Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 70 (2014) 843 – 852

Procedia Engineering

www.elsevier.com/locate/procedia

12th International Conference on Computing and Control for the Water Industry, CCWI2013

Real-time data assimilation in urban rainfall-runoff models

C.J. Hutton[a*], Z. Kapelan[a], L. Vamvakeridou-Lyroudia[a], D. Savić [a]

[a]*Centre for Water Systems, University of Exeter,Harrison Building,North Park Road, Exeter, EX4 4QF, UK*

Abstract

Real-time control of urban rainfall-runoff systems can help limit flooding, and minimise combined sewerage overflow. To improve the ability of runoff models to inform this control decision, a data assimilation methodology is presented where downstream prediction errors are used to update upstream model states at an earlier time step. The methodology led to improved, 'corrected' predictions after model re-propagation to the current time, and improved discharge forecasts. Assimilation performance was sensitive to the update lag time, and the presence of control structures in the model, which affect the ability of assimilation procedures to map observation information to state space.

Selection and peer-review under responsibility of the CCWI2013 Committee

Keywords: Data-Assimilation, Error-Correction, Urban Rainfall-Runoff, Sewer System, Real-Time.

1. Introduction

Numerical modelling of urban rainfall runoff systems, when coupled with optimisation algorithms, can facilitate the identification of optimal control options in real-time (Butler and Schutze 2005; Puig et al. 2009; Schutze et al. 2004; Vanrolleghem et al. 2005). In doing so, such models can help militate against the impacts of excess rainfall and balance the potentially competing requirements of limiting urban flooding, whilst also minimising Combined Sewerage Overflow (CSO). Whilst the availability of (near) real-time observations has improved the potential to apply such models, significant uncertainties in model structure, input rainfall and model parameters, all affect the

* Corresponding author.
E-mail address: c.j.hutton@ex.ac.uk

Selection and peer-review under responsibility of the CCWI2013 Committee
doi:10.1016/j.proeng.2014.02.092

quality of model forecasts, and in turn the reliability of derived control decisions (Deletic et al. 2011; Molini et al. 2005). A range of methods have been developed to quantify and reduce uncertainties in water systems' models during off-line calibration (Dotto et al. 2012; Hutton et al. Submitted; Hutton et al. 2012a), and also in real-time, whereby system observations are assimilated to correct model predictions, states and/or parameters (for a review see: Hutton et al. 2012a). Such Data Assimilation techniques include the Ensemble Kalman Filter and Particle Filter (Hutton et al. 2012b; Pauwels and De Lannoy 2009; van Leeuwen 2009), which use an ensemble of models to quantify uncertainty in real-time predictions. The downside to these approaches is that running an ensemble of models is computationally expensive, particularly in real-time, making them potentially unsuitable for application with distributed, hydrodynamic urban runoff models, such as MOUSE (MIKE URBAN), SWMM and INFOWORKS. Furthermore, multiple model runs would be required to quantify uncertainty over the forecast time horizon in each potential scenario of an optimisation procedure, which becomes computationally expensive, particularly in real-time.

Deterministic error correction procedures have been developed that use system observations to update the states of a single system model, and in doing so are potentially better suited to fully hydrodynamic urban runoff models. These methods include MIKE UPDATE (Hansen et al. 2011), which adjusts node water levels using local system observations; real-time update of model parameters (Leonhardt et al. 2012); and an approach developed by Borup et al. (2011) where downstream measurements are used to update the slow changing flow components upstream of the observation point at the same model time-step. Whilst the residual error at the downstream observation point in an urban runoff model will reflect errors in upstream model components, these errors will propagate downstream to the observation point from previous time-steps, depending on the lag time of the system. Such a temporal lag has been considered in the application of data assimilation methods applied to natural catchment systems (Pauwels and De Lannoy 2009). Urban rainfall-runoff systems, however, may contain control structures, such as tanks, valves and gates, which introduce additional nonlinearity, and threshold behaviour, which may complicate the ability of data assimilation techniques to map observational information to state space. This paper presents the development of a deterministic data assimilation correction methodology, based on the principle that differences between measured and predicted system state variables results from errors in upstream model components at previous time-steps.

