Journal of Hydrology 472–473 (2012) 36–52

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

A formal statistical approach to representing uncertainty in rainfall–runoff modelling with focus on residual analysis and probabilistic output evaluation – Distinguishing simulation and prediction

Anders Breinholt [a,*], Jan Kloppenborg Møller [b], Henrik Madsen [b], Peter Steen Mikkelsen [a]

[a] Department of Environmental Engineering (DTU Environment), Technical University of Denmark, Building 113, DK-2800 Lyngby, Denmark
[b] Department of Informatics and Mathematical Modelling (DTU Informatics), Technical University of Denmark, Building 305, DK-2800 Lyngby, Denmark

ARTICLE INFO

Article history:
Received 3 March 2012
Received in revised form 30 August 2012
Accepted 5 September 2012
Available online 17 September 2012
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:
Conceptual urban drainage model
Infiltration inflow
Maximum likelihood estimation
Output error method
Stochastic differential equations
Interval skill score

SUMMARY

While there seems to be consensus that hydrological model outputs should be accompanied with an uncertainty estimate the appropriate method for uncertainty estimation is not agreed upon and a debate is ongoing between advocators of formal statistical methods who consider errors as stochastic and GLUE advocators who consider errors as epistemic, arguing that the basis of formal statistical approaches that requires the residuals to be stationary and conform to a statistical distribution is unrealistic. In this paper we take a formal frequentist approach to parameter estimation and uncertainty evaluation of the modelled output, and we attach particular importance to inspecting the residuals of the model outputs and improving the model uncertainty description. We also introduce the probabilistic performance measures sharpness, reliability and interval skill score for model comparison and for checking the reliability of the confidence bounds. Using point rainfall and evaporation data as input and flow measurements from a sewer system for model conditioning, a state space model is formulated that accounts for three different flow contributions: wastewater from households, and fast rainfall–runoff from paved areas and slow rainfall-dependent infiltration–inflow from unknown sources. We consider two different approaches to evaluate the model output uncertainty, the output error method that lumps all uncertainty into the observation noise term, and a method based on Stochastic Differential Equations (SDEs) that separates input and model structure uncertainty from observation uncertainty and allows updating of model states in real-time. The results show that the optimal simulation (off-line) model is based on the output error method whereas the optimal prediction (on-line) model is based on the SDE method and the skill scoring criterion proved that significant predictive improvements of the output can be gained from updating the states continuously. In an effort to attain residual stationarity for both the output error method and the SDE method transformation of the observations were necessary but the statistical assumptions were nevertheless not 100% justified. The residual analysis showed that significant autocorrelation was present for all simulation models. We believe users of formal approaches to uncertainty evaluation within hydrology and within environmental modelling in general can benefit significantly from adopting the evaluation measures applied here, so the probabilistic performance of their models can be assessed properly.

1. Introduction

Within rainfall–runoff (RR) modelling many different models exist, common to all of them is however that they are approximative representations of reality, and their outputs should hence be accompanied with an uncertainty estimate. Uncertainties in RR models arise because of errors and biases from spatial and temporal averaging of distributed and heterogeneous model inputs and internal fluxes, model structural limitations and errors as well as observation errors and biases (Renard et al., 2010; Beven et al., 2011; Willems, 2012). However, when trying to deal with uncertainties in RR models we enter a foggy minefield since there is currently no consensus as to how uncertainty should be addressed and quantified, nor to how much physicality our models should accommodate (Todini, 2011; Beven et al., 2011; Clark et al., 2011). For uncertainty quantification two distinct alternatives can be identified. On the one hand we have the formal statistical methods mostly represented by Bayesian approaches that recognise the existence of several probable parameter sets, but also by the frequentist approach that searches for a single best performing parameter set (see comparison of the two methods in Gallagher and Doherty (2007) and Dotto et al. (2009)). On the other hand we have the Generalized Likelihood Uncertainty Estimation (GLUE)

* Corresponding author.
E-mail address: anbre@env.dtu.dk (A. Breinholt).

0022-1694/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2012.09.014

methodology (Beven and Binley, 1992; Beven, 2006; Beven et al., 2011), that questions and rejects the validity of formal residual assumptions due to the existence of epistemic errors and non-stationary model-residuals that are unsuitable for statistical likelihood functions.

When evaluating model residuals it is important to distinguish a "prediction model" from a "simulation model". A model tailored for simulation (referred to also as an off-line model) is suitable for long-term investigations and should describe the uncertainty of the outputs over a longer period, whereas a model tailored for prediction or forecasting (referred to also as a real-time- or on-line model) by accommodating an updating step (referred to also as a data-assimilating step) should outdo the simulation model on the short-term by generating more narrow but reliable confidence bounds (Refsgaard, 1997) in the case when real-time observations are available. Quite commonly the terms prediction or forecasting are used in a context where they actually refer to simulation. As noted by Andréassian et al. (2007) past misunderstandings on the uncertainty estimation issue would have been avoided if authors had clearly defined what type of model application they were discussing. That simulation and prediction models are fundamentally different was the subject of debate in (Beven, 2009b; Vrugt et al., 2009b) following a paper by Vrugt et al. (2009a) in which the prediction uncertainty of a formal statistical (Bayesian) approach was compared to the simulation uncertainty of an informal (GLUE) approach and used to conclude that the Bayesian approach gave smaller spread and higher coverage than the GLUE generated bounds. This is however an inappropriate comparison of uncertainty evaluation methods because in the prediction model an updating step ensures that the model is continuously updated according to the measurements, whereas in the simulation model there is no updating step. If the two uncertainty methods are to be fairly compared they should thus be compared in either simulation or prediction mode (Beven, 2009b) or it should at least be specified that a comparison is being made between a prediction and a simulation model.

To be able to assess and compare models and uncertainty methods we need a common yardstick. According to Montanari (2007) the typical information an end user would require is an indication of a confidence (or a prediction) interval associated with some modelled output and hence this interval should be evaluated. However some authors choose only to show plots of uncertainty intervals and observations for visual evaluation (Yang et al., 2007; McMillan et al., 2009) while others prefer quantitative performance measures. Common practice when comparing uncertainty intervals (umbrella term for prediction and simulation intervals) is to include the width of the bounds and the coverage of observations (Vrugt et al., 2009a), also referred to as sharpness and reliability (Georgakakos et al., 2004; Smith and Marshall, 2010; Engeland et al., 2010), or related indices such as the Average Relative Interval Length (ARIL) (Jin et al., 2010), the r-factor and the Continuos Rank Probability Score (CRPS) (Yang et al., 2008). In a GLUE study Xiong et al. (2009) introduces seven measures (they call them indices) for characterising the prediction bounds from different perspectives and suggest that they be employed for assessing and comparing the uncertainty bounds in a more comprehensive and objective way. These measures take the coverage, the width of the bounds and the symmetry of the bounds into account; however none of them can individually be used to determine which model of several candidate models that provides the best predictions/simulations and clearly there seems to be a need for such a measure. We evaluate the derived uncertainty bounds using the reliability, sharpness and interval skill scoring criterion of Gneiting and Raftery (2007) and Thordarson et al. (2012) and we compare the performance of a model tailored for prediction and a model tailored for simulation to clearly distinguish the difference between the two.

In this paper we take a formal frequentist approach to parameter estimation and draw special attention to evaluating whether the residuals can be properly represented by statistical likelihood functions. This is an important but often neglected analysis by authors who apply formal statistical approaches for model conditioning (Beven et al., 2011). The modelling philosophy adopted here is that of the grey box modelling principle that unites prior physical knowledge with information from data, and uses statistical tools for parameter significance testing and estimation and thus adhere to the principle of parsimony (Kristensen et al., 2004a; Kristensen et al., 2004b; Breinholt et al., 2011).

As rainfall–runoff modelling case we consider a conceptual model of a sewer system that receives wastewater from households, fast rainfall–runoff from paved areas and slow rainfall-dependent infiltration–inflow from unknown sources, and we use rain and evaporation data as inputs and flow measurements for calibration and evaluation. Due to a poor representation of the spatial rainfall that is measured by point gauges, the complexity of the sewer system and the three flow contributions that are solely observed by a flow meter downstream from the catchment, large output uncertainty is expected.

This paper is organised as follows. After this brief introduction we present the case study area, the calibration and validation data, and the deterministic model in Section 2. This is followed by an elaboration of the applied uncertainty analysis methodology in section 3 that distinguishes two likelihood functions to be used for parameter estimation, one that is tailored for prediction models, and one that is tailored for simulation models. Then Section 4 shows how the confidence bounds can be obtained in prediction and simulation mode and the probabilistic performance measures reliability, sharpness and interval skill score are defined. Section 5 evaluates the statistical assumptions of the models and uses this information for developing transformations that allow satisfying the formal statistical assumptions. Finally the models' probabilistic performances are compared in Section 6 using the interval skill scoring criterion, the optimal simulation model is found and the performance of the simulation models are compared to that of a prediction model. Conclusions are drawn in Section 7.

2. Case study and deterministic model

2.1. Catchment, drainage system and measurements

The catchment is located in the western part of greater Copenhagen in Ballerup Municipality, see Fig. 1. The drainage system consists of both combined and separate sewer pipes, however most of the catchment is separated, i.e. wastewater and stormwater runoff is drained in separate pipes. Some catchment characteristics are given in Table 1. The drainage system contains a few detention basins and some pumping stations. When the pipe capacity is exceeded water is stored in the basins and when the storage capacity is exceeded wastewater overflows from the sewer system. A flow meter has been installed downstream the catchment with the aim to detect unintended infiltration inflow contribution. The flow meter is placed in an intercepting concrete pipe ($d = 1.4$ m) and logs every 5 min. The flow meter is an Isco[1] 2150 area-velocity type flow module that uses continuous wave Doppler technology to measure the mean velocity over the cross section area. A differential pressure transducer in the sensor measures liquid depth to determine the flow area and the flow rate is then calculated by multiplying the area of the flow stream by its average velocity.

The two nearest rain gauges from the national Danish tipping bucket network (Jørgensen et al., 1998) with approved measure-

[1] www.isco.com.

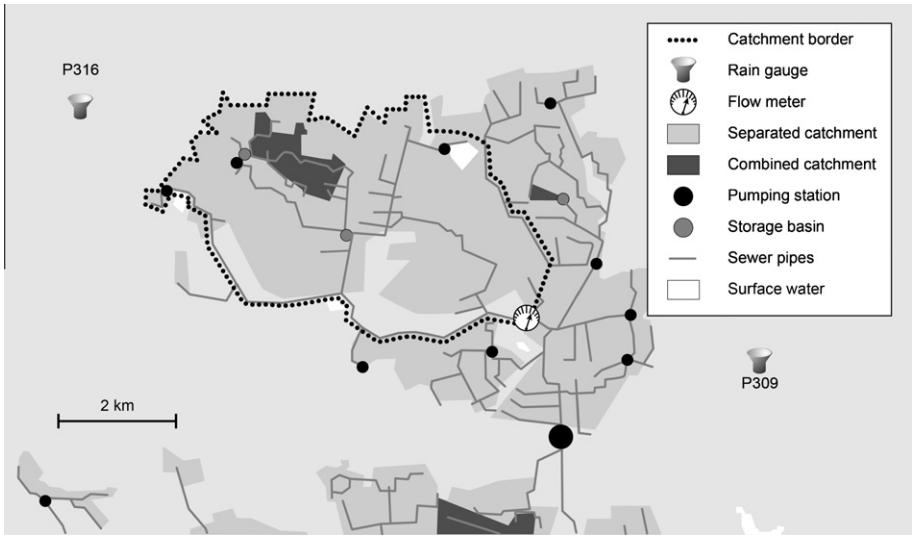

Fig. 1. The Ballerup catchment area.

Table 1
Catchment characteristics (Tomicic, 2006).

Ballerup Municipality	Total area		Imp. area	
	(km^2)	(%)	(km^2)	(%)
Combined	0.92	7	0.33	77
Separated	12.27	93	0.10	23
Total	13.20	100	0.43	100

ments from the considered period are P_{316} and P_{309}, see Fig. 1. Both are located outside the studied catchment area approximately 12 km apart.

Monthly evaporation data was available from 2007 but although we consider a 2010 period the data are still considered useful since the monthly evaporation is not expected to vary that much over the years. The inputs and observation data used for this study are presented in Fig. 2. For use in the models which run in 15-min time steps the evaporation data was linearly interpolated between the months, rain data was summed in 15 min windows and every third flow measurement was used. The same data period are used for both calibration of parameters and for generating simulation and prediction bounds, i.e. for validation. For discussion of the analysis performed and the results we found it beneficial to distinguish between two weather regimes; dry and wet weather periods. The regimes were differentiated from each other by a flow threshold of 450 m^3/h.

