European Water 57: 299-306, 2017.

Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models

J.A. Torres-Matallana[1,3*], U. Leopold[2] and G.B.M. Heuvelink[3]
[1] *Luxembourg Institute of Science and Technology, Department for Environmental Research and Innovation, Chemical Pollution Control and Impact Assessment Unit, 41 Rue du Brill, L-4422, Belvaux, Luxembourg*
[2] *Luxembourg Institute of Science and Technology, Department for Environmental Research and Innovation, eScience Unit, 41 Rue du Brill, L-4422, Belvaux, Luxembourg*
[3] *Wageningen University and Research Centre, Soil Geography and Landscape Group, PO Box 47, 6700AA, Wageningen, The Netherlands*
[*] *e-mail: arturo.torres@list.lu*

Abstract: Precipitation is the most active flux and major input of hydrological systems. Precipitation controls hydrological states (soil moisture and groundwater level), and fluxes (runoff, evapotranspiration and groundwater recharge). Hence, precipitation plays a paramount role in urban water systems. It controls the fluxes towards combined sewer tanks and the dilution of chemical and organic compounds in the wastewater. Furthermore, small catchments (i.e., areas of about 20 ha) have a fast response to precipitation. Therefore, catchment average precipitation is a key component in urban water models. However, average catchment precipitation is not always accurately known when measured at rain gauges, because the location of the gauge might be outside the catchment boundaries or does not reflect the entire catchment. The objective of this paper is to develop a method to estimate the precipitation in a catchment given a known precipitation time series at a location outside of the catchment, while quantifying the uncertainty associated with the estimation. We developed a multivariate autoregressive time series model for conditional simulation of precipitation time series. The case study is a small sub-catchment (16.5 ha) in Luxembourg. The time series of precipitation outside of the sub-catchment are available for two stations and cover the year 2010. We calibrated the model using the R-package 'mAr' and applied the developed conditional simulation algorithm to generate multiple realisations of precipitation time series. The results show that the proposed method is suitable to estimate time series of precipitation at ungauged sites and can quantify the associated uncertainty.

Key words: multivariate model, autoregressive model, precipitation uncertainty, input uncertainty, time series simulation

1. INTRODUCTION

Precipitation is the most important flux and major input of hydrological systems. Precipitation controls hydrological states as soil moisture and groundwater level, and fluxes as runoff, evapotranspiration and groundwater recharge (Guan *et al.*, 2009). Hence, precipitation plays a paramount role in urban water systems. It controls the fluxes towards storage tanks of combined sewer overflows (CSOs) and the dilution of chemical, organic and biological compounds in the wastewater. Furthermore, small catchments (i.e. areas of about 20 ha or smaller) have a fast response to precipitation input. Therefore, catchment average precipitation is a key component in urban water models. However, average catchment precipitation is not always accurately known when measured at rain gauges, because the location of the gauges might be outside the catchment boundaries or do not reflect the entire catchment at one location.

The objective of this paper is to develop and present a method to estimate the precipitation in a specific catchment given a known precipitation time series in a location outside of the catchment under analysis, while also quantifying the uncertainty associated with the estimation. After model implementation, simulated precipitation time series can then be used in urban water models for Monte Carlo uncertainty propagation analysis to reach more robust results for urban water system design and associated potential environmental and economic impacts.

2. MATERIALS AND METHODS

2.1 Multivariate autoregressive time series modelling

The set-up of the case study provides two measured time series of precipitation, one at the rain gauge 1, *RM(t)*, inside the catchment, and one outside the catchment, rain gauge 2, *RM2(t)* (see Figure 1). However, we need to estimate the precipitation at the sub-catchment level, *R(t)*. We model the precipitation in the sub-catchment as:

$$R(t) = RM(t) \cdot \delta(t) \tag{1}$$

where $\delta(t)$ is defined as the ratio of *RM(t)* and *R(t)*, the statistical properties of which are derived using measured time series of *RM(t)* and *RM2(t)*. We assume that *R(t)*, *RM(t)* and $\delta(t)$ are log-normally distributed stochastic processes, so that:

$$\log[R(t)] = \log[RM(t)] + \log[\delta(t)] \tag{2}$$

which we write as :

$$LR(t) = LRM(t) + L\delta(t) \tag{3}$$

LR(t), *LRM(t)* and $L\delta(t)$ are modelled using a first-order multivariate autoregressive process model. Given Equation 3, we only need to model *LRM(t)* and $L\delta(t)$ as (Luetkepohl, 2005):