2. Methodology

The Data Assimilation method is based on a general methodology, outlined in Figure 1, and based on the concept presented in Pauwels and De Lannoy (2009). First, the runoff model, with measured rainfall as an input, is propagated forwards in time from time step $t\text{-}1$ to time step t, the point at which a system observation (e.g. pipe flow rate) becomes available. Second, the observation is compared to the equivalent model prediction (e.g. predicted pipe flow rate) at the same point in the system, to derive the residual error between model prediction and observation. The model states (e.g. pipe flow rates in the sewer model, or upstream catchment model components such as the surface store) are then re-set to their values saved at time $t\text{-}n$, where n is the system lag between a given upstream model state to be corrected, and the downstream observation. The upstream state (e.g. linear reservoir level representing surface storage) is then corrected using the downstream residual error, and the runoff model re-propagated forwards from $t\text{-}n$ to t to derive a 'corrected' estimate of the system states, including at the measurement location (e.g. of the downstream pipe flow rate). The corrected model states at time t then set the initial conditions from which to make a forecast.

The general approach is based on the fact that urban runoff systems are not in steady state; that is, rainfall enters the system and is propagated downstream, through the catchment, into the sewer network, and to a potential outlet (e.g. Waste Water Treatment Plant, or CSO). Differences between downstream measurements and model predictions in for example pipe flow rate result from errors in input rainfall, and also errors in the model structure and parameters, which lead to errors in model state estimates (e.g. linear reservoir levels and pipe flow rates).

Corrected these states requires that the lag time in the system is considered between the state to be updated, and the measurement location.

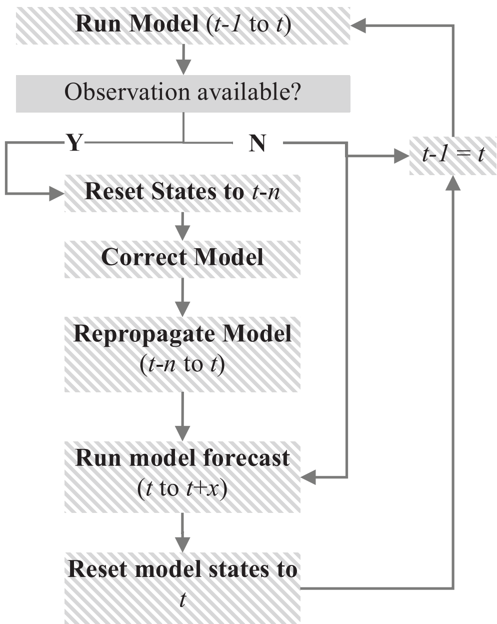


Figure 1. Flow diagram illustrating the Data Assimilation Methodology.

In the study presented here, the general methodology evaluated employs a deterministic Kalman Filter type update (Kalman, 1960) to an upstream model state (x_{t-n}), at a previous time-step $(t-n)$, based on the residual error between forecasted (superscript *f*), and observed (superscript *o*) downstream state (q_t) at the current time-step (t):

$$x_{t-n}^{a} = x_{t-n}^{f} + \left.\frac{\partial x_{t-n}^{f}}{\partial q_t^{f}}\right|_{q_t^{f}} \left(q_t^{o} - q_t^{f}\right) \tag{1}$$

Where the superscript a indicates the analysed (or corrected) state estimate, and the final right hand term is the correction to the forecasted upstream model state, derived from the product of downstream residual error and the first derivative between x_{t-n} and q_t. The method to determine this first order relation is considered below.