2.2. Deterministic hydrological model

The conceptual hydrological model is depicted in Fig. 3 and a nomenclature is given in Table 2.

The model accounts for three separate flow contributions:

1. Wastewater from households.
2. Fast rainfall–runoff from paved areas of the catchment.
3. Slow unintended infiltration water from permeable surfaces.

The modelling of each flow component is detailed in the following.

2.2.1. Wastewater fluctuations

The wastewater flow variation D_t is modelled by a harmonic function reflecting the diurnal pattern of water discharge from households

$$D_t = \sum_{i=1}^{2}\left(s_i \sin\frac{i2\pi t}{L} + c_i \cos\frac{i2\pi t}{L}\right), \quad (1)$$

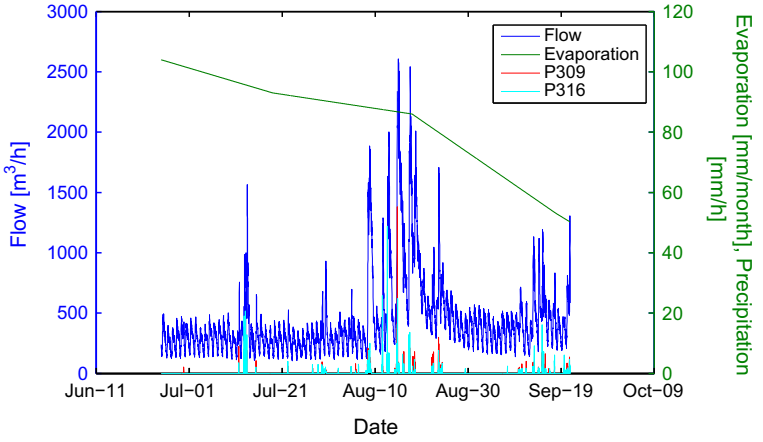

Fig. 2. Observed data used: flow, precipitation from two rain gauges and evaporation.

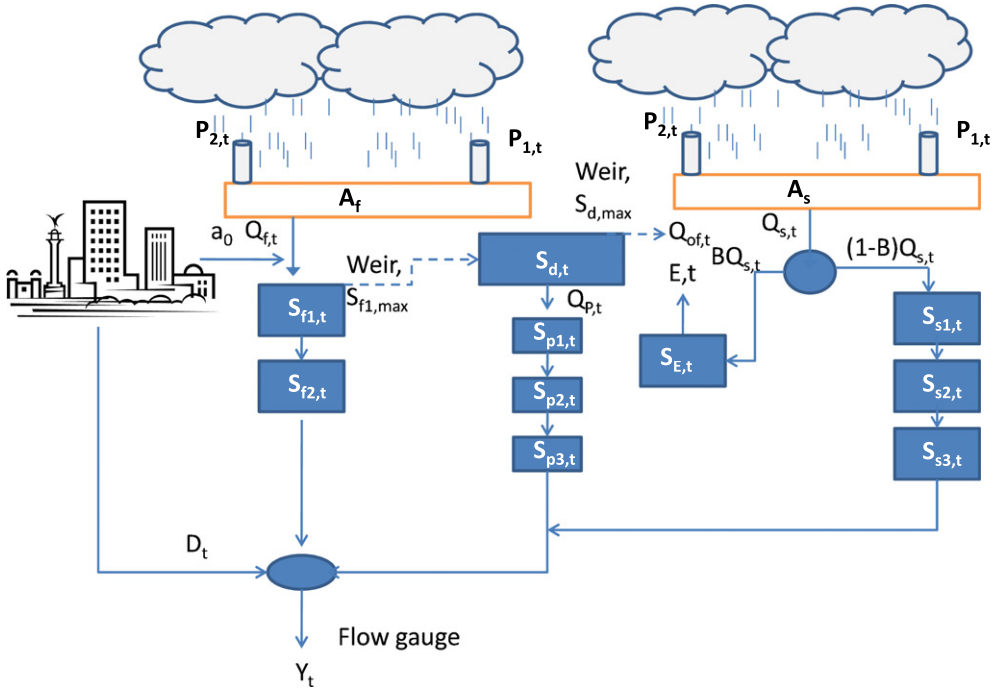

Fig. 3. The conceptual hydrological model.

where L is the period of 24 h, the parameters s_1, c_1, s_2 and c_2 are non-physical parameters and t is time. To fully account for the wastewater flow a constant average flow a_0 has to be added to D_t.

2.2.2. Fast rainfall–runoff from paved areas

The fast rainfall–runoff contribution Q_f from paved surfaces is modelled by

$$Q_{f,t} = \alpha A_f P_{1,t} + (1-\alpha) A_f P_{2,t}, \tag{2}$$

where P_1 and P_2 are the measured rain inputs (respectively P_{316} and P_{309} in Fig. 1), α is a rain gauge weighting factor and A_f represent the total paved area of the catchment.

Two linear reservoirs S_{f1} and S_{f2} and a retention time constant K_f are used to describe the storage and transportation of water on paved surfaces as well as inside the sewer system. As shown, in Fig. 3, a_0 is connected to S_{f1} to imitate reality in which wastewater always passes the sewer system. To account for a maximum internal water storage in detention basins during heavy rainfall a constraint C_1 on the capacity of S_{f1} is implemented and denoted $S_{f1,max}$. If the capacity of S_{f1} is exceeded, i.e. if $S_{f1} > S_{f1,max}$, water is diverted to a detention basin S_d. This restriction can be approximated by a continuous sigmoid function as follows

$$C_1(S_{f1,t}) = \frac{1}{1+\exp(-\zeta(S_{f1,max} - S_{f1,t}))}, \tag{3}$$

where ζ is used to tune the transition around the threshold $S_{f1,max}$; the larger the ζ the sharper the threshold. In order for the gradient-based parameter estimation method (see later Sections) to work properly this threshold should not be too sharp; we found that $\zeta = 0.05$ worked well during parameter estimation and hence this constant was applied for all other constraints in the system to be introduced in the following. The parameter estimation method is detailed in Section 3. In practice the aim is to secure that C_1 is equal to or close to one when $S_{f1} < S_{f1,max}$ and equal to or close to zero when the opposite holds. C_1 can then be used to control the inlet to S_{f1} such that inlet takes place only when $S_{f1} < S_{f1,max}$, and the mass balance for S_{f1} becomes

$$dS_{f1,t} = \left[C_{1,t}(Q_{f,t} + a_0) - \frac{2}{K_f} S_{f1,t}\right] dt. \tag{4}$$

This also means that when the opposite holds, i.e. when $S_{f1} > S_{f1,max}$, we can utilise $(1 - C_1)$ to ensure that water is diverted to the detention basin S_d instead of to S_{f1}.

In the sewer system the detention basin has a maximum capacity $S_{d,max}$ that can be estimated using another continuous constraint

$$C_2(S_{d,t}) = \frac{1}{1+\exp(\zeta(S_{d,max} - S_{d,t}))}. \tag{5}$$

When $S_d > S_{d,max}$ holds C_2 becomes approximately one and zero when the opposite holds and the overflow from the system becomes

$$Q_{of,t} = C_{2,t} \frac{(S_{d,t} - S_{d,max})}{dt}. \tag{6}$$

A pumping station is emptying the detention basin once the system regains vacant capacity which is consistent with the real system. In this simplified conceptual model this is assumed to happen when the volume of S_{f1} is below some proportion γ of $S_{f1,max}$, and a new constraint is required

$$C_3(S_{f1,t}) = \frac{1}{1+\exp(-\zeta(\gamma S_{f1,max} - S_{f1,t}))}. \tag{7}$$

The pumping is assumed turned off once the volume in the detention basin falls below a minimum volume $S_{d,min}$ that is fixed to 500 m^3. This minimum volume has no practical meaning but is simply inserted to avoid numerical instability which might occur had the minimum volume been fixed to zero. This constraint is modelled by

$$C_4(S_{d,t}) = \frac{1}{1+\exp(\zeta(S_{d,min} - S_{d,t}))}. \tag{8}$$

The pumping rate Q_p is assumed constant, however three linear reservoirs (S_{p1}, S_{p2} and S_{p3}) with a retention constant K_p are applied to simulate the slow startup and slowdown of the pumping rate. All

Table 2
Model nomenclature.

Symbol	Description	Unit
Inputs		
P_1	Rain	mm/h
P_2	Rain	mm/h
E	Evaporation	mm/h
Model states		
S_{f1}, S_{f2}	Fast reservoirs	m^3
S_d	Detention storage	m^3
S_{p1}, S_{p2}, S_{p3}	Pumping station reservoirs	m^3
S_E	Evaporation reservoir	m^3
S_{s1}, S_{s2}, S_{s3}	Infiltration reservoirs	m^3
Wastewater parameters		
s_1, s_2	Sine constants	–
c_1, c_2	Cosine constants	–
a_0	Average wastewater flow	m^3/h
Fast rainfall–runoff parameters		
A_f	Impermeable run-off area	ha
K_f	Retention time, fast runoff	h
α	Rain gauge weighting coef.	–
S_{f1max}	Max. vol. of fast reservoir	m^3
γ	Stop/start of pumps	–
S_{dmax}, S_{dmin}	Max. and min. vol. det. basin	m^3
Q_p	Pumping rate	m^3/h
K_p	Start/stop pumping time constant	h
Other		
ζ	Transition constant	–
Slow infiltration parameters		
A_s	Impermeable slow-runoff area	ha
K_s	Retention time, infilt. runoff	h
$S_{E,min}$	Min. vol. of evap. reservoir	m^3
ς	Saturation constant	m
Uncertainty parameters		
σ_{f1}, κ	Diffusion for S_{f1}	–
σ_{f2}	Diffusion for S_{f2}	–
σ_d	Diffusion for S_d	m^3
σ_{p1}	Diffusion for S_{p1}	m^3
σ_{p2}	Diffusion for S_{p2}	m^3
σ_{p3}	Diffusion for S_{p3}	–
σ_E	Diffusion for S_E	m^3
σ_{s1}	Diffusion for S_{s1}	m^3
σ_{s2}	Diffusion for S_{s2}	m^3
σ_{s3}	Diffusion for S_{s3}	–
S, S_z	Observation noise	m^3/h, –
y_{min}, y_{max}	Observation transformation	–
Time		
k	Discrete time steps	0.25 h
t	Continuous time	h
Outputs		
y_k	Observed flow k	m^3/h

the necessary state equations for modelling the fast runoff are shown in Eq. (12).

2.2.3. Slow infiltration inflow

The infiltration inflow Q_s is modelled in a similar way as the runoff from paved areas

$$Q_{s,t} = \alpha A_s P_{1,t} + (1-\alpha) A_s P_{2,t}, \tag{9}$$

where A_s represents the measured catchment area (13.2 km^2, see Table 1) that drains to the sewer system. The slow infiltration inflow to the sewer system is modelled using three linear infiltration reservoirs in series denoted S_{s1}, S_{s2} and S_{s3} with the average retention constant, K_s. The amount of rain water entering the first reservoir, S_{s1}, is assumed to depend on the saturation of the infiltration surface. The saturation storage is modelled by a reservoir, S_E, that is emptied by evaporation E. The amount of infiltration water being diverted to S_E is governed by an exponential relationship defined as

$$B_{E,t} = \exp\left(\frac{-S_{E,t}}{A_s \varsigma}\right) \tag{10}$$

which means that for a less saturated storage, less water flows to the infiltration runoff, and for a more saturated storage, more water flows to the infiltration reservoirs. The parameter ς is a saturation constant. To prevent S_E from becoming negative a minimum volume of S_E is defined $S_{E,min}$ similarly to $S_{d,min}$ which was fixed to 2000 m^3 and yet another constraint C_5 is therefore necessary

$$C_5(S_{E,t}) = \frac{1}{1+\exp(\zeta(S_{E,min} - S_{E,t}))}. \tag{11}$$

2.2.4. State-space formulation of the model

A state-space representation of the conceptual hydrological model can now be formulated in continuous time by

$$d\begin{bmatrix} S_{f1,t} \\ S_{f2,t} \\ S_{d,t} \\ S_{p1,t} \\ S_{p2,t} \\ S_{p3,t} \\ S_{E,t} \\ S_{s1,t} \\ S_{s2,t} \\ S_{s3,t} \end{bmatrix} = \begin{bmatrix} C_{1,t}(Q_{f,t}+a_0) - \frac{2}{K_f} S_{f1,t} \\ \frac{2}{K_f} S_{f1,t} - \frac{2}{K_f} S_{f2,t} \\ (1-C_{1,t})(Q_{f,t}+a_0) - C_{3,t}C_{4,t}Q_{p,t} - Q_{of,t} \\ C_{3,t}C_{4,t}Q_{p,t} - \frac{3}{K_p} S_{p1,t} \\ \frac{3}{K_p} S_{p1,t} - \frac{3}{K_p} S_{p2,t} \\ \frac{3}{K_p} S_{p2,t} - \frac{3}{K_p} S_{p3,t} \\ B_{E,t}Q_{s,t} - C_{5,t}A_s E_t \\ (1-B_{E,t})Q_{s,t} - \frac{3}{K_s} S_{s1,t} \\ \frac{3}{K_s} S_{s1,t} - \frac{3}{K_s} S_{s2,t} \\ \frac{3}{K_s} S_{s2,t} - \frac{3}{K_s} S_{s3,t} \end{bmatrix} dt, \tag{12}$$

and the discrete time observation equation becomes

$$\hat{y}_k = \frac{2}{K_f} S_{f2,k} + \frac{3}{K_p} S_{p3,k} + \frac{3}{K_s} S_{s3,k} + D_k. \tag{13}$$

The time index k denotes the time instances t_k where observations y_k are available.

