists 27 rain gauges, (2) the reduced network has 14 rain gauges and (3) the complete synthetic network has 100 rain gauges. Removal of the rain gauges from the actual observation network follows a method (Goudenhoofdt and Delobbe, 2009), which keeps the spatial distribution of remaining gauges as uniform as possible. First, the sum of the inverse distance between the four nearest gauges is calculated for each gauge and then half of the gauges with the highest values are removed. The complete synthetic network for the catchment's pixels without any real rain gauge is obtained by randomly drawing x- and y-coordinates from a uniform distribution.

Because of the higher computational costs of the time-dependent multivariate conditional simulations at high resolution grids, the simulation domain is reduced for that purpose to $100 \times 100\,\mathrm{km}^2$ with a $10 \times 10\,\mathrm{km}^2$ raster resolution (dashed box in Fig. 5). The analysis includes 24 ensemble realizations and the length of the rainy periods varies between four and 18 h, depending on the advection speed.

2.5.3 Conditional simulations: real world experiment

Finally, the real world experiment will focus on the three events described by Hazenberg et al. (2011) (see Sect. 2.1). For these events, the impact of time and area as described in Sect. 2.5.2 are analyzed using the actual observation network (27 rain gauges).

HESSD
9, 3087–3127, 2012

3 Results

3.1 Interpretation of direct and cross variograms

Empirical and modelled spherical direct and cross variograms (Sect. 2.2) are calculated for two cases of a synthetic spherical rainfall event with an intensity of 2 mm h^{-1}. The first case consists of two rain fields (see 1 and 2 in Fig. 7), which are complementary. The second case consists of four advected rainfall fields (Fig. 7, right panel). The intersection of the third and the fourth rainfall field (Fig. 7, right panel) yields exactly half of the rainfall area. Since the third and the fifth rainfall fields are tangent, the intersection of the fourth and the fifth rainfall field covers around 28 % of the rainfall area (Glassner, 1998). Moreover, the intersection of the fifth and the sixth rainfall field yields exactly half of the rainfall area. We chose a rather small sized spherical rainfall cell with a 36 km diameter, to minimize the impact of boundary effects.

Figure 8a and b show the direct and cross-variograms of the two complementary rainfall fields in Fig. 7 (left and middle panels). These have been square-root transformed, for the two sampling densities: a dense synthetic network (21 025 points covering all raster pixels) and the real world network (42 rain gauges). We can observe that the two direct variograms are identical and symmetrical with respect to their cross-variogram. This holds for both sampling densities. As expected, the real world rain gauge network has a higher sill and a larger scatter in the empirical variograms than the much denser synthetic observation network. The empirical variogram is estimated very well by the fitted spherical variogram model.

For the advected cell of Fig. 7 (right panel), the direct and cross-variograms are shown in Fig. 8c and d. For decreasing intersected area between two rainfall fields, the fitted spherical sills decrease proportionally using the dense synthetic network (Table 2). Zero overlapping area (i.e. three combinations: rainfall fields 3 and 5, 3 and 6, 4 and 6), gives a sill of about zero. For the real world network similar behaviour can be observed, although the scatter in the fitted models through the empirical variograms is much larger. Note that even though the spherical rainfall cell is identical for all rainfall

HESSD
9, 3087–3127, 2012

fields, the shape of the empirical variograms differs from each other, because the rain gauge configuration is not spatially uniform and effectively changes for each rainfall field.

3.2 Conditional simulations: synthetic experiment

As a first example, the impact of advection speed and the number of hours used as part of the conditional simulation was identified for the setup presented in (Fig. 5). As such, precipitation is "registered" by the real-world rain gauge network ($N = 27$ in Fig. 6), while a total of up to eight hours were used as part of the simulation memory (0–7 h). An example of such a simulation for one pixel is shown in Fig. 9. We can observe that for the time-independent univariate case (i.e. conditioned on 0 h of previously simulated fields), there is no temporal consistency for ensemble realizations over time, since no information between individual time steps is taken into account. However, for the time-dependent multivariate cases (i.e. conditioned on 1–7 h of previously simulated fields), the temporal consistency for ensemble realizations becomes more clear. Overall, the spread in simulated precipitation increases when a larger number of previously simulated fields is included as part of the simulations.

From the event-based perspective (see bold line in Fig. 4), the scatter plots between the simulated mean precipitation over time ($\hat{\mu}_{T,j}$) and corresponding standard deviation ($\hat{\sigma}_{T,j}$) for the first four ensemble realizations are shown in Fig. 10 separately, where $\hat{\mu}_{T,j}$ and $\hat{\sigma}_{T,j}$ are plotted for all the individual pixels within the 4900 km^2 catchment (see Fig. 6b). The panels indicate that the spread in $\hat{\mu}_{T,j}$ gradually increases from the time-independent univariate conditional simulation to the most complex time-dependent multivariate scenario. This is in agreement with the simulation results shown in Fig. 9. Because we are not only interested in the mean simulated values, but also in their temporal variability, standard deviations are plotted against the corresponding means (Fig. 11). The slope of the fitted linear regression line represents the mean temporal coefficient of variation. The coefficients of variation are gradually decreasing

HESSD
9, 3087–3127, 2012

for longer simulation memories, which indicates lower temporal variability and larger temporal coherence for longer simulation memories.

For the across-ensemble perspective (see □ in Fig. 4), the scatter plots between simulated mean precipitation across-ensemble ($\hat{\mu}_{J,t}$) and their corresponding standard deviations ($\hat{\sigma}_{J,t}$) for the four time steps ($t = 7, 8, 9, 10$) are shown in Fig. 11, where $\hat{\mu}_{J,t}$ and $\hat{\sigma}_{J,t}$ are plotted for all individual pixels within the 4900 km^2 catchment. The figures reflect a rather constant spread in $\hat{\mu}_{J,t}$ for all simulation memories during those four time steps. Nevertheless, the fitted coefficients of variation are gradually increasing for longer simulation memories. This is in line with higher ensemble spread for longer simulation memories (see Fig. 9).

Since the purpose of this paper is to assess the impact of precipitation uncertainty estimation across the ensemble (i.e. $\hat{\mu}_{J,t}$ vs. $\hat{\sigma}_{J,t}$), which is especially of key interest for hydrological data assimilation applications, the next step is to evaluate it in a lumped manner over all time steps. This was done by overlapping the individual sub-plots (partially depicted for $t = 7, 8, 9, 10$ in Fig. 11) for all time steps. An example is shown in Fig. 12, where the coefficients of variation increase with rainfall simulations conditioned on longer simulation memories.