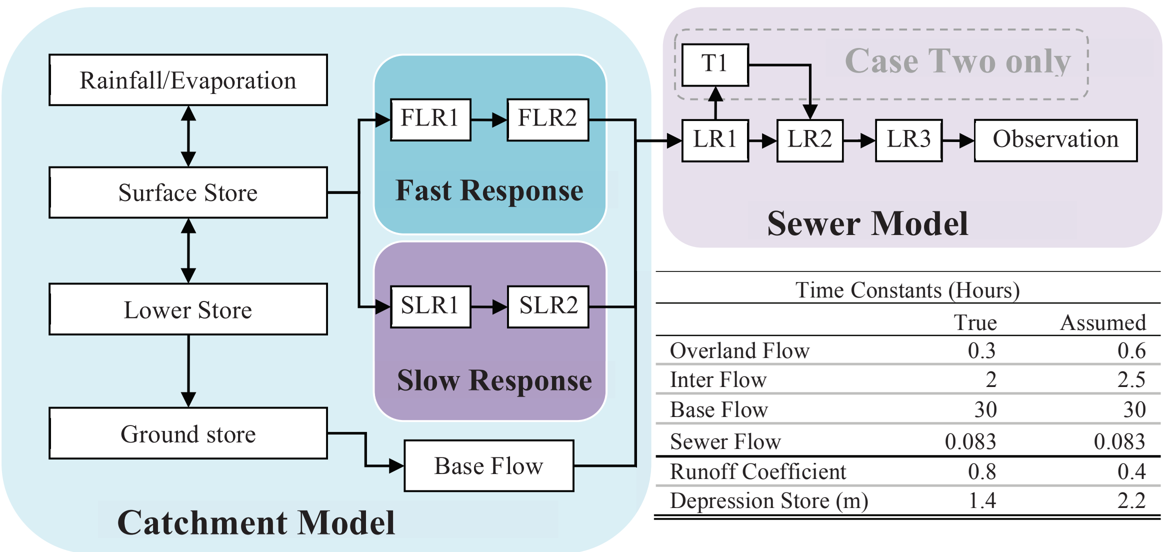
3. Case Study

The methodology is tested using a conceptual rainfall runoff model of a 20ha catchment (Figure 2). The catchment is represented in the model using a series of reservoirs, as applied in many rainfall-runoff models to simulate different water storage and catchment runoff responses (e.g. Madsen, 2000). The model consists of a surface store (linear reservoir) which routes flow via linear reservoirs representing fast and slow runoff responses to a simplified sewer system, consisting of a further three linear reservoirs. The model parameters consist of four time constants governing the response time of the linear reservoirs, and a Depression Store and Runoff Coefficient, which govern surface store flow routing to the fast and slow runoff response linear reservoirs. For a set of ‘true’ model parameters (Figure 2), the model is driven by a 16 day rainfall time-series to generate discharge observations at the downstream end of the sewer model, the “Observation” location depicted in Figure 2. For the assimilation runs, model parameters are adjusted to a set of ‘assumed’ parameters (Figure 2) to represent rainfall-runoff model error (e.g. parameter uncertainty). No additional noise is added to the observations, and no additional uncertainty is added to the model structure. To represent rainfall uncertainty, the rainfall time-series is perturbed by removing a 25% heteroscedastic bias (to represent under-estimation of rainfall, as occurs for tipping bucket rain-gauges, particularly at higher rainfall rates; Molini et al., 2005), and a uniform random noise component added, sampling between 0 and 25% of the 'true' rainfall.

4.1 Case One

In the first case considered, the data assimilation procedure is applied to the runoff model considered above to evaluate the sensitivity of the assimilation procedure to the frequency of observations/updates (between 1 and 40 minutes), and also to the time lag n in the system where the state is updated, prior to re-propagation of the model to the current time step (between 1 and 250 minutes), giving a total of 10,000 sensitivity analysis runs.

In order to correct the surface store at time-lag n, a first order relation between the downstream discharge observation at time t and the Surface Store at t-n need to be established. The rainfall-runoff model was run offline to generate a time-series of Surface Store states and corresponding downstream discharge observations from which to establish this relationship. For each potential time lag used in the sensitivity analysis, the gradient of the linear relationship was calculated between the time-series, as used in equation (1). The coefficient of determination for these relationships varied from 0.5 at a lag of 250 minutes to a peak of 0.77 at a lag of 94 minutes.



Time Constants (Hours)	True	Assumed
Overland Flow	0.3	0.6
Inter Flow	2	2.5
Base Flow	30	30
Sewer Flow	0.083	0.083
Runoff Coefficient	0.8	0.4
Depression Store (m)	1.4	2.2

Figure 2. Runoff Model Structure alongside the true and assumed model parameters used in the experiments. In the second case study considered, a storage tank is added to the system.