3. Uncertainty estimation

In this section we add uncertainty terms to the deterministic model formulation of Section 2.2, write the model using a general notation and outline the parameter estimation method. These steps were also described in Breinholt et al. (2011) and are only summarised here to provide some methodology background. Three different approaches to quantify the inherent uncertainties of the system are incorporated into three models denoted M1, M2 and M3. All three models are based on the same deterministic hydrological model structure described in Section 2.2. An overview of how they differ from each other with respect to their incorporated error assumption and application is given in Table 3. All three models rely on the Maximum Likelihood (ML) estimation method for parameter estimation (Kristensen et al., 2004b) however in M1 and M2 all uncertainty is described by an output error term and no data assimilation takes place. Therefore M1 and M2 are suitable for simulation only. In M3 we apply data assimilation using Stochastic Differential Equations (SDE's) and the Extended Kalman Filter (EKF) to continuously update some of the model states during parameter estimation. This means that M3 is suitable for making predictions. However, M3 is validated for both prediction (using real-time updating/data assimilation) and simulation (without real-time updating) and compared to M1 and M2. The open source software CTSM[2] is used to estimate the parameters of the models, M1–M3. In Breinholt et al. (2011) CTSM was used for

[2] Continuous-Time Stochastic Modelling – www.imm.dtu.dk/ctsm.

Table 3
Overview of the three different models and their approach to uncertainty estimation.

Model	Uncertainty est. method	Parameter est. method	Obs. transf. $\mathcal{Z}$	Error minimisation	SDE	Application
M1	Output error	ML	No	$\hat{y}_k - y_k$	No	Sim
M2	Output error	ML	Yes	$\hat{z}_k - z_k$	No	Sim
M3	Prediction error	ML	Yes	$\hat{z}_{k\|k-1} - z_{k\|k-1}$	Yes	Sim./Pred.

ML: maximum likelihood. SDE: stochastic differential equations. Sim.: simulation. Pred.: prediction.

parameter estimation in prediction only. A detailed account of the mathematics behind the program is given in Kristensen and Madsen (2003).

3.1. General model formulation

The hydrological state-space model given in (12) and (13) can be written in a more compact form using a general notation that also accounts for uncertainty due to model structural errors/limitations and input by adding a state noise term (diffusion term) in Eq.(12) and by adding an observation noise term e_k in Eq. (13)

$$d\boldsymbol{X}_t = \underbrace{\boldsymbol{f}(\boldsymbol{X}_t, \boldsymbol{U}_t, t, \boldsymbol{\theta})}_{\text{drift term}}\, dt + \underbrace{\boldsymbol{\sigma}(\boldsymbol{X}_t, \boldsymbol{U}_t, t, \boldsymbol{\theta})}_{\text{diffusion term}}\, d\boldsymbol{\omega}_t, \tag{14}$$

$$y_k = \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{\theta}) + \underbrace{e_k}_{\text{obs. noise term}}. \tag{15}$$

Here $\boldsymbol{X}_t \in \mathbb{R}^{10}$ represents the state variables of the model $(S_{f1}, S_{f2}, S_d, S_{p1}, S_{p2}, S_{p3}, S_E, S_{s1}, S_{s2}, S_{s3})$, $\boldsymbol{U}_t \in \mathbb{R}^3$ the input variables (P_1, P_2, E) and $\boldsymbol{\theta} \in \mathbb{R}^p$ the parameters. The number of parameters p is the sum of the number of parameters in the system equation and the observation equation, which will depend on how the uncertainty is described. $\boldsymbol{f}(\cdot)$ represents the functions in the drift term, i.e. the functions in Eq. (12). Note that the state uncertainty is modelled by the diffusion term $\boldsymbol{\sigma}(\cdot)$ which can be a function of the states, the inputs, the time and the parameters, and finally $\boldsymbol{\omega}_t$ is a 10-dimensional standard Wiener process. In the observation Eq. (15), $\boldsymbol{h}(\cdot)$ is a short notation of the relationship between the modelled states and the observations.

3.2. Parameter estimation

In the case of M1 and M2 the observation noise term e_k accounts for all the uncertainty and is assumed to be a Gaussian white noise process with variance S and the likelihood function can be written

$$L(\boldsymbol{\theta}; \boldsymbol{\mathcal{Y}}_N) = \frac{1}{(\sqrt{2\pi S})^N} \exp\left[-\sum_{k=1}^{N} \frac{1}{2S}(y_k - \hat{y}_k)^2\right], \tag{16}$$

where y_k and $\hat{y}_k$ are respectively the measured and modelled outputs at time instances k, N is the number of measurements and $\boldsymbol{\mathcal{Y}}_N$ a series of observations.

In the case of M3 data assimilation takes place in the states by the diffusion noise term and the EKF. By introducing the notation $\boldsymbol{\mathcal{Y}}_k = [y_k, y_{k-1}, \ldots, y_1, y_0]$, the likelihood function is expressed as a product of conditional densities

$$L(\boldsymbol{\theta}; \boldsymbol{\mathcal{Y}}_N) = P(\boldsymbol{\mathcal{Y}}_N|\boldsymbol{\theta}) = \left(\prod_{k=1}^{N} P(y_k|\boldsymbol{\mathcal{Y}}_{k-1}, \boldsymbol{\theta})\right) P(y_0|\boldsymbol{\theta}), \tag{17}$$

where Bayes theorem $P(A \cap B) = P(A|B)P(B)$ is repeatedly used at each time step to formulate the likelihood function as a product of the one step ahead conditional densities and $P(y_0|\boldsymbol{\theta})$ is a parameterisation of the initial conditions. It is assumed that the system equations are driven by a Wiener process which have Gaussian increments and thus the conditional probabilities in (17) can be approximated by Gaussian densities.

The Gaussian density is completely characterised by the mean and covariance of the one step prediction, which are denoted by $\hat{y}_{k|k-1} = E\{y_k|\boldsymbol{\mathcal{Y}}_{k-1}, \boldsymbol{\theta}\}$ and $\boldsymbol{R}_{k|k-1} = V\{y_k|\boldsymbol{\mathcal{Y}}_{k-1}, \boldsymbol{\theta}\}$, respectively, and, by introducing an expression for the innovation formula, $\boldsymbol{\varepsilon}_k = y_k - \hat{y}_{k|k-1}$ the likelihood function can be rewritten as (Madsen, 2008)

$$L(\boldsymbol{\theta}; \boldsymbol{\mathcal{Y}}_N) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}_k^{\top} \boldsymbol{R}_{k|k-1}^{-1} \boldsymbol{\varepsilon}_k\right)}{\sqrt{\det(\boldsymbol{R}_{k|k-1})}\left(\sqrt{2\pi}\right)^l}\right) P(y_0|\boldsymbol{\theta}), \tag{18}$$

where the conditional mean and covariance are calculated using an EKF. The EKF is needed (rather than the Kalman filter) due to the non-linearity of the model. The EKF is preferred to a particle filter and the ensemble Kalman filter because of the relatively limited number of model states and to avoid large computational burdens.

The parameter estimates can be obtained by conditioning on the initial values and solving the optimisation problem

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \{\log(L(\boldsymbol{\theta}; \boldsymbol{\mathcal{Y}}_N|y_0))\}. \tag{19}$$

Numerical methods are needed to optimise the likelihood function (Kristensen and Madsen, 2003).

The maximum likelihood method also provides an assessment of the uncertainty for the parameter estimates in Eq. (19) since the maximum likelihood estimator is asymptotically normal distributed with mean $\boldsymbol{\theta}$ and covariance matrix

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = \boldsymbol{H}^{-1}.$$

The matrix $\boldsymbol{H}$ is the Fisher Information Matrix (Madsen and Thyregod, 2011) given by

$$h_{ij} = -E\left\{\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log(L(\boldsymbol{\theta}|\boldsymbol{\mathcal{Y}}_{k-1}))\right\} \quad i,j = 1, \ldots, p. \tag{20}$$

In practice an approximation for $\boldsymbol{H}$ is obtained by the observed Hessian h_{ij} evaluated for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Due to the asymptotic Gaussianity of the estimator in Eq. (19), a t-test can be performed to check if the estimated parameters are statistically significant (Madsen and Thyregod, 2011).

3.3. Output error method (M1) assuming variance homogeneity

The first method we explore to estimate the simulation uncertainty is the so-called output error method in which the diffusion term is ignored and a weighted sum of squared errors between observations and simulated outputs are minimised for the whole observation series. The implicit assumption is that errors can be attributed to flow meter inaccuracy alone and hence state and input uncertainty is neglected. This means that the system equation is identical to (12) but the observation Eq. (13) is extended to include an observation noise term e_k and we get

$$y_k = \frac{2}{K_f} S_{f2,k} + \frac{3}{K_p} S_{p3,k} + \frac{3}{K_s} S_{s3,k} + D_k + e_k, \tag{21}$$

where $e_k \in N(0, S)$ is assumed to be a Gaussian white noise process with constant variance S.

3.4. Output error method with observation transformation (M2)

The assumption of variance homogeneity might be unreasonable as uncertainty is expected to be higher in wet weather than in dry weather periods because of biases introduced when non-representative rain inputs are used for running the models and because flow meters typically become less accurate with increasing flow (Salamon and Feyen, 2010; Willems, 2012). Usually some kind of observation transformation will be required to stabilize the variance. A common way of dealing with heteroscedasticity in hydrological models is to resort to Box–Cox transformations, see e.g. Yang et al. (2007), Feyen et al. (2007), Jin et al. (2010), Franchini et al. (2011) and Beven (2009a) also provide examples of possible transformations. The necessary transformation can be deduced from a range-mean plot (Madsen, 2008) that examines how the size of the errors changes with the flow rate and a range mean plot can be obtained by grouping the observations into subsets of increasing flow range and plot them against the Root Mean Square Errors (RMSEs) of the residuals belonging to each subset, thereby identifying the transformation needed.

A transformation of the observations $\mathcal{Y}$ to $\mathcal{Z} = \Psi(\mathcal{Y})$ and $\mathcal{Z}_N = [z_1, \ldots, z_N]^T$ will change the observation equation (Eq. (13)) to

$$z_k = \Psi(\hat{y}_k) + \epsilon_k, \tag{22}$$

where $\epsilon_k \in N(0, S_z)$ is now assumed to be a Gaussian white noise process. The parameters are estimated using the same likelihood function as in (16), however with transformed model outputs and observations. Possible transformations Ψ are discussed in Section 5.

3.5. The SDE method (M3)

Using an SDE approach as formulated in (14) and (15)and the EKF for parameter estimation the model is extended to include uncertainty in both the system and observation equation. This is attractive because uncertainty due to spatio-temporal rainfall heterogenity and model inadequacies are addressed directly by the state noise term and separated from uncertainty that origins from flow gauge errors that are accounted for by the observation noise term. Hence in addition to the system Eq. (12) a diffusion term is added

$$\boldsymbol{\sigma}(\boldsymbol{X}_t, \boldsymbol{U}_t, t, \boldsymbol{\theta}) = \mathrm{diag}\left[\sigma_{f1}, \sigma_{f2}, \sigma_d, \sigma_{p1}, \sigma_{p2}, \sigma_{p3}, \sigma_E, \sigma_{s1}, \sigma_{s2}, \sigma_{s3}\right]^T \tag{23}$$

In Breinholt et al. (2011) it was concluded that a state dependent diffusion term could provide an adequate description of the one step ahead uncertainty. However, state dependent diffusion terms are not possible in CTSM, because they require higher order filtering techniques that have been shown to become numerically unstable (Vestergaard, 1998). Instead a Lamperti transformation of the states is required, see Breinholt et al. (2011).