The behaviour of the coefficients of variation (CV_J) for individual catchment sizes, advection speeds and simulation memories is summarized in Fig. 13. We can observe that for a larger advection speed the coefficient of variation rises faster and approaches its upper level earlier than for slower moving systems. Note that for the faster events, with advection speeds of 25 and 17 km h^{-1} lasting four and six hours respectively (Fig. 5), the coefficients of variation can be obtained only up to the three and five previously simulated hours. Additionally, the catchment size does not have a large influence on the coefficients of variation except for very small catchment sizes consisting of only few pixels, where the estimation is significantly affected by sampling uncertainty. Because small catchments are nested, most variability is smoothed out for larger catchments sizes.

HESSD
9, 3087–3127, 2012

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◂◂ | ▸▸
◂ | ▸
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

The before-mentioned examples were based on the real world rain gauge network ($N = 27$). When the number of rain gauges is decreased to the half of its original density ($N = 14$), the general shape of the fitted spherical variogram through the estimated coefficients of variation remains very similar. However, both y-intercept and horizontal asymptote values become higher (Table 3). This indicates a slight increase in the across-ensemble variability. For a dense synthetic network ($N = 100$) the opposite occurs. Both, the y-intercept and horizontal asymptote values decrease, which means lower across-ensemble variability.

3.3 Conditional simulations: real world experiment

The real rain gauge observations were analyzed within the same conditional simulation framework as was done within the synthetic experiment. The resulting coefficients of variation (CV_J) for the different catchment sizes and simulation memories are shown in Fig. 14. They correspond well with the synthetic experiment. For the two fast moving systems (22 October 2002 and 1 January 2003), there is a steep increase in the across ensemble spread (CV_J), which becomes more or less steady after simulation memories of 1–2 h. This means that no uncertainty is added to simulated precipitation fields by conditioning on more than two hours of previous simulations. For a large-scale stratiform system (22 December 2002) moving very slowly, a gentle rise in CV_J is observed. For this event, the horizontal asymptote is reached when a simulation memory of >5 h is taken into account. Moreover, these values of CV_J are considerably smaller than for the faster systems.

4 Synthesis and discussion

The overall temporal correlation structure of the simulated precipitation field can be quantified using the first order autocorrelation coefficient r_{1h}, which expresses the correlation of a precipitation time series for a time lag of 1 h. Figure 15 shows boxplots of r_{1h} for time series of areal precipitation for different catchment sizes and simulation memories. For robust investigation of the autocorrelation, it is preferable to have a long time series. Therefore, it was chosen here to use the data for the slowest synthetic event with a duration of 18 h (see Fig. 5). It can be observed that r_{1h} increases when moving from time-independent univariate ($M = 0$) to the time-dependent multivariate conditional simulations ($M = 1$–7 h). Nevertheless, the impact of this discontinuity decreases for larger catchments, which is caused by a limited spatial extent of the synthetic spherical rainfall cell advecting over the catchment.

The across-ensemble uncertainty was quantified using the lumped CV_J and its shape for both the synthetic experiment (Fig. 13) as well as the real world experiment (Fig. 14) clearly resembles the spherical variogram model (Eq. 2). By fitting the spherical model, the range can be obtained, which represents a simulation memory threshold of the system, after which no additional precipitation uncertainty is added by including more previous information.

For the synthetic experiment, this leads to a non-linear relation between the advection speed and its corresponding fitted range (Fig. 16):

$$\text{Range} \times \text{Speed} \approx 50\,\text{km}. \tag{10}$$

This result indicates that for typical advection speeds (>20 km h^{-1}) no uncertainty in terms of across ensemble spread is added to the simulated precipitation fields by conditioning it on more than two previous simulations.

We need to bear in mind, that the synthetic case analyzed here is the most simplistic example of a precipitation cell, which assumes a known and constant advection speed, has a rather small dimension, and results in a constant rainfall intensity. For real world examples, however, a much higher spatial and temporal variability in the dynamics of precipitation systems can be expected. This means that the effective ranges of previous information are expected to increase. This corresponds well with the results obtained for the real world experiment, which are shown in Fig. 16. To quantify the apparent uncertainty in the fitted ranges for the three real rainfall events, the values of

HESSD
9, 3087–3127, 2012

the ranges for 80 %, 90 % and 100 % of the partial sills are shown. For the large scale stratiform system (red triangles), the fitted ranges vary around 5–7 h, which is considerably longer than for the two faster systems. This can be caused by the combination of (1) the gentle increase and in general lower values of coefficient of variation, which are about half of the values for the two remaining events (recall Fig. 14) and (2) the size of the observed precipitation system, which for the real world case has a larger dimension. Because of this latter property, we repeated the synthetic experiment for rainfall cells with twice the original diameter size, which made its area four times larger. For these cases, indeed, a higher coefficient of 63 km is obtained (gray line in Fig. 16).

Overall, for the time-dependent multivariate conditional simulations with longer simulation memories, we observed larger across-ensemble spread. The commonly defined time-independent rainfall perturbations used for the hydrological data assimilation applications reach the errors with a standard deviation up to 50 % of the observed precipitation (e.g. Pauwels and De Lannoy, 2006; Weerts and El Serafy, 2006; Clark et al., 2008). This corresponds well with our simulations, in which the maximum slope of the fitted coefficients of variations is about 0.5 (Fig. 11). However, in comparison with the aforementioned references, we were able to additionally capture the temporal coherence for each realization in space.

Finally, conditional simulation methods increase computational costs quite dramatically in comparison with interpolation methods. Fortunately, this problem can be partly circumvented by decreasing the temporal (Δt [min]) or spatial (Δr [km]) resolution of the simulation model. From an applied hydrological point of view, an hourly time step is usually recommended for regions with an area of ~10 000 km^2 (Berne et al., 2004). Therefore, the choice for a rather coarse 10 × 10 km^2 grid resolution, as was chosen in this study, can be supported by the analysis carried out by Berne et al. (2004), who reported the spatial rainfall resolution to be $4.5\sqrt{\Delta t}$, which yields a decorrelation distance of about 35 km.

5 Summary and conclusions

In this paper, a rain gauge precipitation ensemble generator at hourly time step using time-dependent multivariate conditional simulations was developed, which were made conditional on previous simulations back in time. As such, a plausible way to generate temporal correlation structures for precipitation for each realization over time was introduced. Next, we identified the uncertainty and the temporal correlation structures in the simulated fields based on (1) the number of previous simulation hours, on which the new simulation is conditioned, (2) the advection speed of the rainfall event, (3) the size of the catchment considered and 4) the rain gauge density within the catchment.