$$\begin{bmatrix} LRM(t+1) \\ L\delta(t+1) \end{bmatrix} = \begin{bmatrix} \mu_R \\ \mu_\delta \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \left(\begin{bmatrix} LRM(t) \\ L\delta(t) \end{bmatrix} - \begin{bmatrix} \mu_R \\ \mu_\delta \end{bmatrix} \right) + \begin{bmatrix} \varepsilon_R(t+1) \\ \varepsilon_\delta(t+1) \end{bmatrix} \tag{4}$$

where μ_R = E(*LRM*); μ_δ = E($L\delta$); A_{11}, A_{12}, A_{21}, A_{22} are the coefficients of the autoregressive (AR) model; ε_R and ε_δ are zero-mean, normally distributed white noise processes.

We need to calibrate this model, i.e. estimate the parameters μ_R, μ_δ, A_{11}, A_{12}, A_{21}, A_{22}, σ_R^2, σ_δ^2 and $\rho_{R\delta}$, where σ_R^2 = Var(ε_R), σ_δ^2 = Var(ε_δ), and $\rho_{R\delta}$ is the correlation between ε_R and ε_δ. The parameters σ_R^2, σ_δ^2 and $\rho_{R\delta}$ are needed to account for cross-correlation. We use the R package 'mAr' (Barbosa, 2015) to calibrate the model given the two time series *LRM* and $L\delta$. For this, we derive $L\delta$ by taking the difference of *LRM* and *LRM2*, in other words we assume that the time series *RM* and *R* have similar multivariate behaviours as time series *RM* and *R*. Note that the rain gauge associated with time series *RM2(t)* is about the same distance from the first rain gauge as the sub-catchment.

We derive a time series of δ by dividing the *RM* data by the *RM2* data, for times when both *RM* > 0 and *RM2* > 0, i.e. we create a bivariate time series of *RM* and δ. Next, we generate time series *LRM(t)* and $L\delta(t)$ by taking log[*RM(t)*] and log[*RM2(t)/RM(t)*].

2.2 Conditional time series simulation

Given the calibrated model we need to simulate from $L\delta(t)$. This simulation should be conditional to *LRM*. We define:

$$X_1(t) = LRM(t) - \mu_R; \qquad \varepsilon_1(t) = \varepsilon_R(t) \tag{5}$$

$$X_2(t) = L\delta(t) - \mu_\delta; \qquad \varepsilon_2(t) = \varepsilon_\delta(t) \tag{6}$$

and therefore we have:

$$\begin{bmatrix} X_1(t+1) \\ X_2(t+1) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t+1) \\ \varepsilon_2(t+1) \end{bmatrix} \tag{7}$$

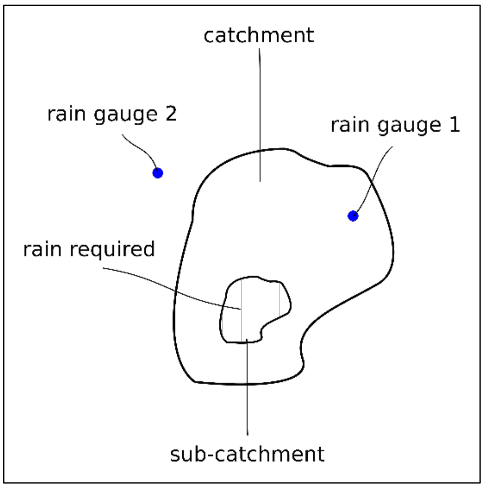

Figure 1. Schematic definition of the case study set-up. Rain gauge 1 = RM(t); rain gauge 2 = RM2(t); rain required = R(t)

At each time step we need to simulate $X_2(t+1)$ given $X_1(t)$, $X_1(t+1)$ and $X_2(t)$. If we assume $\rho_{R\delta} = 0$, then $X_2(t+1)$ is given by:

$$X_2(t+1) = A_{21} \cdot X_1(t) + A_{22} \cdot X_2(t) + \varepsilon_2(t+1) \tag{8}$$

where $X_1(t)$, $X_2(t)$, A_{21}, A_{22} are known, and $\varepsilon_1(t+1)$ is a normally distributed white noise process with mean 0 and variance σ_2^2. If we assume $\rho_{R\delta} \neq 0$, this case is not straightforward, because $X_1(t+1)$ and $X_2(t+1)$ are also 'directly' correlated so that $X_1(t+1)$ has to be included in the conditional distribution. We can write:

$$Y = \begin{bmatrix} X_1(t) \\ X_1(t+1) \\ X_2(t) \\ \hline X_2(t+1) \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \tag{9}$$

with

$$Y_1 = \begin{bmatrix} X_1(t) \\ X_1(t+1) \\ X_2(t) \end{bmatrix}; \qquad \text{and} \qquad Y_2 = \begin{bmatrix} X_2(t+1) \end{bmatrix} \tag{10}$$