The transformation of the states changes the system Eq. (14) and the observation Eq. (15) to

$$d\boldsymbol{Z}_t = \tilde{\boldsymbol{f}}(\boldsymbol{Z}_t, \boldsymbol{U}_t, t, \boldsymbol{\theta})dt + \tilde{\boldsymbol{\sigma}}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t, \tag{24}$$

$$y_k = \tilde{\boldsymbol{h}}(\boldsymbol{Z}_k, \boldsymbol{U}_k, t_k, \boldsymbol{\theta}) + e_k, \tag{25}$$

where the functions $\boldsymbol{f}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ and $\boldsymbol{h}(\cdot)$ have been reformulated, respectively to $\tilde{\boldsymbol{f}}(\cdot)$, $\tilde{\boldsymbol{\sigma}}(\cdot)$ and $\tilde{\boldsymbol{h}}(\cdot)$ in relation to the transformation of the state-space. The parameters of the drift term and the input–output relations are unaffected by the transformation.

Regarding the observation noise term, it is again anticipated that some transformation of the observations is required. However, since the diffusion term now accounts for most of the noise, at least in rainy periods, the observation transformation needed will presumably be different from the observation transformation needed in Model M2.

4. Model evaluation

4.1. Generating predictions and confidence bounds

The objective with the proposed models is to generate reliable mean values and confidence bounds for both the one step prediction $\left(\hat{y}_{k+1|k}, \left[y^{(l)}_{k+1|k}; y^{(u)}_{k+1|k}\right]\right)$ and for simulation $\left(\hat{y}_{k|0}, \left[y^{(l)}_{k|0}; y^{(u)}_{k|0}\right]\right)$, where the upper cases (l) and (u) refer to respectively the lower and upper confidence bounds. It is only meaningful to distinguish the one step prediction uncertainty from the simulation uncertainty in the case of model M3.

4.1.1. Model M1

For the case of M1 the mean and confidence bounds are directly obtained from (21) recalling that the confidence bounds are estimated using a normal distribution with zero mean and variance S.

4.1.2. Model M2

For the case of M2 the mean and the confidence bounds in the $\mathcal{Z}$-domain are given by (22) recalling that the confidence bounds are estimated from a normal distribution with zero mean and variance S_z, however to obtain the mean and the confidence bounds of the $\mathcal{Y}$-domain additional calculations are needed. To obtain $\hat{y}_{k|0}$ the inverse transformation $\Psi^{-1}(\hat{z}_k)$ will have to be taken and then the confidence bounds can be derived from $\left[y^{(l)}_{k|0}; y^{(u)}_{k|0}\right] = \left[\Psi^{-1}\left(z^{(l)}_{k|0}\right); \Psi^{-1}\left(z^{(u)}_{k|0}\right)\right]$.

4.1.3. Model M3

For the case of M3 in prediction mode the mean and variance in the $\mathcal{Z}$-domain are given by $\hat{z}_{k+1|k} = \tilde{h}(\hat{\boldsymbol{Z}}_{k+1|k}, \boldsymbol{u}_{k+1}, t_{k+1}, \boldsymbol{\theta})$ and the output variance prediction is given by (Jazwinski, 2007)

$$\boldsymbol{R}_{k|k-1} = \boldsymbol{C}\boldsymbol{P}_{k|k-1}\boldsymbol{C}^T + \boldsymbol{S}, \tag{26}$$

where in our case $\boldsymbol{h}(\cdot) = \boldsymbol{C}\boldsymbol{Z}_k, \boldsymbol{P}_{k|k-1} = V[\boldsymbol{Z}_{t_{k+1}}|\boldsymbol{Z}_{t_k}]$ and $\boldsymbol{S}$ is the observation variance. How to derive confidence bounds in simulation was described in detail in Thordarson et al. (2012) and will not be repeated here. Again the inverse transformation will have to be taken as described above in Section 4.1.2 once the confidence bounds of the $\mathcal{Z}$-domain has been found.

4.2. Probabilistic performance measures

To measure each models ability to generate useful confidence bounds we adopt the performance measures described in Thordarson et al. (2012). We are generally interested in the width of the derived confidence bounds referred to as the sharpness and the correspondence between nominal and observed coverage of the confidence bounds referred to as the reliability. The evaluation of the sharpness and the reliability can be combined into a single scoring criterion denoted the interval skill score. In the following the performance measures are formulated for simulation uncertainty evaluation, however it should be kept in mind that the notation is slightly different when formulated for the one step prediction evaluation (see Thordarson et al. (2012)).

4.2.1. Sharpness

The sharpness is defined as the average width of the confidence bounds at any given β

$$\overline{\delta}^{(\beta)} = \frac{1}{K}\sum_{k=1}^{K}\left(y_{k|0}^{(u)} - y_{k|0}^{(l)}\right), \tag{27}$$

where $y_{k|0}^{(u)}$ and $y_{k|0}^{(l)}$ represent, respectively, the upper and lower confidence bound at any given time step and quantile level β of the entire simulation period K. When comparing interval simulations generated from different models, the one with the most narrow confidence bounds is the sharpest.

4.2.2. Reliability

For a given confidence bound and measured output an indicator variable is introduced

$$n_k^{(\beta)} = \begin{cases} 1, & \text{if } y_k \in \left[y_{k|0}^{(l)}; y_{k|0}^{(u)}\right] \text{ for } k \leqslant K, \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

The reliability is the percentage time that the observations are within the confidence bounds for the whole time series.

The reliability bias is defined as

$$b^{(\beta)} = 1 - \beta - \bar{n}^{(\beta)}, \tag{29}$$

where perfect reliability is given by $b^{(\beta)} = 0$. When discrepancy between the empirical and nominal coverage is observed, the coverage of the confidence bound is biased and $\bar{n}^{(\beta)} > 1-\beta$ corresponds to a positive bias in the coverage and $\bar{n}^{(\beta)} < 1-\beta$ a negative bias.

4.2.3. Interval skill score

The sharpness and reliability can be combined into a performance measure denoted the interval skill score Sc (Gneiting and Raftery, 2007) that rewards narrow and reliable confidence bounds. For any given quantile level for evaluation and K measurements the interval skill score is defined as

$$Sc^{(\beta)} = \frac{1}{K}\sum_{k=1}^{K}\left(y_{i,k|0}^{(u)} - y_{i,0|k}^{(l)}\right) + \frac{2}{\beta}\left[\left(y_{i,k|0}^{(l)} - y_k\right)\mathbf{1}\left\{y_k < y_{i,k|0}^{(l)}\right\} + \left(y_k - y_{i,k|0}^{(u)}\right)\mathbf{1}\left\{y_k > y_{i,k|0}^{(u)}\right\}\right]. \tag{30}$$

Hence the interval skill score is the average width of the confidence bounds plus the sum of deviances outside those bounds weighted by the inverse of the quantile level. In case the observation misses the quantile interval a penalty is incurred which depends on the considered quantile level, meaning that an observation miss at the 95% confidence interval is much more expensive than a miss at e.g. the 5% confidence interval. Thus, the smaller the interval skill score the better the probabilistic prediction performance of the model, and so $Sc^{(\beta)}$ can be used to evaluate the different models probabilistic performance and eventually to select a preferred model.

5. Formulation, estimation and evaluation of models

5.1. Estimation and evaluation of M1

5.1.1. Parameter estimates for M1

Table 4 presents the parameter estimates for M1 and their associated standard deviances. All estimated parameters were significant. It is noted that the parameters $S_{d,min}$, A_s and $S_{E,min}$ were fixed as explained in Section 2.2. The contributing fast area was found to 0.6 km^2 (exp(4.21) hectares and units converted) which was much larger than found by Tomicic (2006) (0.33 km^2), see Table 1.

The fast runoff constant K_f was estimated to 5.15 h whereas the slow infiltration constant K_s was estimated to almost 80 h. The maximum capacity of S_{f1} was found to 10,788 m^3 and the size of the detention basin estimated to approx. 24,500 m^3 (subtracted $S_{d,min} = 500$ m^3). The emptying by the pumps of the detention basin starts when the volume of S_{f1} has fallen to 67% of maximum capacity. The pumping rate was estimated to 818 m^3/h and the time constant K_p to 4.2 h. The standard deviation of the observation noise was estimated to 88.7 m^3/h, corresponding to $\ln(S) = 8.97$ as shown in Table 4.

Table 4
Estimation results.

Parameter	M1		M2		M3	
	$\hat{\theta}$	sd($\hat{\theta}$)	$\hat{\theta}$	sd($\hat{\theta}$)	$\hat{\theta}$	sd($\hat{\theta}$)
Wastewater						
s_1	−38.7	0.8	−44.8	1.0	−49.1	3.25
c_1	−94.0	1.5	−89.9	1.1	−67.1	3.43
s_2	−49.9	1.0	−43.4	1.33	−38.4	2.1
c_2	12.9	1.3	22.6	1.3	18.1	19.9
a_0	275.2	1.13	277.2	0.8	278.3	5.7
Fast rainfall–runoff						
$\ln(A_f)$	4.21	$6e^{-3}$	4.19	$5.9e^{-3}$	4.21	0.04
K_f	5.15	0.05	4.93	0.05	4.55	0.04
α	0.39	$6e^{-3}$	0.36	$9.1e^{-3}$	0.25	0.03
$S_{f1,max}$	10,788	134	8827	191	8827[a]	–
γ	0.67	0.05	0.61	0.03	0.57	0.03
$S_{d,max}$	25,059	1063	21,989	1912	19,500	1431
$S_{d,min}$	500[a]	–	500[a]	–	500[a]	–
Q_p	818	8.8	890	2.7	850	64.5
K_p	4.2	0.3	2.5	0.2	3.5	0.7
Slow infiltration						
$\ln(A_s)$	7.19[a]	–	7.19[a]	–	7.19[a]	–
K_s	79.7	8.0	85.1	1.2	133.1	13.5
$S_{E,min}$	2000[a]	–	2000[a]	–	2000[a]	–
ς	1546	14.8	1672	19.2	1176	169
Uncertainty						
κ	–	–	–	–	1.2[a]	–
$\ln(\sigma_{f1})$	–	–	–	–	−10.76	0.07
$\ln(\sigma_{f2})$	–	–	–	–	−2.26	0.03
$\ln(\sigma_d)$	–	–	–	–	10^{-10}[a]	–
$\ln(\sigma_{p1})$	–	–	–	–	10^{-10}[a]	–
$\ln(\sigma_{p2})$	–	–	–	–	10^{-10}[a]	–
$\ln(\sigma_{p3})$	–	–	–	–	−2.73	0.12
$\ln(\sigma_E)$	–	–	–	–	10^{-10}[a]	–
$\ln(\sigma_{s1})$	–	–	–	–	10^{-10}[a]	–
$\ln(\sigma_{s2})$	–	–	–	–	10^{-10}[a]	–
$\ln(\sigma_{s3})$	–	–	–	–	−2.73	0.12
$\ln(S), \ln(S_z)$	8.97	0.017	−2.95	0.015	−6.39	0.04
y_{max}	–	–	3000[a]	–	–	–
y_{min}	–	–	30[a]	–	–	–
Other						
ζ	0.05[a]	–	0.05[a]	–	0.05[a]	–

[a] Fixed parameter value.