The synthetic experiment shows that for typical advection speeds of >20 km h^{-1} no uncertainty in terms of across ensemble spread lumped over time (expressed using the coefficient of variation) is added to simulated precipitation fields by conditioning them on more than one or two previous simulations. In the real world experiment, which exhibits a larger spatial and temporal variability, the time-dependent simulations require somewhat longer simulation memories. Furthermore, by halving the observation network, i.e. using 14 rain gauges, the uncertainty in the synthetic experiment increases only slightly. Finally, the first order autocorrelation coefficient showed temporal coherence in the time series of the areal precipitation using the time-dependent multivariate conditional simulation in comparison with the time-independent univariate conditional simulations. Nevertheless, this coherence decreased with increased catchment area.

The presented technique to generate spatial precipitation ensembles can be easily implemented within a hydrological data assimilation framework to be used as an improvement over currently used simplistic approaches to perturb the interpolated point or spatially distributed estimates (as referred to in the introduction). As shown, using the time-dependent rainfall simulations with at least one hour simulation memory, but preferably longer, we were able to reach this goal and obtain precipitation ensembles with temporal correlation structures that are plausible from a hydro-meteorological perspective. Therefore, the corresponding simulated spatially distributed model states obtained by that ensemble should inherit this temporal aspect as well. The advantage of having the temporal coherence in model states is that it eliminates smoothing of possible extreme state values, which can be the case when neglecting it.

Acknowledgements. We thank the Hydrological Service of the Walloon Region of Belgium (MET-SETHY) and the Royal Meteorological Institute of Belgium (KMI) for providing the hydrological and meteorological data. This research is financially supported by the Flood Control 2015 program (www.floodcontrol2015.nl) and the EU-FP7 Project IMPRINTS (FP7-ENV-2008-1-226555). This support is gratefully acknowledged.

References

Berne, A., Delrieu, G., Creutin, J.-D., and Obled, C.: Temporal and spatial resolution of rainfall measurements required for urban hydrology, J. Hydrol., 299, 166–179, doi:10.1016/j.jhydrol.2004.08.002, 2004. 3092, 3103

Berne, A., ten Heggeler, M., Uijlenhoet, R., Delobbe, L., Dierickx, Ph., and de Wit, M.: A preliminary investigation of radar rainfall estimation in the Ardennes region and a first hydrological application for the Ourthe catchment, Nat. Hazards Earth Syst. Sci., 5, 267–274, doi:10.5194/nhess-5-267-2005, 2005. 3091

Beven, K.: Rainfall-runoff modelling: the primer, John Wiley & Sons, 2001. 3088

Bivand, R. S., Pebesma, E. J., and Gmez-Rubio, V.: Applied spatial data analysis with R, Springer, 2008. 3089, 3090

Ciach, G. and Krajewski, W.: Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma, Adv. Water Resour., 29, 1450–1463, doi:10.1016/j.advwatres.2005.11.003, 2006. 3089

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, Adv. Water Resour., 31, 1309–1324, doi:10.1016/j.advwatres.2008.06.005, 2008. 3090, 3103

Cole, S. J. and Moore, R. J.: Distributed hydrological modelling using weather radar in gauged and ungauged basins, Adv, Water Resour., 32, 1107–1120, 2009. 3089

Driessen, T. L. A., Hurkmans, R. T. W. L., Terink, W., Hazenberg, P., Torfs, P. J. J. F., and Uijlenhoet, R.: The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model, Hydrol. Earth Syst. Sci., 14, 651–665, doi:10.5194/hess-14-651-2010, 2010. 3091

Germann, U., Berenguer, M., Sempere-Torres, D., and Zappa, M.: REAL-Ensemble radar precipitation estimation for hydrology in a mountainous region, Q. J. Roy. Meteorol. Soc., 135, 445–456, doi:10.1002/qj.375, 2009. 3089

Glassner, A. S., ed.: Graphics gems, Elsevier, 1998. 3098

Goovaerts, P.: Geostatistics for Natural Resources Evaluation, Applied geostatistics series, Oxford University Press, 1997. 3089, 3090, 3093

Goudenhoofdt, E. and Delobbe, L.: Evaluation of radar-gauge merging methods for quantitative precipitation estimates, Hydrol. Earth Syst. Sci., 13, 195–203, doi:10.5194/hess-13-195-2009, 2009. 3089, 3097

Hazenberg, P., Leijnse, H., and Uijlenhoet, R.: Radar rainfall estimation of stratiform winter precipitation in the Belgian Ardennes, Water Resour. Res., 47, W02507, doi:10.1029/2010WR009068, 2011. 3089, 3091, 3097, 3112

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P. J., van Dijk, A., van Velzen, N., He, M., Lee, H., Jin Noh, S., Rakovec, O., and Restrepo, P.: Toward advancing data assimilation in operational hydrologic forecasting and water resources management: a review of the current status, challenges, and emerging opportunities, Hydrology and Earth System Sciences Discussions, Preface for the Special Issue on “Latest advances and developments in data assimilation for operational hydrologic forecasting and water resources management”, submitted, 2012. 3090

Pauwels, V. R. N. and De Lannoy, G. J. M.: Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge, J. Hydrometeorol., 7, 458–477, doi:10.1175/JHM490.1, 2006. 3090, 3103

Pebesma, E. J.: Multivariable geostatistics in S: the gstat package, Comput. Geosci., 30, 683–691, doi:10.1016/j.cageo.2004.03.012, 2004. 3093

R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, http://www.R-project.org, last access: 7 December 2011, ISBN 3-900051-07-0, Vienna, Austria, 2011. 3093

HESSD

9, 3087–3127, 2012

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., and Franks, S. W.: Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, Water Resour. Res., 47, W11516, doi:10.1029/2011WR010643, 2011. 3090, 3093

Rossiter, D.: Technical Note: Co-kriging with the gstat package of the R environment for statistical computing, Version 2.1, Self-published online, http://www.itc.nl/personal/rossiter/teach/R/R_ck.pdf (last access: 1 February 2012), 2007. 3093

Schuurmans, J. M.: Hydrological now- and forecasting, Ph.D. thesis, Utrecht University, The Netherlands, 2008. 3090, 3093

Schuurmans, J. M. and Bierkens, M. F. P.: Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model, Hydrol. Earth Syst. Sci., 11, 677–693, doi:10.5194/hess-11-677-2007, 2007 3089

Schuurmans, J. M., Bierkens, M. F. P., Pebesma, E. J., and Uijlenhoet, R.: Automatic prediction of high-resolution daily rainfall fields for multiple extents: The potential of operational radar, J. Hydrometeorol., 8, 1204–1224, doi:10.1175/2007JHM792.1, 2007. 3089, 3092

Seo, D.-J. and Breidenbach, J.: Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements, J. Hydrometeorol., 3, 93–111, doi:10.1175/1525-7541(2002)003<0093:RTCOSN>2.0.CO;2, 2002. 3089