Y follows a multivariate normal distribution with mean vector μ and variance-covariance matrix Σ (Box *et al.,* 2008):

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \tag{11}$$

where μ_1 = 3x1, μ_2 = 1x1, Σ_{11} = 3x3, Σ_{12} = 3x1, Σ_{21} = 1x3, Σ_{22} = 1x1. Solving Equation 11, we then know:

$$\{Y_2|Y_1 = a\} \sim N\left(\mu_2 + \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot (a - \mu_1), \qquad \Sigma_{22} - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot \Sigma_{12}\right) \tag{12}$$

so we can simulate from $Y_2 = X_2(t+1)$ by sampling from this conditional normal distribution. Therefore, we need to derive vector $[\mu_1, \mu_2]$ and the variance-covariance matrix Σ. The first is straightforward because we centred X_1 and X_2 on zero:

$$\begin{bmatrix} \mu_1 \\ \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{13}$$

Regarding the calculation of the variance-covariance matrix Σ, we assume stationarity for variances and covariances, not depending on time, so the initial effect fades out. Therefore, we can define:

$$\sum = \begin{bmatrix} C_{11} & C_{12} & C_{13} & |C_{14} \\ C_{21} & C_{22} & C_{23} & |C_{24} \\ C_{31} & C_{32} & C_{33} & |C_{34} \\ - & - & - & - \\ C_{41} & C_{42} & C_{43} & |C_{44} \end{bmatrix} \tag{14}$$

We can demonstrate that the components of the Σ matrix are defined as:

$$C_{11} = C_{22} = \frac{A_{12}^2}{1 - A_{11}^2} \cdot \frac{A_{21}^2\sigma_1^2 + \sigma_2^2}{A_{11}^2 + A_{11}^2 A_{22}^2 - A_{12}^2 A_{21}^2} + \sigma_1^2 = \mathrm{Var}(X_1) \tag{15}$$

$$C_{33} = C_{44} = \frac{A_{21}^2\sigma_1^2 + \sigma_2^2}{A_{11}^2 + A_{11}^2 A_{22}^2 - A_{12}^2 A_{21}^2} = \mathrm{Var}(X_2) \tag{16}$$

$$C_{13} = C_{31} = C_{24} = C_{42} = \frac{A_{11}A_{21}\mathrm{Var}(X_1) + A_{12}A_{22}\mathrm{Var}(X_2) + \rho\sigma_1\sigma_2}{1 - A_{11}A_{22} - A_{12}A_{21}} = \mathrm{Cov}(X_1, X_2) \tag{17}$$

$$\rho\sigma_1\sigma_2 = \mathrm{Cov}(\varepsilon_1, \varepsilon_2) \tag{18}$$

$$C_{12} = C_{21} = A_{11}\mathrm{Var}(X_1) + A_{12}\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}[X_1(t+1), X_1(t)] \tag{19}$$

$$C_{34} = C_{43} = A_{21}\mathrm{Cov}(X_1, X_2) + A_{22}\mathrm{Var}(X_2) = \mathrm{Cov}[X_2(t+1), X_2(t)] \tag{20}$$

$$C_{23} = C_{32} = A_{11}\mathrm{Cov}(X_1, X_2) + A_{12}\mathrm{Var}(X_2) = \mathrm{Cov}[X_1(t+1), X_2(t)] \tag{21}$$

$$C_{14} = C_{41} = A_{21}\mathrm{Var}(X_1) + A_{22}\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}[X_2(t+1), X_1(t)] \tag{22}$$

2.3 Case study

The case study is located at Goesdorf, a small sub-catchment (16.5 ha) of the Haute-Sure system in the Northwest of Luxembourg. We know the time series of precipitation outside of the Goesdorf sub-catchment at two locations: Dahl (around 2 Km from the CSO tank Goesdorf) and Esch-sur-Sûre (around 3.5 km from the CSO tank Goesdorf). The time series of precipitation provided by the Luxemburgish *Administration des services techniques de l'agriculture* (ASTA) (http://www.asta.etat.lu), covers the year 2010 with measurements at 10 minutes resolution. The precipitation stations are provided with Lambrecht 15188 tipping bucket rain gauges with a resolution of 0.1 mm per tip, and a surface of the round reception area of 200 cm^2. We have 52,556 observations for each time series. The total precipitation at Esch-sur-Sûre is 658.6 mm, and 758.7 mm at Dahl. The time series of precipitation were validated by the Observatory for Climate and Environment (OCE) of the Luxembourg Institute of Science and Technology (LIST).