5.1.2. Investigating the model residuals for M1

Fig. 4a shows a plot of the standardised residuals r_s (residuals divided by their standard deviance) as a function of $\hat{y}$ and clearly r_s do not look randomly distributed over the flow interval. This is also evident from Fig. 4c, in which a range-mean plot of how the Root Mean Square Error (RMSE) of r_s varies with $\hat{y}$ is depicted, an expression of how the uncertainty varies with the observed flow rate. The observations were divided into 6 sub-sets of increasing flow magnitude and the RMSE calculated. Clearly, the uncertainty is small in dry weather and then increases as the flow rate increases up to a flow rate around 1500 m^3/h and then starts to decrease as the flow rate approaches the maximum flow capacity of the system. This seems reasonable since the system can transport only a limited flow due to the constraints of the system. Thus,

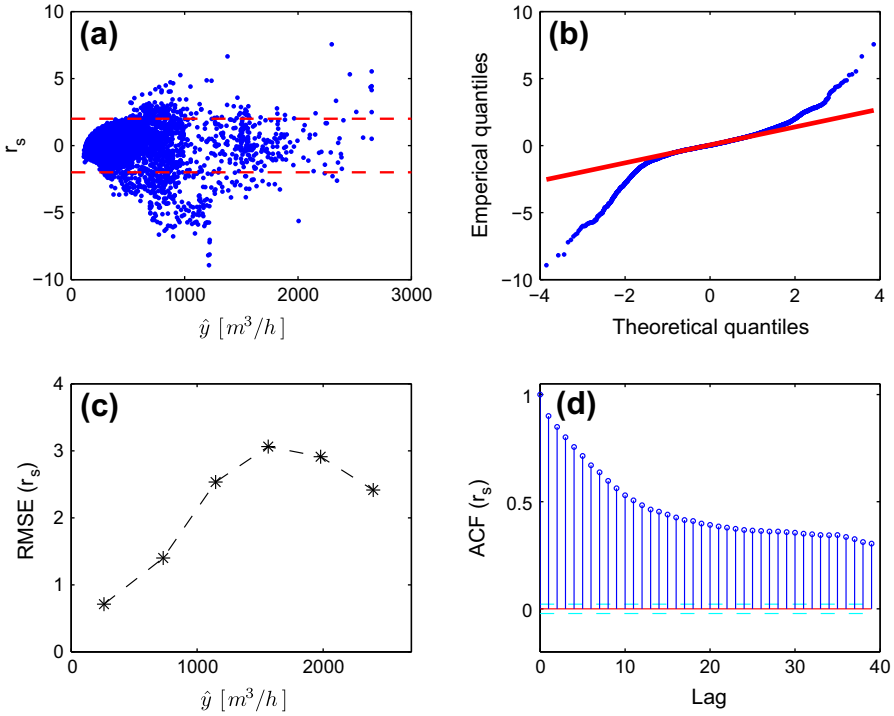

Fig. 4. Investigation of standardised residuals of Model M1.

the assumption of homoscedasticity which is one of the statistical preconditions for using model M1 is violated. Fig. 4b shows a Quantile–Quantile (Q–Q plot) comparing the model residuals on the vertical axis to a standard normal population on the horizontal axis. The deviation from linearity indicates that the residuals are not normally distributed and hence that the underlying statistical assumptions of the output error for model M1 are violated. The autocorrelation function (ACF) in Fig. 4d shows that considerable autocorrelation is present, which is also a clear violation of the statistical preconditions that demands residuals to be serially independent. Due to the rejection of the error assumptions of this model we should not expect the parameter estimates in Table 4 to be correct even though the *t*-test showed parameter significance.

Fig. 5 shows a plot of the generated 90% confidence bounds and the observations in both wet weather (left plot) and dry weather (right plot) periods. The lower confidence bound actually turns negative in dry weather periods which is due to the assumption that residuals are symmetric about the observed value. This problem is exacerbated if a larger quantile is considered, e.g. the 95% confidence bounds or higher. In wet weather periods the confidence bounds look better but perhaps too narrow. It should be recalled that the figure depicts a subperiod of the whole period only and therefore cannot be expected to enclose exactly 90% of the observations in the subperiod shown. We will return to this in Section 6 where we investigate the reliability of the confidence bounds.

The problem with heteroscedasticity suggests that a transformation of the observations are needed to obtain a homoscedastic Gaussian structure of the residuals.

5.2. Formulation, estimation and evaluation of model M2

5.2.1. Seeking a reasonable observation transformation

The pattern in Fig. 4c suggests that a logistic transformation could be feasible, i.e. a transformation of the observations to the $\mathcal{Z}$-domain by

$$\mathcal{Z} = \log\left(\frac{\mathcal{Y} - y_{min}}{y_{max} - \mathcal{Y}}\right). \tag{31}$$

This of course requires minimum y_{min} and maximum y_{max} values of $\mathcal{Y}$ to be specified, and good guestimates can be inferred from considering Fig. 4c. Such an observation transformation will change the observation equation to

$$z_k = \log\left(\frac{\hat{y}_k - y_{min}}{y_{max} - \hat{y}_k}\right) + \epsilon_k, \tag{32}$$

cf. Eq. (22). To illustrate how a logistic transformation of the observations affects the observation noise, consider Fig. 6.

Fig. 6a creates using a logistic $\mathcal{Z}$-transformation according to (31) of 1000 randomly generated flow values ($\hat{y}_k$) in the range 30–3000 m^3/h and adding a constant standard deviation of $\sqrt{S_z} = 0.3$ that is homoscedasticly distributed, see Fig. 6b. When returning to the original $\mathcal{Y}$-domain by taking the inverse

$$y_k = \frac{e^{z_k} y_{max} + y_{min}}{1 + e^{z_k}}, \tag{33}$$

we conclude from Fig. 6c, that the flow observations are now unequally scattered around $\hat{y}$ (show heteroscedastic behaviour). The scattering increases over the approximated range 30–1500 m^3/h and then decreases again above 1500 m^3/h. This trend is easier to recognise from the standard deviance $\sqrt{S}$ plotted in Fig. 6d that is arc-shaped. It is noticed that the shape of Figs. 6d and 4c look somewhat similar and hence a logistic transformation is proposed and $\sqrt{S_z}$ can be estimated with CTSM.

5.2.2. Parameter estimates for M2

The parameter estimates and their associated standard deviations for models M1 and M2 do not differ much from each other although minor deviations are detected for most of the wet weather parameters, see Table 4. The most noticeable estimation difference concerns $S_{f1,max}$ that is estimated almost 2000 m^3 smaller for M2. The estimated output error variance S_z has also changed considerably which is due to the logistic transformation of the

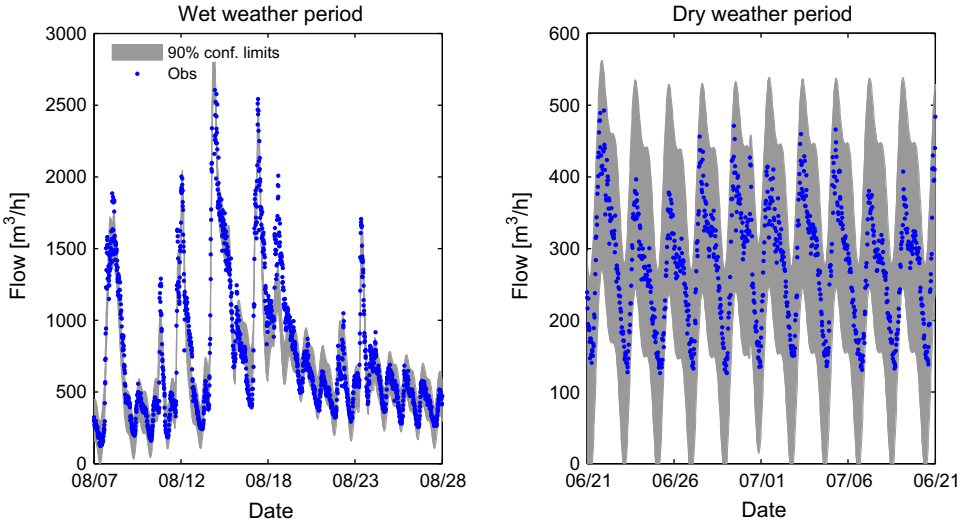

Fig. 5. 90% Confidence bounds of M1 vs. observations in a selected dry and wet weather calibration period.

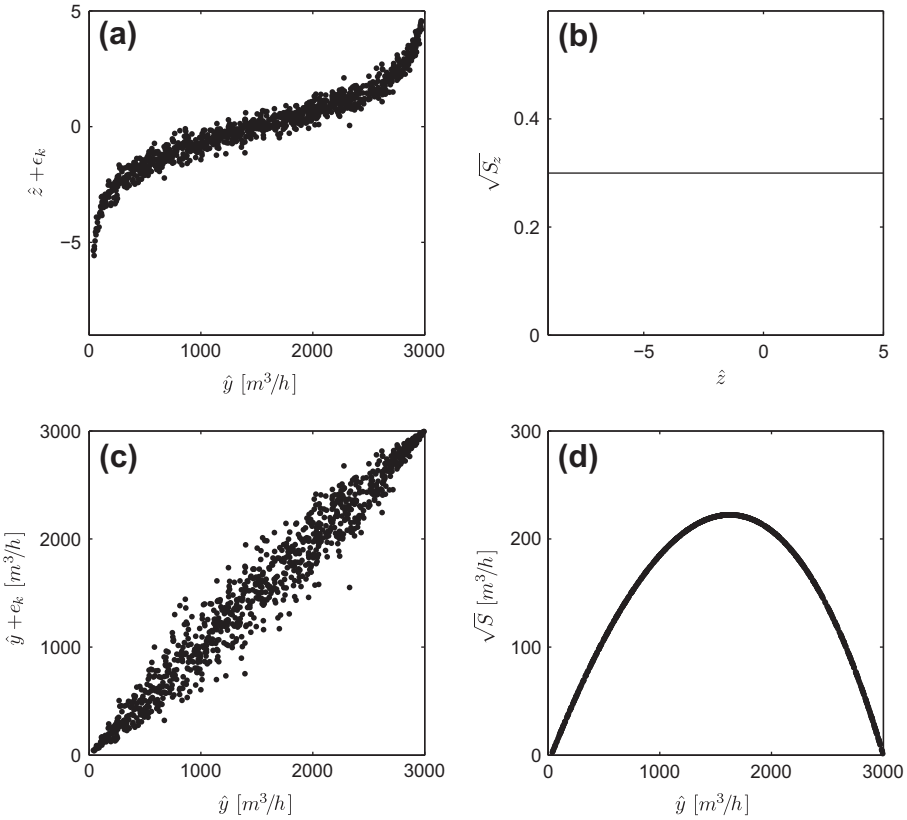

Fig. 6. Illustration of how an additive observation noise (with $\sqrt{S_z} = 0.3$) in the $\mathcal{Z}$-domain leads to a logistic observation noise term in the $\mathcal{Y}$-domain. Illustrated with 1000 randomly generated flow values in the interval 30–3000 m^3/h.

observations to the $\mathcal{Z}$-scale (Table 4). All estimated parameters were statistically significant.

5.2.3. Investigating the model residuals for M2

Fig. 7a shows that the r_s in the $\mathcal{Z}$-domain looks much more equally scattered than was the case for M1 and the distribution of the residuals is also much closer to a Gaussian distribution, which is visible from Fig. 7b. From Fig. 7c that now shows the RMSE of r_s as a function of $\hat{z}$, it is seen that there is still some heteroscedasticity present as the uncertainty is seen to vary with the transformed flow. However the variations are not as widespread as was the case for M1. Significant autocorrelation of the residuals is also present with model M2 (see Fig. 7d), and consequently the parameter estimates of Table 4 cannot be assumed 100% valid.

A plot of the 90% confidence bounds and the observations are shown in Fig. 8 for a selected wet and dry weather period. The dry weather confidence bounds looks much better than for M1, cf. Fig. 4. The confidence bounds do not enclose negative values and seem to cover most of the observed data.

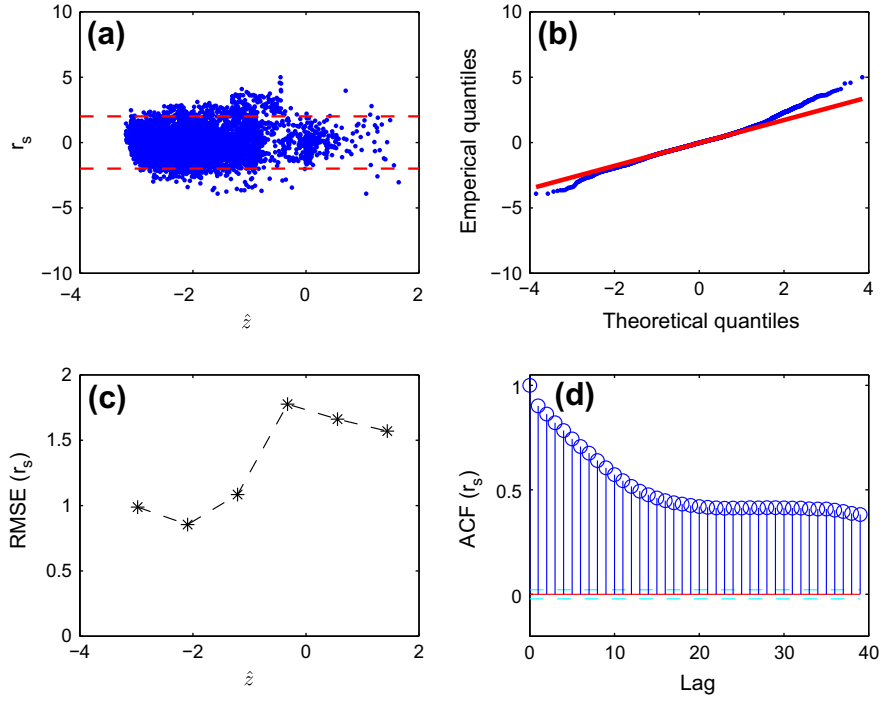

Fig. 7. Investigation of the standardised residuals of Model M2.