Seo, D.-J., Breidenbach, J., and Johnson, E.: Real-time estimation of mean field bias in radar rainfall data, J. Hydrol., 223, 131–147, doi:10.1016/S0022-1694(99)00106-7, 1999. 3089

Steiner, M., Smith, J., Burges, S., Alonso, C., and Darden, R.: Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation, Water Resour. Res., 35, 2487–2503, 1999. 3089

van de Beek, C. Z., Leijnse, H., Torfs, P. J. J. F., and Uijlenhoet, R.: Climatology of daily rainfall semi-variance in The Netherlands, Hydrol. Earth Syst. Sci., 15, 171–183, doi:10.5194/hess-15-171-2011, 2011. 3092

Verworn, A. and Haberlandt, U.: Spatial interpolation of hourly rainfall – effect of additional information, variogram inference and storm properties, Hydrol. Earth Syst. Sci., 15, 569–584, doi:10.5194/hess-15-569-2011, 2011. 3092

Villarini, G., Mandapaka, P., Krajewski, W., and Moore, R.: Rainfall and sampling uncertainties: A rain gauge perspective, J. Geophys. Res.-Atmos., 113, D11102, doi:10.1029/2007JD009214, 2008. 3089

HESSD

9, 3087–3127, 2012

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀◀ | ▶▶
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Vischel, T., Lebel, T., Massuel, S., and Cappelaere, B.: Conditional simulation schemes of rain fields and their application to rainfallrunoff modeling studies in the Sahel, J. Hydrol., 375, 273–286, 2009. 3090, 3093

Wackernagel, H.: Multivariate geostatistics: an introduction with applications, Springer, 2003. 3092, 3096

Webster, R. and Oliver, M.: Geostatistics for environmental scientists, Statistics in practice, John Wiley & Sons, 2001. 3089, 3090

Weerts, A. H. and El Serafy, G. Y. H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, 1–17, doi:10.1029/2005WR004093, 2006. 3090, 3103

Willems, P.: Stochastic description of the rainfall input errors in lumped hydrological models, Stoch. Environ. Res. Risk Assess., 15, 132–152, 2001. 3089

Zawadzki, I.: Errors and fluctuations of raingauge estimates of areal rainfall, J. Hydrol., 18, 243–255, 1973. 3089

HESSD

9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.

Table 1. Characteristics of three representative rainfall events.

Starting time	Duration [h]	Speed [$km\,h^{-1}$]	Mean[a] [mm]	St. dev.[b] [mm]
22 Oct 2002	10	54	12.2	7
22 Dec 2002	10	21	16.3	3.9
1 Jan 2003	10	33	17.5	5

[a]Mean of the precipitation sums for all 42 rain gauges.
[b]Standard deviation of the 42 precipitation sums.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.

Table 2. Synthetic experiment. Fitted sills (mm) using spherical model for direct and cross-variograms for rainfall fields 3–6 (see Fig. 7) derived for the dense synthetic observation network (Fig. 8c).

		Rainfall field			
		3	4	5	6
Rainfall field	3	0.31			
	4	0.16	0.32		
	5	−0.03	0.08	0.33	
	6	−0.03	−0.03	0.16	0.31

HESSD
9, 3087–3127, 2012

Table 3. Synthetic experiment. Fitted y-intercept and horizontal asymptote values of coefficients of variation (CV_J, in Fig. 13) for different rain gauge networks (Fig. 6b).

Number of gauges	y-intercept	Horizontal asymptote
14	0.5	1
27	0.45	0.9
100	0.4	0.7

HESSD
9, 3087–3127, 2012

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

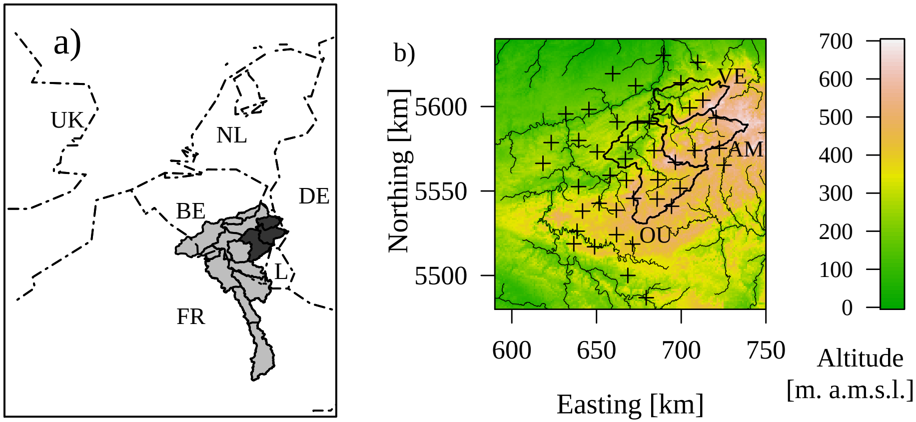

Fig. 1. (a) The Meuse River basin upstream of Borgharen (~21 000 km^2), the entrance point into the Netherlands, and its 15 sub-catchments (gray polygons) including the Upper Ourthe, Ambleve and Vesdre sub-catchments (black polygons). **(b)** Topographic map of the Belgian Ardennes including the Upper Ourthe (OU), Ambleve (AM) and Vesdre (VE) sub-catchments, the channel network (blue lines), rain gauges (plusses). Projection is in the Universal Transverse Mercator (UTM) 31N coordinate system. After Hazenberg et al. (2011).

HESSD

9, 3087–3127, 2012

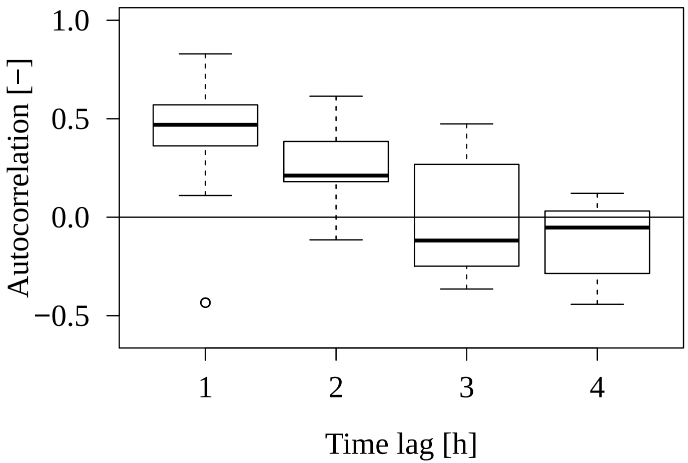

Fig. 2. Autocorrelation coefficients of the catchment averaged precipitation of the Upper Ourthe for 30 rainy events between 1 October 2002 and 31 March 2003.