3. RESULTS

3.1 Multivariate autoregressive time series modelling

Two observed ASTA time series, Esch-sur-Sûre and Dahl (Figure 2), were used for the calibration of the multivariate autoregressive model. Before we could calculate of $\delta(t)$ (Eq. 1) we applied a Daniell Kernel (R Core Team and contributors worldwide, 2017) to smooth the time series and avoid the sudden tipping bucket effect in the measurements (Table 1). Only those values of precipitation above 0.1 mm were smoothed. Then, the time series were filtered and posteriorly computed the ratio between them. The filter creates two new time series for the cases where the value of precipitation is higher than 0.01 mm for both time series to compute the ratio. The length of the resulting filtered time series was 6,454 observations. Finally, the ratio between the time series was computed.

Table 1. Daniell Kernel for smoothing the observed time series of precipitation

Index	Factor
-2	0.1111
-1	0.2222
0	0.3333
1	0.2222
2	0.1111

We defined the log-transform of the observed filtered time series, $LRM(t)$, and the ratio, $L\delta(t)$, and checked the normality assumption. The log-transform of the time series is fairly normal for the observed, $LRM(t)$, and the ratio, $L\delta(t)$, time series. Also, we checked the autocorrelation function (ACF) of these time series and both follow a similar pattern.

Given the $LRM(t)$ and $L\delta(t)$ we calibrated the multivariate autoregressive model for order one (Equation 4) using the 'mAr' R-package. Table 2, summarises the calibrated parameters of the model.

3.2 Conditional time series simulation

Upon calibration of the multivariate autoregressive model, we proceeded with the conditional simulation of Y_2 or $X_2(t+1)$ (Equation 12). For this, we compute the Σ matrix given the parameters of the model (Table 2). Table 2 also presents the calculated values of the components of Σ.

Once the components of Σ matrix are calculated, we developed an algorithm for simulating Y_2 given the known values of Y_1 i.e. $X_1(t)$, $X_1(t+1)$ and $X_2(t)$. Then, we added the mean to the time series Y_2, backtransformed the lognormal values to compute the required precipitation time series at the sub-catchment $R(t)$ according to Equation 1. Figure 3 shows the observed, $RM(t)$ at Esch-sur-Sûre, and the simulated, $R(t)$ at Goesdorf, precipitation time series of the case study.

4. DISCUSSION

According to McMillan *et al*. (2011), traditional calibration methods do not take into account input error, which leads to bias in parameter estimation and possible misleading model predictions. Several studies have proposed precipitation multipliers for overcoming this issue and taking into account precipitation input uncertainty in hydrological model calibration and prediction (McMillan *et al.*, 2011; Leta *et al.*, 2015; Del Giudice *et al.*, 2016). Based on data from a dense gauge and radar network, McMillan *et al.* (2011) validated and confirmed the suitability of a multiplicative error model. Moreover, they showed that the lognormal multiplier distribution is a good approximation to the true error characteristics.

Table 2. Calibrated parameters of the multivariate autoregressive model for LRM(t) and Lδ(t) and components of Σ

Parameter	Value	Component	Value
μ_R	2.8550080	C_{11}	0.00639559
μ_δ	0.1019426	C_{33}	0.00388534
A_{11}	0.9564952	C_{13}	-0.00135028
A_{12}	0.0397994	C_{12}	0.00617109
A_{21}	0.0242934	C_{34}	0.00339810
A_{22}	0.8830385	C_{23}	-0.00144617
σ_R^2	0.0724132	C_{14}	-0.00103698
σ_δ^2	0.0795191		
$\rho_{R\delta}$	-0.0387570		