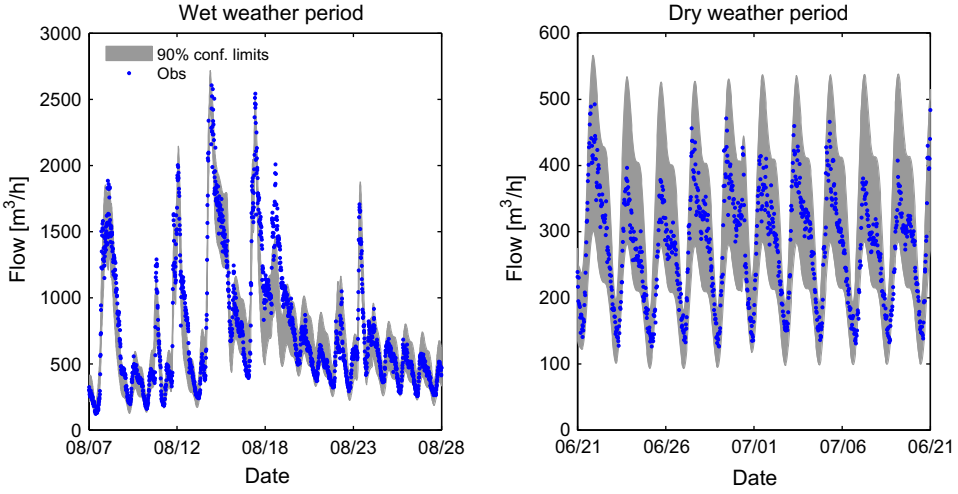

Fig. 8. 90% Confidence bounds of M2 vs. observations in a selected dry and wet weather period.

5.3. Formulation, estimation and evaluation of model M3

5.3.1. Seeking a reasonable diffusion term and observation transformation

In M3 the state-space SDE approach is used to account for both model structural limitations and input uncertainty by the diffusion term and flow meter inaccuracy by the observation noise term. The input uncertainty is particularly expected to affect S_{f1} since the uncertain rainfall–runoff contribution from paved areas enters this reservoir quickly during rain. Instead of accounting for decreasing uncertainty near maximum system capacity by transforming the observations as in M2, this is accounted for in the diffusion term directly. In Breinholt et al. (2011) a state proportional uncertainty described the one step prediction uncertainty adequately. However, a new flow meter capable of measuring the flow during extreme events, the data of which underlie this study, made it possible to identify the system's maximum conveyance capacity and a different relationship between state and uncertainty is therefore sought.

Fig. 9 displays two state dependent diffusion terms for S_{f1}. An arc-shaped state varying proposal shown with red[3] lines given by

$$\sigma_{1,t} = \sigma_{f1} S_{f1,t}(S_{f1,max}\kappa - S_{f1,t}), \tag{34}$$

resembles the logistic observation transformation shown in Fig. 6d, however instead of describing all the uncertainty by the observation noise term as in M2, most of the uncertainty is now accounted for by the state diffusion term. This is more reasonable since the natural constraints in the system should be accounted for in the states and not by the observation noise term that ideally accounts

[3] For interpretation of colour in Fig. 9, the reader is referred to the web version of this article.

for noise from inaccurate flow measurements only. σ_{f1} is a parameter to be estimated in CTSM whereas $\kappa > 1$ is a constant that ensures that state uncertainty is present also at maximum system capacity $S_{f1,max}$, in this example $S_{f1,max} = 8500\,\mathrm{m}^3$. It is noted that a small change in σ_{f1} (shown in Fig. 9a from $3e^{-5}$ to $4e^{-5}$) results in a significant displacement of the state dependent uncertainty. Likewise a small change in κ has a substantial impact on the state dependent uncertainty (see Fig. 9b). For comparison, the figure also demonstrates how a state-proportional diffusion term given by

$$\sigma_{1,t} = \sigma_{f1} S_{f1,t}, \tag{35}$$

would affect the state uncertainty in S_{f1} as S_{f1} increases (blue lines). In this example $\sigma_{f1} = 0.3$.

To implement the proposed state dependencies in CTSM it is necessary to apply the Lamperti transformation as mentioned in Section 3.5.

For the rest of the states we decided to use a state-proportional diffusion term and refer to Breinholt et al. (2011) for the derivation of the transformed system equation.

In order to avoid over-parameterisation in the diffusion term, we decided in addition to σ_{f1} to estimate only the state variances that appear in the observation equation, i.e. $\{\sigma_{f2}, \sigma_{p3}, \sigma_{s3}\}$ using a state-proportional diffusion term and refer to (Breinholt et al., 2011) for the derivation of the transformed system equation as well as to the advantages of using this transformation. The rest of the states were fixed at a small value, i.e. 10^{-10} (considered deterministic) and the diffusion term is then given by

$$\sigma_1(\boldsymbol{X}_t, \boldsymbol{U}_t, t, \boldsymbol{\theta}) = \mathrm{diag}\Big[\sigma_{f1}(S_{f1,max}\kappa - S_{f1,t}), \sigma_{f2} S_{f2,t}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, \sigma_{p3} S_{p3,t}, 10^{-10}, 10^{-10}, 10^{-10}, \sigma_{s3} S_{s3,t}\Big]^T. \tag{36}$$

In (Breinholt et al., 2011) a flow-proportional observation noise term is argued and is also implemented here.

5.3.2. Parameter estimates for M3

To be able to estimate the parameters we found it necessary to fix both $S_{f1,max}$ and κ. We used the value obtained with model M2 for $S_{f1,max}$ and reasoned that $\kappa = 1.2$ was a fair guestimate considering that we did not want the state uncertainty to vanish completely at full flow capacity. Even though the parameter estimates are found from the one step conditional densities we observe from Table 4 that most of the parameter estimates agreed fairly well with the estimates of M1 and M2. One noticeable estimation difference is seen for K_s that is larger in M3 meaning that the infiltration is taken to be slower in M3. It is also observed that the incorporation of the diffusion term enlarges the uncertainty of the parameter estimates considerably, compare the columns of $\mathrm{sd}(\hat{\theta})$ in Table 4. Especially the standard deviance of Q_p and ς are higher for M3, but the parameter estimates where nevertheless all significant. It is furthermore noticed that the selected diffusion parameters all could be estimated and that the variance of the observation noise deviates from the estimated one of M1 and M2, however this is not surprising because the estimate was conducted in a new $\mathcal{Z}$-domain.

5.3.3. Investigating the model residuals for M3

Model M3 is evaluated with respect to both the one step prediction and the simulation ability. Initially we address the models one step prediction ability since the model is tailored for this. It should be recalled that in prediction mode the states are updated in accordance with the observations through the EKF, whereas no updating takes place in simulation mode.

In Fig. 10 the residuals of the one step prediction mode are investigated. It is observed from Fig. 10a that the standardised residuals are nicely dispersed around zero and that most of the residuals appear inside the 95% confidence bounds. However, the Q–Q plot in Fig. 10b suggests that the residuals are non-Gaussian distributed as departures are clearly seen at the tails. The range-mean plot of Fig. 10c indicates that we are very close to the assumption of homoscedasticity although a small deviation from a straight line remains. In contrary to what was seen for the simulation models M1 and M2 (Fig. 4d and Fig. 7d), autocorrelation is almost absent in prediction mode, see Fig. 10d and hence the assumption of independent residuals seems justified.

In Fig. 11a the r_s of model M3 in simulation mode is shown and evidently they look heteroscedasticity dispersed. This is backed up by the range-mean plot of Fig. 11d and it looks like we still have a challenge in describing the state-dependent uncertainty satisfyingly well. The residuals resemble a Gaussian distribution relatively well which is seen from Fig. 11b although minor deviations are observed at the tails of the Q–Q plot. Autocorrelation is also present for model M3 in simulation mode, see Fig. 11d.

Comparing the confidence bounds of Fig. 12 with Fig. 13 the difference between confidence bounds generated in prediction mode with updating and simulation mode without updating is clear-cut. The confidence bounds of the one step ahead predictions are nar-

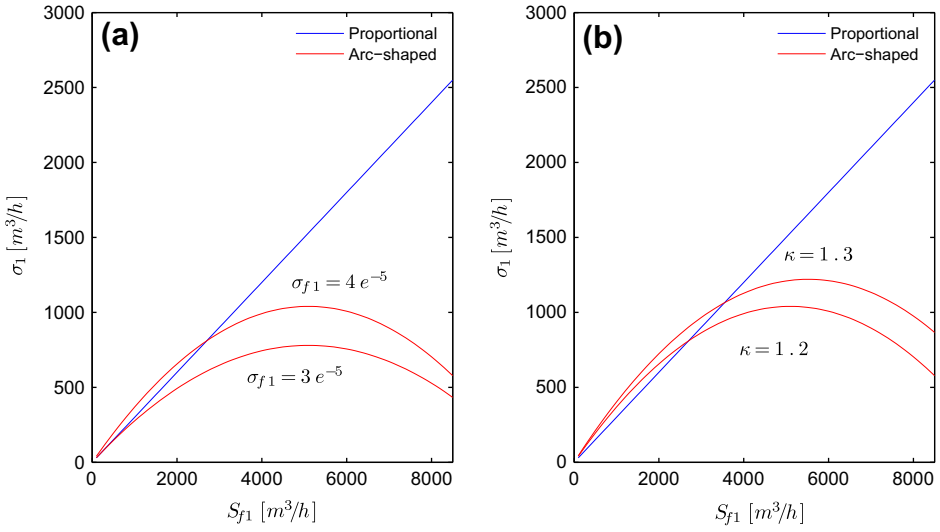

Fig. 9. Investigation of potential state dependencies.

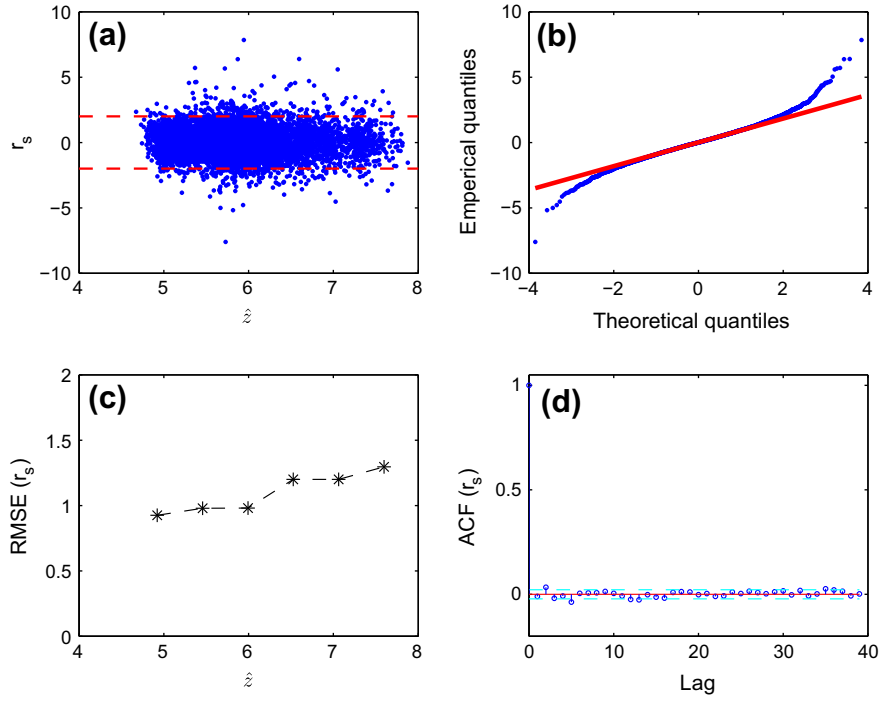

Fig. 10. Investigation of the standardised one step prediction residuals of Model M3.