For each of the 10,000 sensitivity analysis runs, the assimilation procedure was initiated after an initial simulation period of 6,000 minutes, following which assimilation performance was evaluated by calculating the Mean Absolute Error (MAE) between the corrected model predictions at time *t* and the observations.

Regardless of the time lag considered in the sensitivity analysis, the performance of the assimilation methodology improved with an increase in the frequency of the observations (Figure 3A). The more frequent observations are assimilated relative to the lag in the system, the closer the methodology approximates an optimization procedure as iterative adjustments are made to the model states more frequently than changes in the system. As observations become less frequent, the errors in the model and rainfall move the system predictions away from the observations.

For a one minute update frequency, optimum performance of the correction procedure occurred at a lag of 126 minutes with a MAE of $2.2 \times 10^{-4} m^3 s^{-1}$, which is approximately 5% of the uncorrected, open loop simulation error of $3.8 \times 10^{-3} m^3 s^{-1}$ (Figure 4). The optimum time-lag for update of the surface store reflects a balance between the time it takes discharge to move through the reservoirs representing the fast and slow runoff responses before reaching the catchment outlet. When only peak discharges were considered ($>0.025 m^3 s^{-1}$), optimal performance occurred at a shorter lag of 96 minutes, with an MAE of $8.9 \times 10^{-5} m^3 s^{-1}$. This time lag is shorter than when considering the error in all discharges as the peak discharges result from the fast runoff response, which has a shorter lag time. Furthermore, the MAE is smaller when only considering the peak discharges, suggesting that the error for all discharges reflects a tradeoff in the optimum time-lag between correcting the errors in the fast and slow runoff responses.

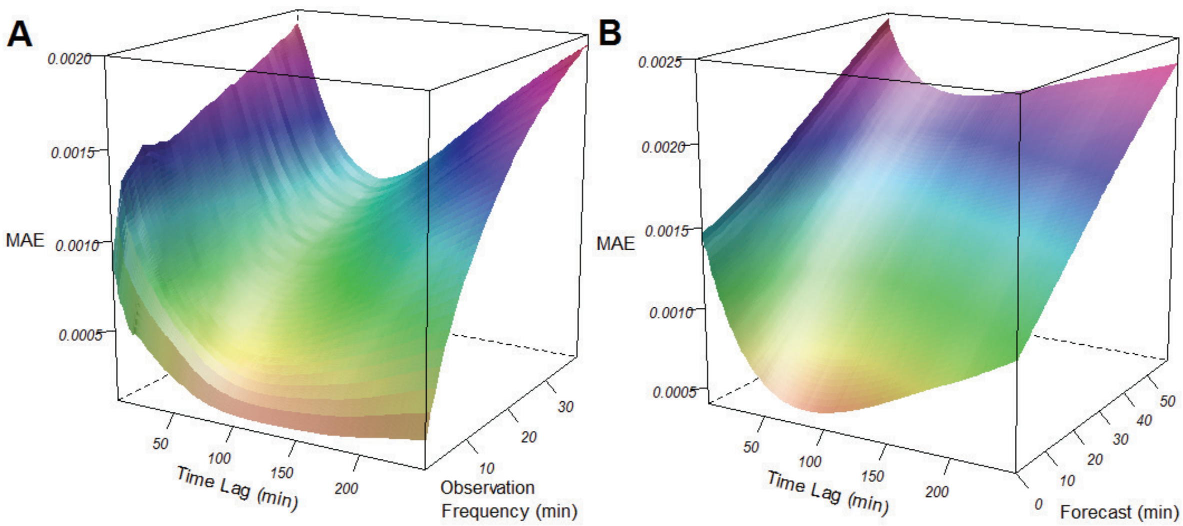


Figure 3. Mean Absolute Error (MAE) of corrected discharge predictions calculated for all corrected observations as a function of the time lag *n* (e.g. observation frequency) and the update frequency (frequency of observations) (A); forecast accuracy as a function of lead time and update time lag for an observation frequency of 10 minutes (B).