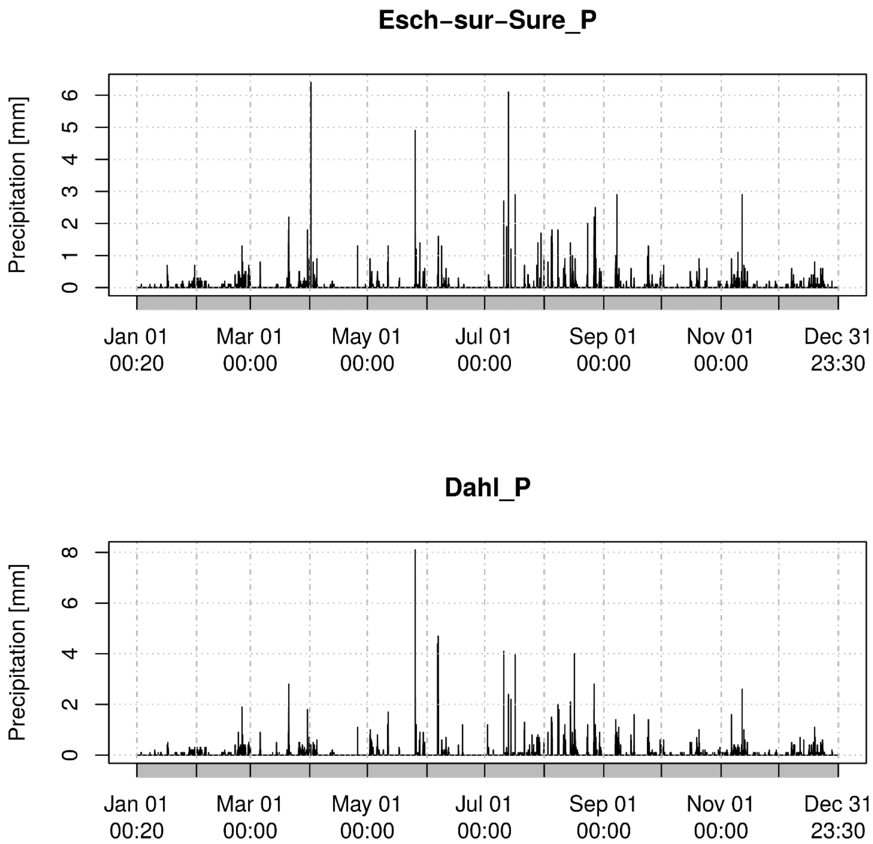

Figure 2. Observed precipitation time series of the case study: Esch-sur-Sûre and Dahl.

We proposed a multivariate autoregressive model for conditional simulation of input precipitation based on a multiplicative error model in the lognormal distribution. This method is essentially the same as the application of a Kalman filter/smoother (Kalman, 1960; Webster and Heuvelink, 2006). From a mathematical-statistical point of view we addressed the same principle of Kalman filter, i.e. to compute the conditional probability distribution given the available time series at each time step, sample from it and move to the next time point. Despite the goodness of the proposed method, some cases show an overestimation of the simulated precipitation, mainly due to high values of the ratio for the multiplicative factor *δ(t)*. This behaviour is also recognised by McMillan *et al*. (2011), who stated that the multiplicative factor used in their study "*does not capture the distribution tails, especially during heavy rainfall where input errors would have important consequences for runoff prediction*"

Note also that we ignore the change-of-support effect because the sub-catchment area is much greater than a point. Future research may address this issue of support.