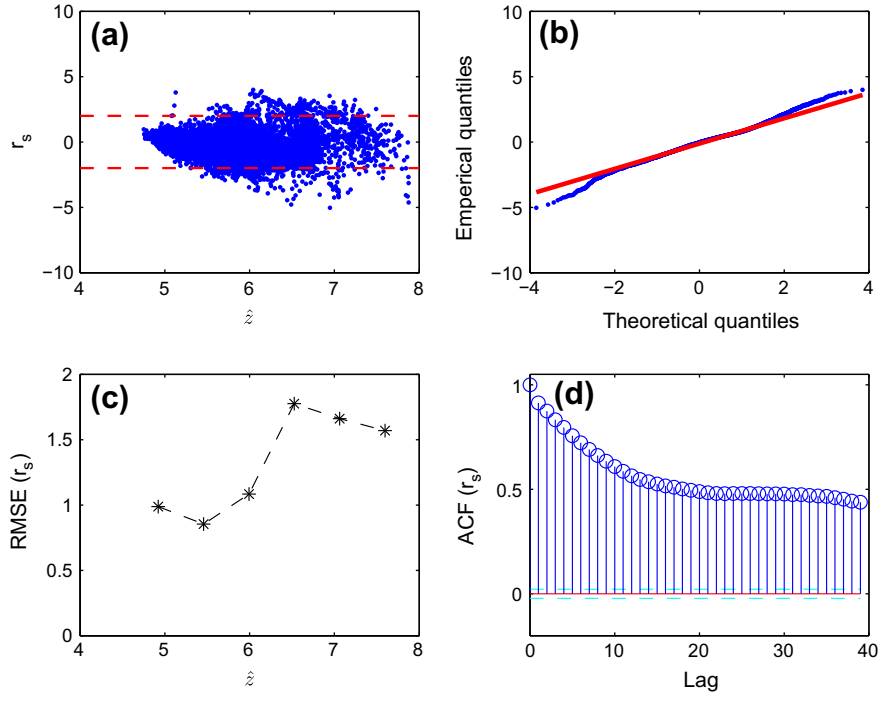

Fig. 11. Investigation of the simulation residuals of Model M3.

row and follow the observations closely, whereas the opposite is true for the confidence bounds in simulation. The width of the bounds during the days 08/18–08/28 is particularly striking. They are caused by a large uncertainty in the diffusion term of the slow infiltration reservoir S_{s3}.

If one looks closer into the data behind this, a rain event apparently passed the catchment at 08/19 since a flow peak is observed, however only a small precipitation was registered by the rain gauges and the model seems to partly compensate for this by letting more rain from a previous rain event pass the slow infiltration component of the model (by increasing K_s). The assumed input deficiency results in quite wide confidence bounds during 08/18–08/28 in simulation mode (Fig. 13).

To sum up we conclude that the statistical assumptions of the uncertainty description of M3 are roughly met for the purpose of generating one-step predictions although a minor departure from Gaussianity is observed. On the other hand, the description of the model uncertainty needs improvement to gain statistically satisfy-

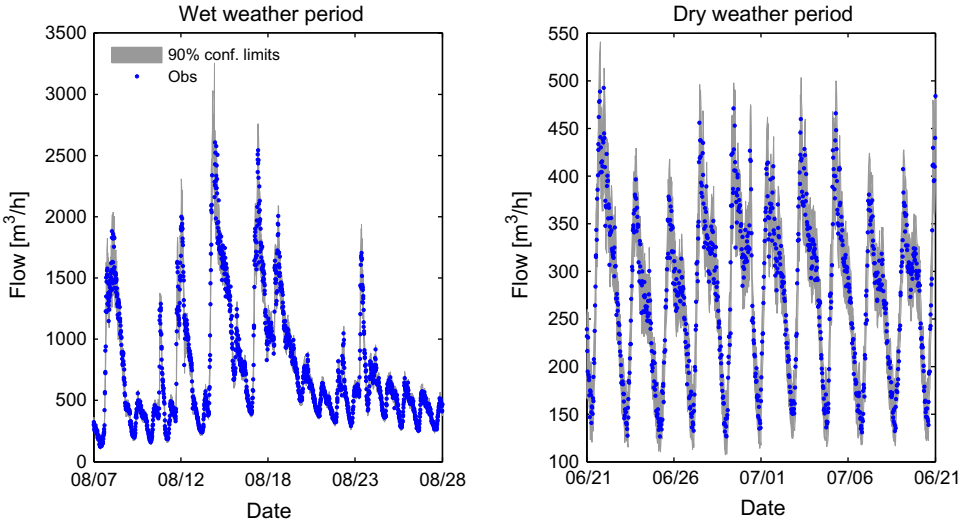

Fig. 12. 90% one step prediction confidence bounds of M3 vs. observations in a selected dry and wet weather period.

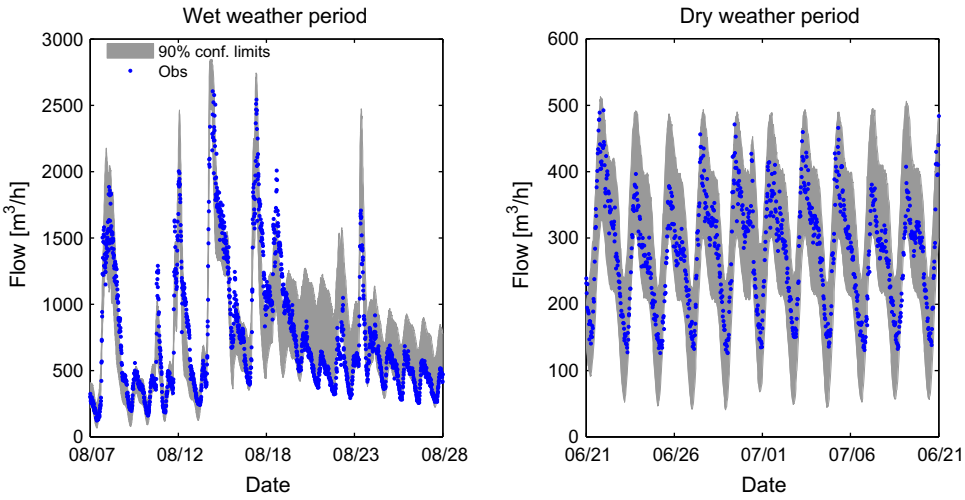

Fig. 13. 90% Simulation confidence bounds of M3 vs. observations in a selected dry and wet weather period.

ing residuals in simulation. The results presented here suggest that this is unrealistic when using a time resolution as high as applied here (15 min). One interesting approach to perhaps alleviate this problem of autocorrelation in simulation models was recently outlined by Reichert and Mieleitner (2009). They suggest to improve the model structure by analysing time-dependent parameters and inputs.

6. Model comparison using quantile performance measures

To be able to compare the models ability to generate useful confidence bounds we now take advantage of the model performance measures introduced in Section 4.2 that do not require the statistical assumptions are fulfilled. A comparison of the models sharpness and reliability as a function of the nominal coverage are shown in Fig. 14.

6.1. Reliability comparison

If we look at the *overall reliability* it is seen that simulation model M3 (denoted M3 sim in Fig. 14) is very close to the nominal coverage to all quantile levels, whereas the poorest is M1. It is somewhat surprising that simulation of M3 turns out to have a higher *overall reliability* than a one step prediction of M3 (M3 one step in Fig. 14), however if we now turn our attention to the *wet weather and dry weather reliabilities* we see that the reliability of M3 is underestimated by as much as 0.2 for nominal coverages of 0.5–0.7 during wet weather periods and overestimated by 0.1 during dry weather periods. Such large variation between dry and wet weather reliability are not seen for the prediction model M3 that is the closest to have a reliability of zero of all the models. Thus, the reason why we observe such a good overall reliability for M3 in simulation mode is simply because the poor reliability obtained in dry and wet weather periods counterbalance each other.

It is furthermore noticed that M1's reliability in dry weather periods is particularly poor, the coverage is overestimated by as much as 0.25, however this is not surprising considering that Fig. 5 already revealed that the generated 90% confidence bounds were too large and thus enclosed a higher percent of observations than intended. If we now compare the *wet weather reliability* of the simulation models, M2 outperforms the other models whereas if we turn to the *dry weather reliability* M2 and M3 show similar reliability performance.

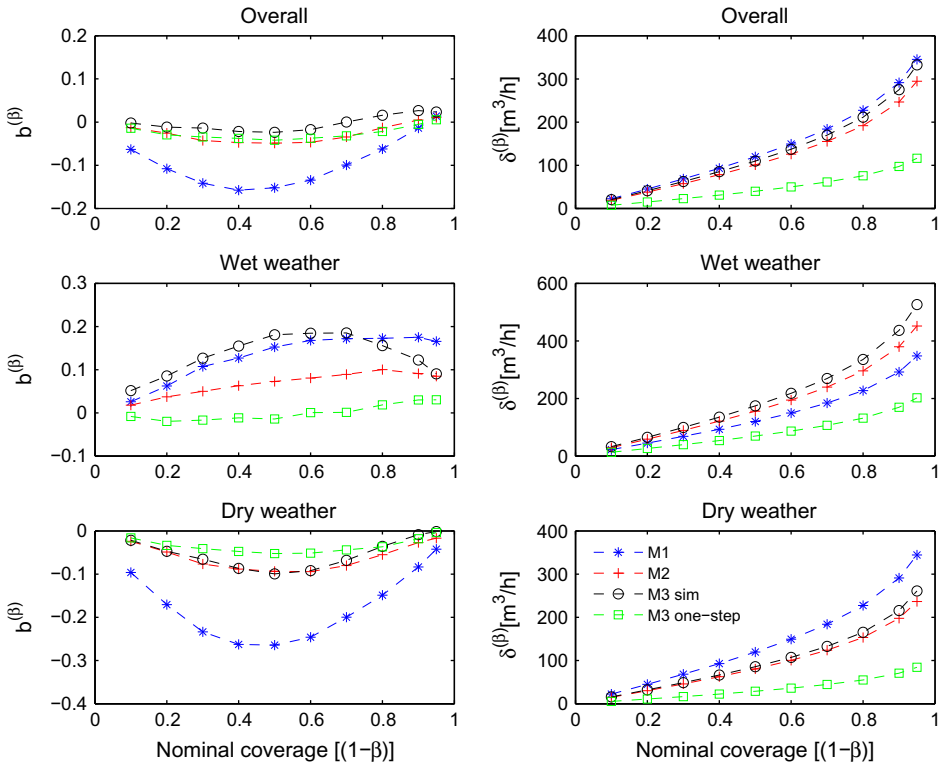

Fig. 14. Reliability bias (left) and sharpness (right) for Models M1–M3 calculated for the whole period altogether and in dry and wet weather periods separately.

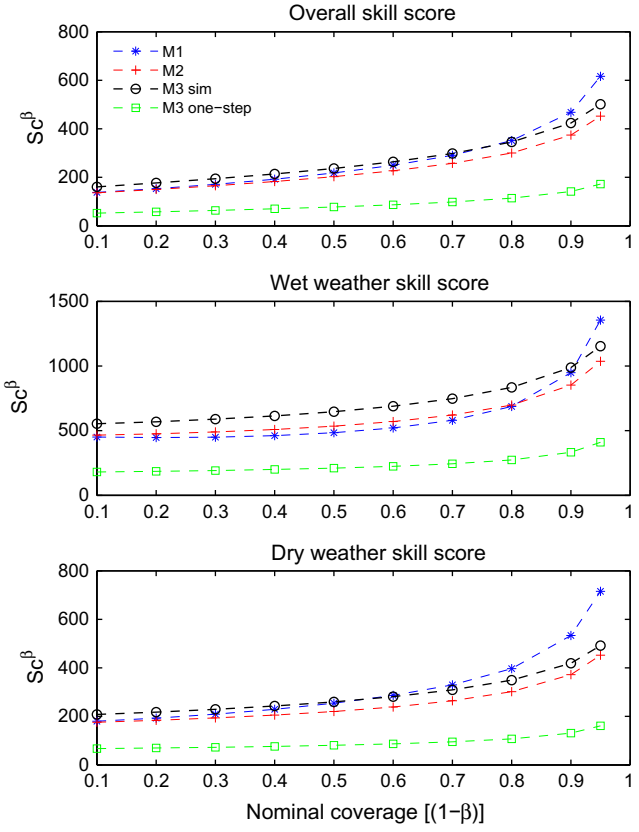

Fig. 15. Interval skill score for all models, whole period and separately for dry and weather periods.