HESSD
9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.

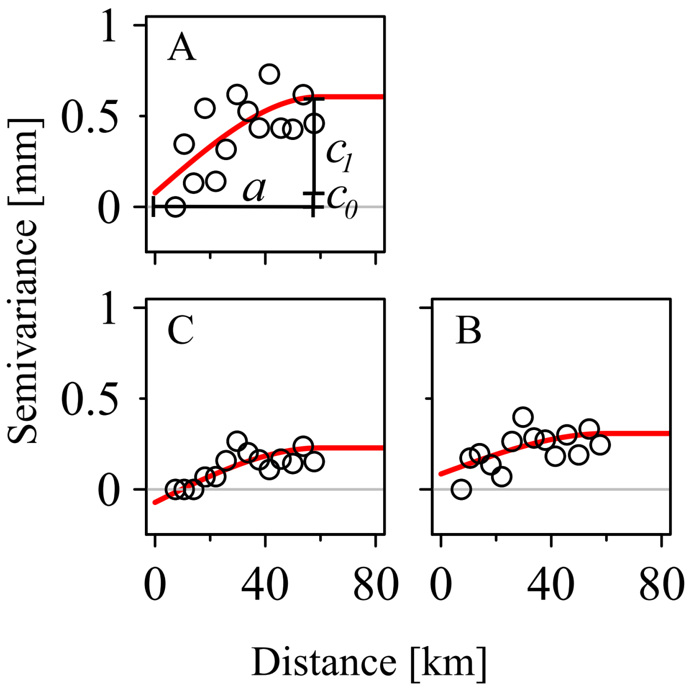

Fig. 3. Two examples – **(A)** and **(B)** – of an experimental variogram (black circles) and the fitted spherical variogram model (red curve) of square-root transformed rain gauge observations. **(C)** Experimental cross-variogram (circles) and fitted spherical cross-variogram model corresponding to **(A)** and **(B)**.

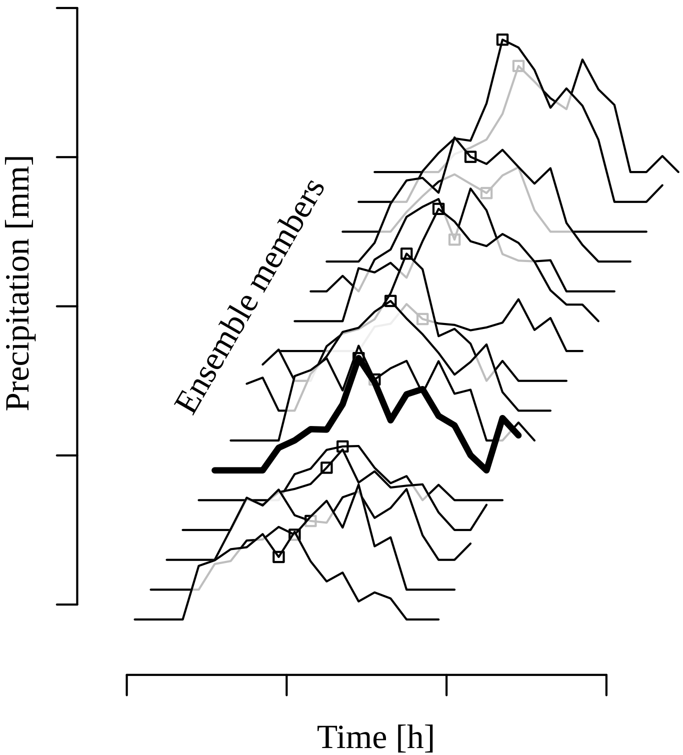

Fig. 4. Two types of evaluating the statistics of precipitation ensembles: hydrological event-based approach (bold line for one ensemble realization) and data assimilation across-ensemble approach (□ for one time).

HESSD
9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.

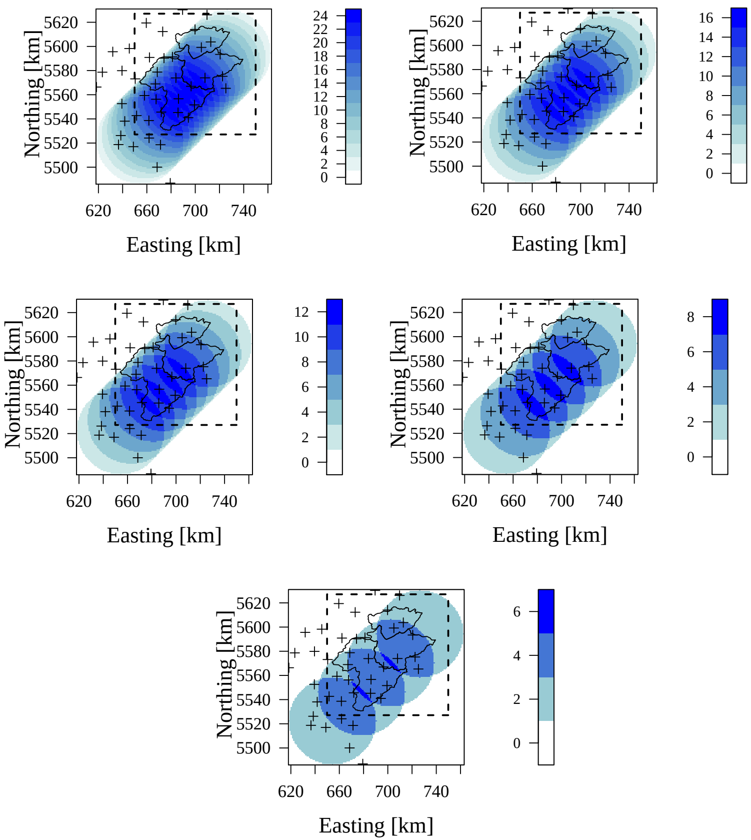

Fig. 5. Synthetic experiment. Five synthetic rainfall events of spherical shape with an intensity of 2 mm h^{-1} and advection speeds of 6, 8, 11, 17, 25 km h^{-1} and duration of 18, 12, 9, 6 and 4 h, respectively (from top panel to bottom panel) over a 145 × 145 km^2 grid with 1 × 1 km^2 resolution. The figures show precipitation sums. The plusses are rain gauges and the dashed box delineates the 100 × 100 km^2 simulation domain.

HESSD

9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.