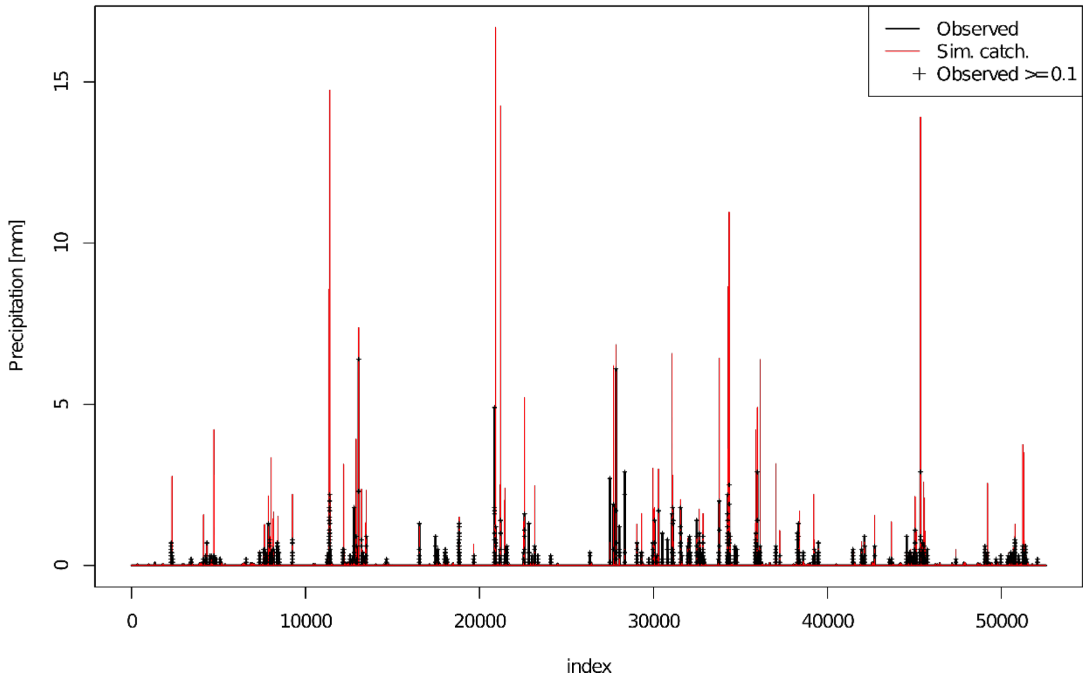

Figure 3. Observed and simulated precipitation time series of the case study: simulation at Goesdorf.

5. CONCLUSIONS

Catchment average precipitation is a major driving force and key component in urban water models. However, average catchment precipitation is not always accurately known when measured at rain gauge, because the location of the gauge might be outside of the catchment boundaries or does not reflect the entire catchment at one location. To overcome this issue, we developed a method to estimate the precipitation in a catchment given a known precipitation time series at a location outside of the catchment, while quantifying the uncertainty associated with the estimation. A first-order multivariate autoregressive model for conditional simulation of input precipitation based on a multiplicative error model was proposed. This method is essentially the same as the application of a Kalman filter/smoother. Although the goodness of the proposed method, some cases show an overestimation of the simulated precipitation due to high values of the ratio for the multiplicative factor. This behaviour is also recognised in the literature.

The simulated precipitation time series can be used in urban water models for uncertainty propagation analysis to reach more robust results for urban water system design and associated potential environmental and economic impacts.

ACKNOWLEDGEMENTS

The work presented is part of the QUICS (Quantifying Uncertainty in Integrated Catchment Studies) project. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No.

607000, and the Luxembourg Institute of Science and Technology (LIST). The authors thank to the Luxemburgish *Administration des services techniques de l'agriculture* (ASTA) for the time series of precipitation provided, and to the Observatory for Climate and Environment (OCE) of the LIST for its technical support.

REFERENCES

Barbosa S.M. 2015. Package 'mAr': Multivariate AutoRegressive analysis. The Comprehensive R Archive Network, CRAN, 1.1-2 edition.

Box G.E.P., Jenkins G.M., Rinsel G.C., 2008. Time series analysis: forecasting and control, 4th ed., John Wiley & Sons, Inc.

del Giudice D., Albert C., Rieckermann J., Reichert P. 2016. Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation. Water Resources Research, 52(4):3162-3186, doi: 10.1002/2015wr017871.

Guan H., Wilson J.L., Xie H. 2009. A cluster-optimizing regression-based approach for precipitation spatial downscaling in mountainous terrain. Journal of Hydrology, 375(3-4):578-588, doi: 10.1016/j.jhydrol.2009.07.007.

Kalman R.E. 1960. A new approach to linear filtering and prediction problems. Transactions of the American Society of Mechanical Engineers: Journal of Basic Engineering, 82D:35–45.

Leta, O.T., Nossent, J., Velez, C., Shrestha, N.K., van Griensven, A. and Bauwens, W., 2015. Assessment of the different sources of uncertainty in a SWAT model of the River Senne (Belgium). Environmental Modelling & Software, 68:129-146.

Luetkepohl H., 2005. New Introduction to Multiple Time Series Analysis. Springer.

McMillan, H., Jackson, B., Clark, M., Kavetski, D. and Woods, R., 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. Journal of Hydrology, 400(1):83-94.

R Core Team and contributors worldwide, 2017. The R Stats Package. The R Project for Statistical Computing, 3.5.0 edition. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html.

Webster, R. and Heuvelink, G.B.M., 2006. The Kalman filter for the pedologist's tool kit. European Journal of Soil Science, 57(6):758-773.