6.2. Sharpness comparison

From the *overall sharpness* we conclude that a prediction model not surprisingly is much sharper than the simulation models. We also observe that M2 is the sharpest of the simulation models due to the fact that system states are continuously updated to the measurements. The overall sharpness reflects large deviations in dry and wet weather periods. The much too large bounds of M1 are once again identified in dry weather periods as the least sharp of the models but M1 is the sharpest in wet weather periods. In both wet and dry weather periods M2 is seen to be the sharpest of the two uncertainty developed models M2 and M3.

6.3. Interval skill score comparison

To be able to determine which model performs best we use the interval skill score. Fig. 15 displays the result hereof, and it is immediately clear that the one step ahead prediction is the best model in all three cases, hence updating improves the model performance significantly compared to a simulation model.

It is also seen from the *overall interval skill score* that M2 is the best simulation model to all nominal coverages. Perhaps surprisingly M1 is the better model in *wet weather periods* at the nominal coverages 0.1–0.8 but for the larger confidence bounds that we are normally more interested in, M2 perform better. In *dry weather periods* M1 is the poorest which reflects the poor sharpness and reliability that was observed in Fig. 14. In dry weather periods M2 is the best to all the considered nominal coverages. Note also from the *overall* skill score that even though M3 was optimised for the one step ahead prediction, M3 still performs rather similarly to the simulation models M1 and M2 in simulation mode.

7. Conclusions

In this study a formal frequentist approach to uncertainty evaluation was taken to assess the output of a conceptual hybrid sewer flow model formulated in state-space. The model accounted for three separate flow contributions: wastewater from households, fast rainfall–runoff from paved areas and slow infiltration inflow from unknown sources. Using point rain gauges and evaporation data as inputs and flow data for model conditioning the model parameters were estimated by maximum likelihood evaluation.

Three different ways to represent uncertainty in the deterministic model were investigated and denoted M1, M2 and M3. In models M1 and M2 all uncertainty sources were accounted for by an output error term assuming the residuals to follow a Gaussian white noise process using a likelihood function that minimised the errors between observed and simulated outputs for parameter estimation. Model M2 was expanded from M1 by introducing a transformation of the observations in an attempt to remove the heteroscedastic residuals structure. In model M3 stochastic differential equations were introduced which involved updating of states in each time step using the EKF, and where a diffusion term accounted for uncertainty related to inputs and model structural limitations and errors, and an observation error term accounted for flow observation errors. In M3 a transformation of the observations was also needed, but because the diffusion term accounted for most of the uncertainty, the required observation transformation was different from the transformation required in M2. For parameter estimation in M3 a likelihood function that minimised the one step prediction error was used in combination with an extended Kalman filter, and hence this model was tailored for prediction because of the continuous updating of model states in accordance with the measurements. M3 was however also evaluated in simulation mode, i.e. without updating the states.

Investigation of the standardised residuals of model M1 revealed that uncertainty from rainy periods was transferred to dry weather periods leading to exaggerated confidence bounds for low flows with absurd negative lower confidence bounds. M2 was closer to complying with the statistical error assumptions although some heteroscedasticity and deviation from Gaussianity remained, but the observation transformation solved the problem with negative lower confidence bounds. The residuals of M3 for the one step ahead prediction showed that the white noise assumption was almost perfectly met. However when applying model M3 for simulation some heteroscedasticity and autocorrelation was found, which was also the case for the other simulation models. We therefore suspect that the statistical assumption that residuals should be independent of each other is unrealistic in simulation mode when using a time resolution as high as applied here (15 min).

Due to the autocorrelated residuals in simulation mode the estimated parameters cannot be taken as the true parameter values despite a *t*-test showed the parameters were significant.

The probabilistic performances of the models were compared using three performance measures: the reliability bias, the sharpness and the interval skill score. For simulation and when considering the overall period (without separating between wet and dry weather periods) the interval skill score showed that model M2 performed best. This was also the case when conditioning on dry weather periods solely, however in wet weather periods M2 performed best only at the higher nominal coverages (>0.8) whereas M1 performed best for the nominal coverages 0.1–0.8. The unsatisfactory lower bounds of M1 in dry weather periods entails however that M1 is ill-suited for simulation. M3 in simulation mode generally performed worse than M2.

An interval skill score comparison of the simulation models with the one step prediction model clearly showed the improving effect updating of model states has on the predictive performance as the one step prediction model proved to yield much sharper but still reliable uncertainty bounds.

This study also proved the importance of checking and learning from the model residuals in the model building phase, and it should be common practise when applying a formal approach to uncertainty evaluation to show how well the actual error structure fits the assumed by the use of residual analysis. Furthermore the study showed that it is important to be explicit about the model purpose, that is, if the model is tailored for prediction or simulation. The use of the probabilistic performance measures sharpness, reliability and interval skill score are useful for checking that the assumed error structure holds in all regions of the output and additionally provides useful information for further model development. It is important to recognise that the conclusion about the invalidated parameter estimates does not imply that the conclusions about the relative model performances derived from the skill score is invalidated. A main advantage of the quantile performance measures (reliability/sharpness/skill score) is that they apply to any models that can provide probabilistic forecasts, irrespective of how the parameters of this model have been estimated and if all statistical assumptions are fulfilled. Hence the skill score can be used to select the model that provides the best predictions/simulations even when we cannot fulfill the traditional statistical requirements to our models. We believe users of formal approaches to uncertainty evaluation within hydrology and within environmental modelling in general can benefit significantly from adopting the evaluation measures applied here so the probabilistic performance of their models can be assessed properly.

Acknowledgements

We wish to thank Jacob Nørremark, Spildevandscenter Avedøere I/S for support with background information and flow meter data, and Lisbeth Brusendorff, DTU Environment for help with the graphical layout. This research was financially supported by a PhD fellowship co-founded by Krüger A/S, DTU Environment and the Ministry of Science, Technology and Innovation through the Urban Water Technology graduate school (www.urbanwatertech.dk), and by the Danish Council for Strategic Research (Storm- and Wastewater Informatics Project).

References

Andréassian, V., Lerat, J., Loumagne, C., Mathevet, T., Michel, C., Oudin, L., Perrin, C., 2007. What is really undermining hydrologic science today? Hydrol. Process. 21 (20), 2819–2822.

Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320 (1–2), 18–36.

Beven, K., 2009a. Environmental Modelling: An Uncertain Future/An Introduction to Techniques for Uncertainty Estimation in Environmental Prediction. Routledge Taylor & Francis Group.

Beven, K., Binley, A., 1992. The future of distributed models – Model calibration and uncertainty prediction. Hydrol. Process. 6 (3), 279–298.

Beven, K., Smith, P.J., Wood, A., 2011. On the colour and spin of epistemic error (and what we might do about it). Hydrol. Earth Syst. Sci. 15 (10), 3123.

Beven, K.J., 2009b. Comment on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" by Jasper A. Vrugt, Cajo J.F. ter Braak, Hoshin V. Gupta and Bruce A. Robinson. Stoch. Environ. Res. Risk Assess. 23 (7), 1059–1060.

Breinholt, A., Thordarson, F.O., Møller, J.K., Grum, M., Mikkelsen, P.S., Madsen, H., 2011. Grey-box modelling of flow in sewer systems with state-dependent diffusion. Environmetrics 22 (8), 946–961.

Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resour. Res., 47.

Dotto, C., Deletic, A., Fletcher, T., 2009. Analysis of parameter uncertainty of a flow and quality stormwater model. Water Sci. Technol. 60 (3), 717–725.

Engeland, K., Renard, B., Steinsland, I., Kolberg, S., 2010. Evaluation of statistical models for forecast errors from the HBV model. J. Hydrol. – Amsterdam 384 (1–2), 142.

Feyen, L., Vrugt, J.A., Nualláin, B.O., van der Knijff, J., DeRoo, A., 2007. Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. J. Hydrol. 332 (3–4), 276–289.
Franchini, M., Bernini, A., Barbetta, S., Moramarco, T., 2011. Forecasting discharges at the downstream end of a river reach through two simple Muskingum based procedures. J. Hydrol. 399 (3–4), 335–352.
Gallagher, M., Doherty, J., 2007. Parameter estimation and uncertainty analysis for a watershed model. Environ. Model. Softw. 22 (7), 1000–1020.
Georgakakos, K.P., Seo, D.-J., Gupta, H., Schaake, J., Butts, M.B., 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. J. Hydrol. 298 (1–4), 222–241.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction and estimation. J. Am. Statist. Assoc. 102 (477), 359–378.
Jazwinski, A.H., 2007. Stochastic Processes and Filtering Theory. Dover Publications, Mineola, New York, USA.
Jin, X., Xu, C.-Y., Zhang, Q., Singh, V., 2010. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. J. Hydrol. – Amsterdam 383 (3–4), 147.
Jørgensen, H.K., Rosenørn, S., Madsen, H., Mikkelsen, P.S., 1998. Quality control of rain data used for urban runoff systems. Water Sci. Technol. 37 (11), 113–120.
Kristensen, N.R., Madsen, H., 2003. Continuous Time Stochastic Modeling – CTSM 2.3 – Mathematics Guide. Technical Unversity of Denmark.
Kristensen, N.R., Madsen, H., Jørgensen, S.B., 2004a. A method for systematic improvement of stochastic grey-box models. Comput. Chem. Eng. 28 (8), 1431–1449.
Kristensen, N.R., Madsen, H., Jørgensen, S.B., 2004b. Parameter estimation in stochastic grey-box models. Automatica 40, 225–237.
Madsen, H., 2008. Time Series Analysis. Chapman & Hall/CRC.
Madsen, H., Thyregod, P., 2011. Introduction to General and Generalized Linear Models. CRC Press (texts in statistical science).
McMillan, H., Clark, M., McMillan, H., Clark, M., 2009. Rainfall–runoff model calibration using informal likelihood measures within a markov chain monte carlo sampling scheme. Water Resour. Res. 45 (4), W04418.
Montanari, A., 2007. What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. Hydrol. Process. 21 (6), 841–845.
Refsgaard, J., 1997. Validation and intercomparison of different updating procedures for real-time forecasting. Nord. Hydrol. 28 (2), 65–84.
Reichert, P., Mieleitner, J., 2009. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resour. Res. 45 (10), W10402. http://dx.doi.org/10.1029/2009wr007814.
Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour. Res. 46 (5), W05521b. http://dx.doi.org/10.1029/2009wr008328.
Salamon, P., Feyen, L., 2010. Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation. Water Resour. Res. 46 (12), W12501.
Smith, T.J., Marshall, L.A., 2010. Exploring uncertainty and model predictive performance concepts via a modular snowmelt–runoff modeling framework. Environ. Model. Softw. 25 (6), 691–701.
Thordarson, F., Breinholt, A., Møller, J., Mikkelsen, P., Grum, M., Madsen, H., 2012. Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion. Stoch. Environ. Res. Risk Assess., 1–12.
Todini, E., 2011. History and perspectives of hydrological catchment modelling. Hydrol. Res. 42 (2–3), 73–85.
Tomicic, B., 2006. Kapacitetsplanlægning. Opdatering af MOUSE/MIKE URBAN Model. Modeltilpasningsrapport for spildevandscenter Avedøre. Tech. Rep., Danish Hydraulic Institute (DHI), Hørsholm, Denmark.
Vestergaard, M., 1998. Nonlinear Filtering in Stochastic Volatility Models, Master Thesis. Technical University of Denmark. Department of Mathematical Modelling, Lyngby, Denmark.
Vrugt, J.A., Braak, C.J.F.t., Gupta, H.V., Robinson, B.A., 2009a. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch. Environ. Res. Risk Assess. 23 (7), 1011–1026.
Vrugt, J.A., ter Braak, C.J.F., Gupta, H.V., Robinson, B.A., 2009b. Response to comment by Keith Beven on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?". Stoch. Environ. Res. Risk Assess. 23 (7), 1061–1062.
Willems, P., 2012. Model uncertainty analysis by variance decomposition. Phys. Chem. Earth 42, 21.
Xiong, L., Wan, M., Wei, X., O'Connor, K.M., 2009. Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. Hydrol. Sci. J. 54 (5), 852–871.
Yang, J., Reichert, P., Abbaspour, K., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a swat application to the Chaohe basin in China. J. Hydrol. 358 (1–2), 1–23.
Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007. Hydrological modelling of the Chaohe basin in china: statistical model formulation and Bayesian inference. J. Hydrol. 340 (3–4), 167–182.