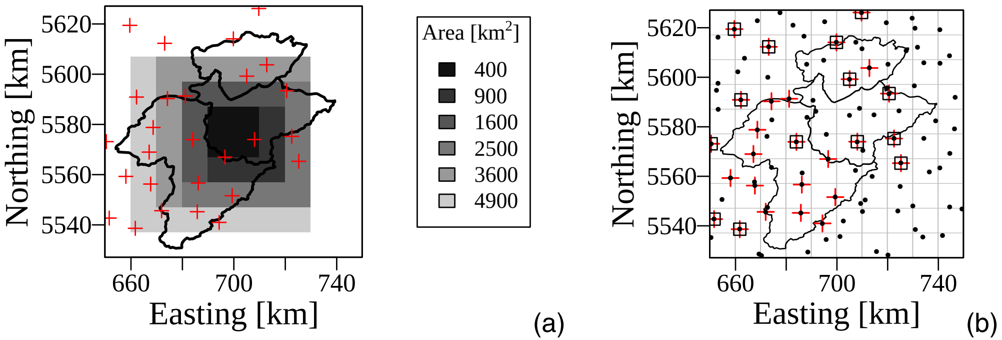

Fig. 6. **(a)** Six nested "sub-catchments" with increasing catchment areas. Rain gauges are indicated by plusses. **(b)** Three types of rain gauge densities: dots ($N = 100$), red pluses ($N = 27$) and squares ($N = 14$). Extent of figure corresponds to the black dashed box in Fig. 5 and the grey lines show the $10 \times 10\,\mathrm{km}^2$ grid resolution.

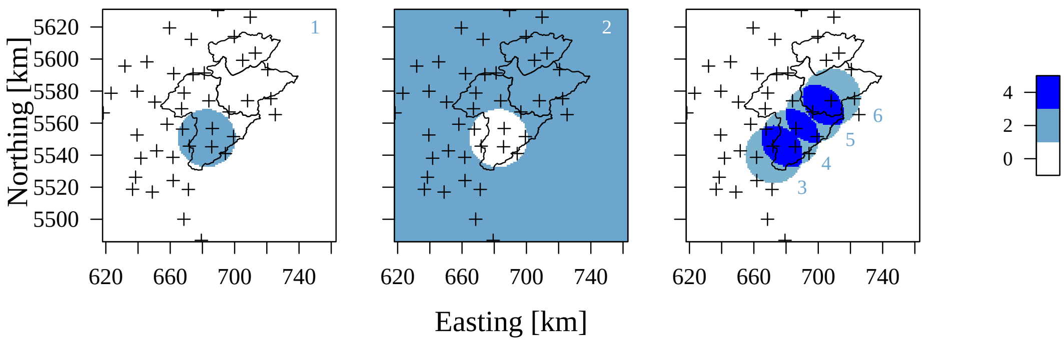

Fig. 7. Synthetic experiment. Spherical rainfall events consisting of six rainfall fields 1–6. Raster resolution is $1 \times 1\,\text{km}^2$. Plusses show the rain gauges and black line delineates the Upper Ourthe, Ambleve and Vesdre catchments.

HESSD
9, 3087–3127, 2012

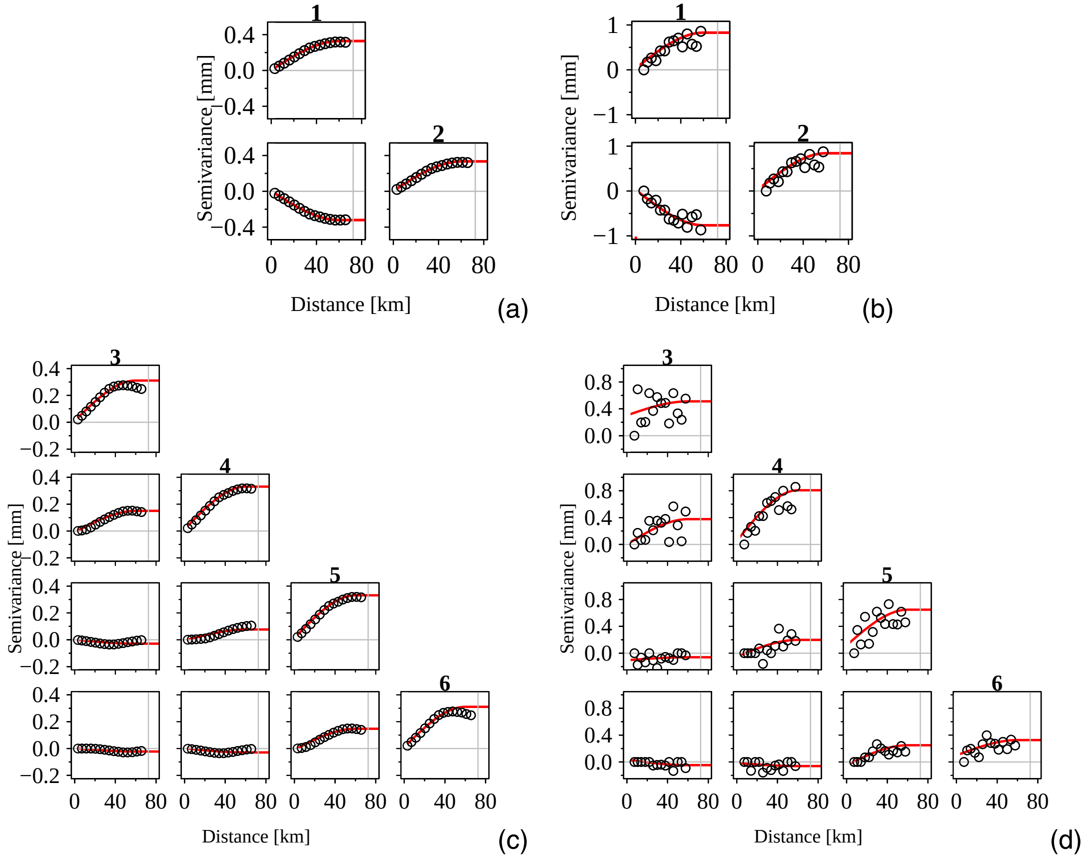

Fig. 8. Synthetic experiment. Direct and cross-variograms for rainfall fields 1 and 2 (Fig. 7) derived from all grid points **(a)** and only the rain gauge grid points **(b)**. Direct and cross-variograms for rainfall fields 3–6 (Fig. 7) derived from all the grid points **(c)** and only the rain gauge grid points **(d)**. Gray vertical line shows the range corresponding to twice the diameter of the rainfall circle.

HESSD
9, 3087–3127, 2012
Hourly precipitation ensemble generator
O. Rakovec et al.

Title Page
Abstract
Introduction
Conclusions
References
Tables
Figures
◄I
►I
◄
►
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

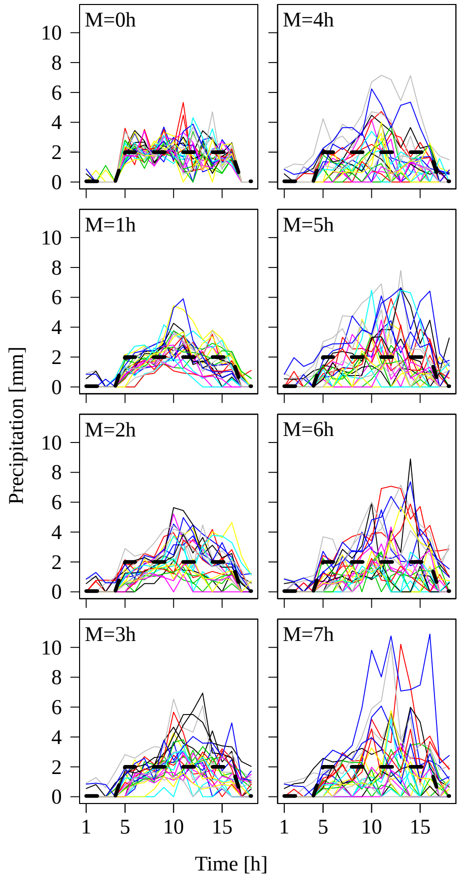

Fig. 9. Synthetic experiment. Precipitation ensemble (24 realizations in colors) for eight types of simulation scenarios given a simulation memory M of 0–7 h, for a pixel located in the centre of the simulation domain. Dashed line stands for the synthetic rain gauge observation. Rainfall scenario for the advection speed of 6 km h^{-1} (see Fig. 5).

HESSD
9, 3087–3127, 2012

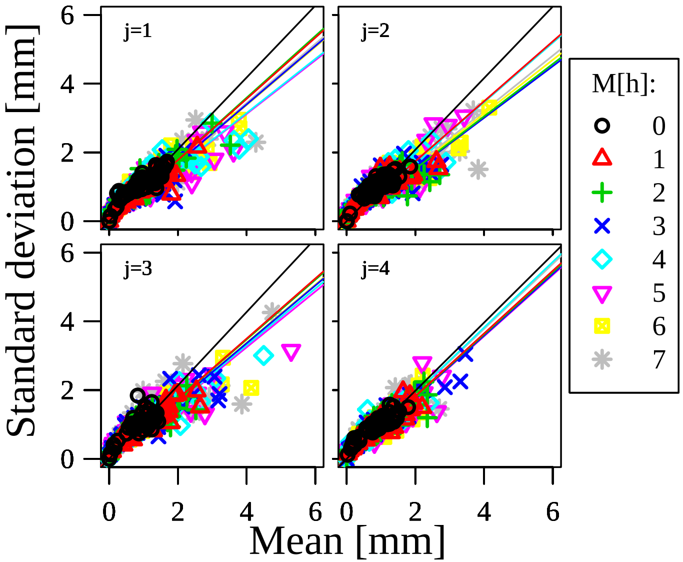

Fig. 10. Synthetic experiment. Scatter plot of $\hat{\mu}_{T,j}$ and $\hat{\sigma}_{T,j}$ with fitted coefficients of variations (CV_T) for four ensemble realizations, an advection speed of 6 km h^{-1} and a catchment area of 4900 km^2.

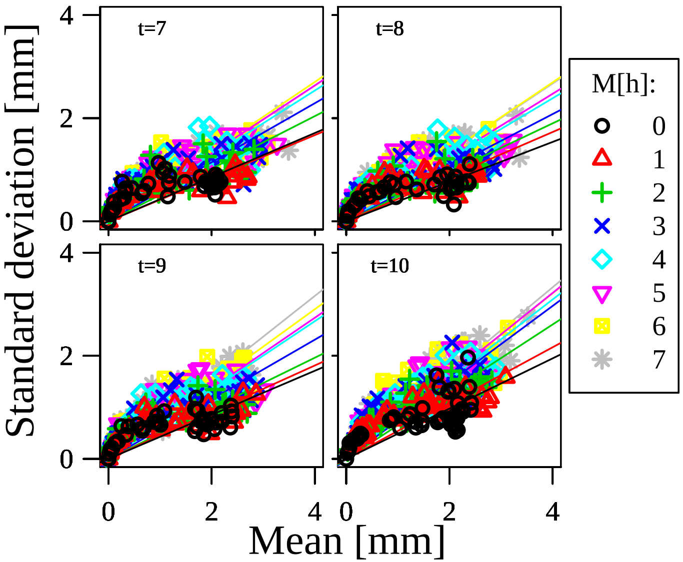

Fig. 11. Synthetic experiment. Scatter plot of $\hat{\mu}_{J,t}$ and $\hat{\sigma}_{J,t}$ with fitted coefficients of variations (CV_J) for four ensemble realizations, an advection speed of 6 km h^{-1} and a catchment area of 4900 km^2.

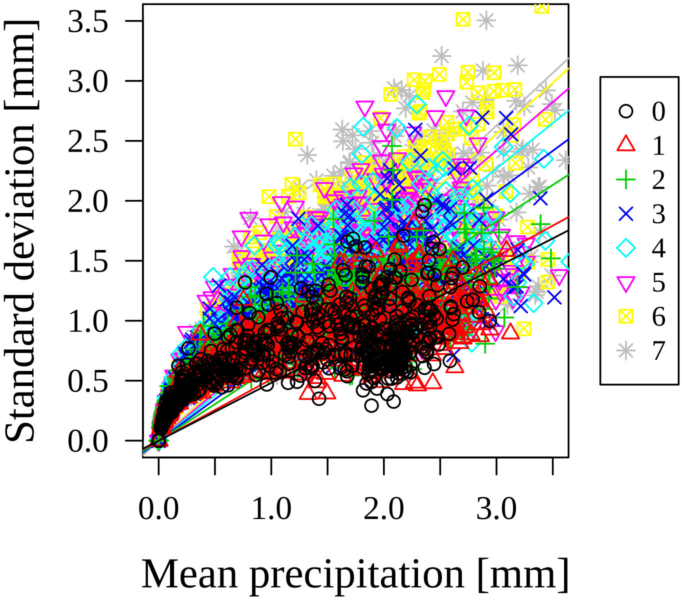

Fig. 12. Synthetic experiment. Scatter plot $\hat{\mu}_{J,t}$ and $\hat{\sigma}_{J,t}$ lumped over time with fitted coefficients of variations (CV_J) for the rainfall scenario with an advection speed of 6 km h^{-1} (see Fig. 5) and a catchment area of 4900 km^2.

HESSD

9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.

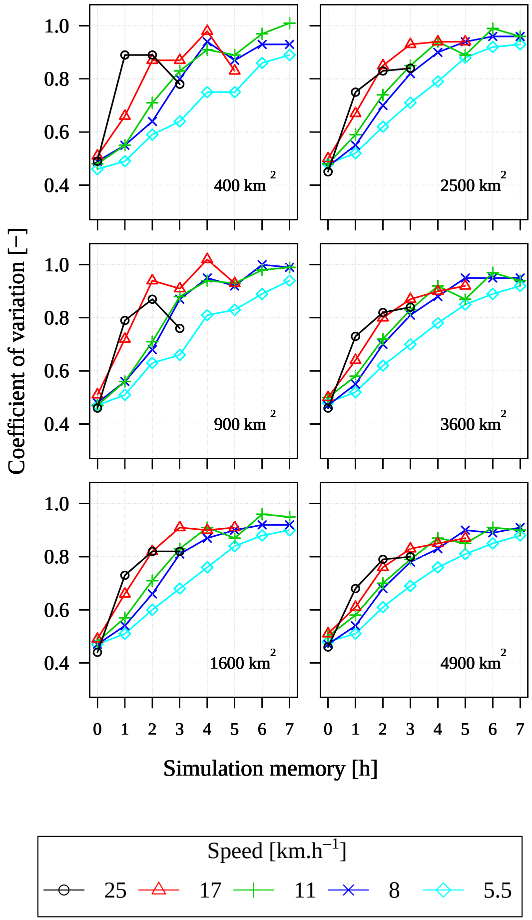

Fig. 13. Synthetic experiment. Coefficient of variation (CV_J) for different catchment sizes and advection speeds.

HESSD
9, 3087–3127, 2012

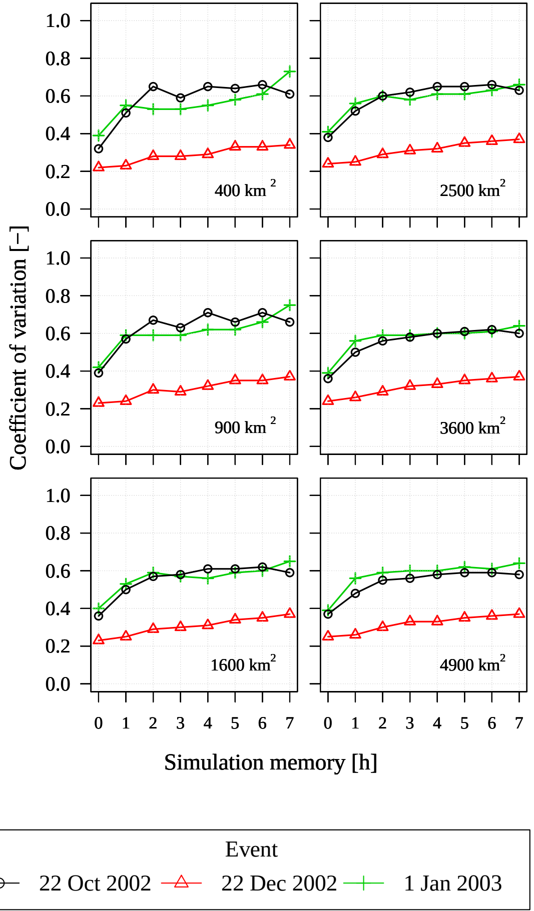

Fig. 14. Real world experiment. Coefficient of variation (CV_J) for different catchment sizes and three rainfall events.

HESSD
9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.

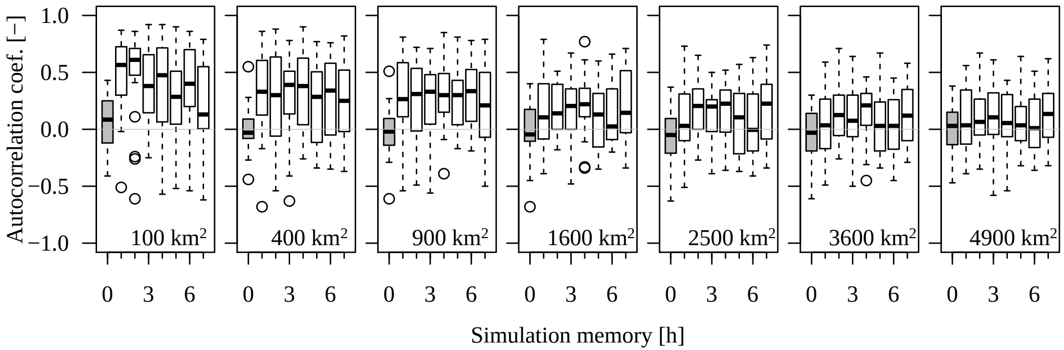

Fig. 15. Synthetic experiment. Boxplots of the first order autocorrelation coefficients (24 realizations) for the time series of the catchment's areal precipitation for different catchment sizes and simulation memories (time-independent univariate simulation in grey, time-dependent multivariate simulations in blank), rainfall scenario with an advection speed of 6 km h^{-1} and duration of 18 h (see Fig. 5).

HESSD
9, 3087–3127, 2012

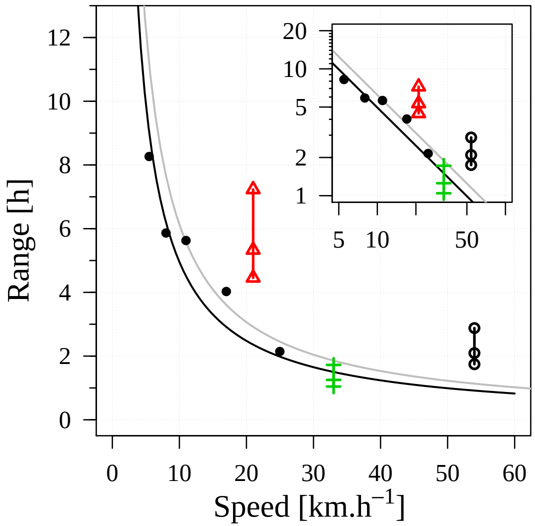

Fig. 16. Non-linear relation between the advection speed and the threshold range for a catchment area of 4900 km^2. Black filled circles represent five synthetic spherical rainfall cells. Black line delineates Eq. (10). Black open circle, red triangle and green plus show three real world rainfall events (see Fig. 14) with the values of ranges for 100 %, 90 % and 80 % (from top to bottom) of the corresponding partial sills. Gray line delineates the fitted non-linear relation for synthetic rainfall cells with double diameter. The identical figure with logarithmic axes is given in the inset.

HESSD
9, 3087–3127, 2012

Hourly precipitation ensemble generator

O. Rakovec et al.