At a more realistic observations frequency of 10 minutes, optimum correction occurred at a lag of 87 minutes, with a MAE of $8.3 \times 10^{-4} m^3 s^{-1}$. This result suggests that a decline in observation frequency allows the predictions to move away from the observations, resulting in a worse assimilation performance after the next observation is assimilated. The shorter optimal time-lag of 87 minutes is closer to the fast runoff response time lag, even though the MAE reflects all discharge magnitudes. Therefore, this movement away from the observations appears to be worse at larger discharges (Figure 4), and hence the optimum time-lag that minimises all errors is closer to the faster runoff response time-lag.

As expected, forecast accuracy declines with an increase in forecast lead time (Figure 3B), yet for up to the 60 minute forecasts accuracies shown, the data assimilation procedure allows the model still to produce better predictions that the open loop simulations. Figure 4 shows selected 60 minute forecasts to illustrate the improvement in model performance. However, on the rising hydrograph limb, despite the update, the model forecasts move away from the observations because of model error introduced when using the "assumed" parameters (Figure 2).

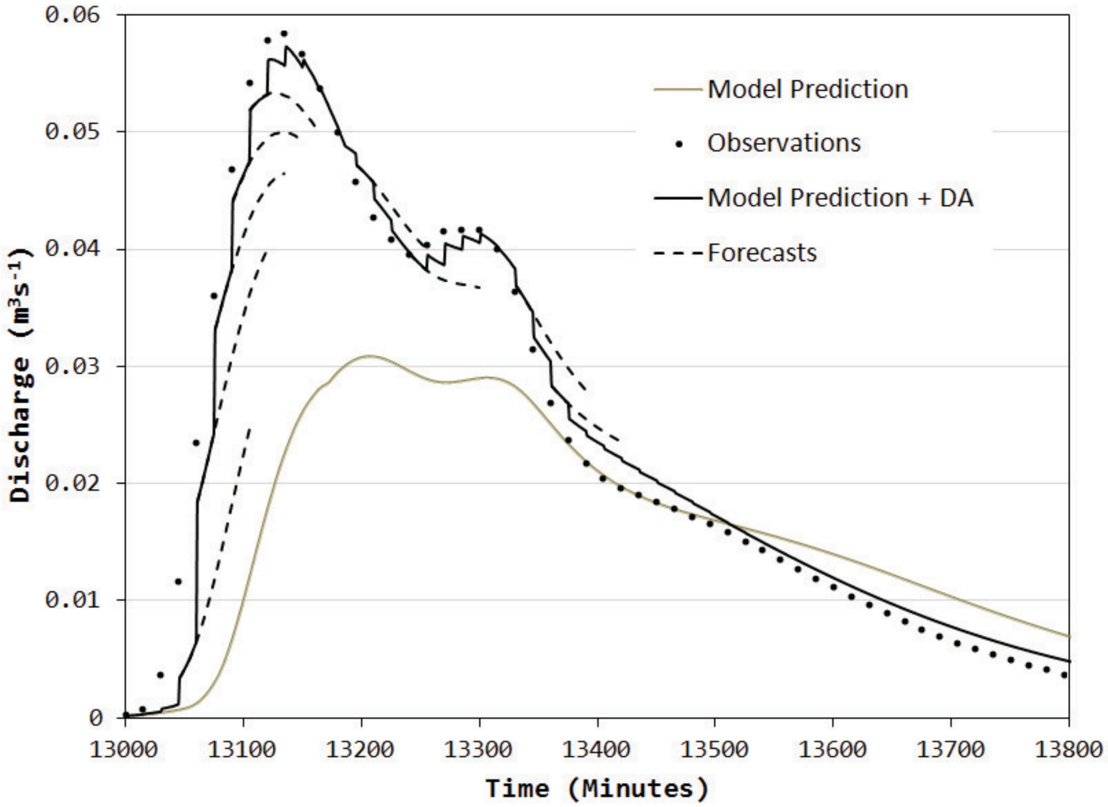


Figure 4. Modelled downstream runoff (Prediction) and modelled downstream runoff with Data Assimilation (Prediction + DA) in comparison to the 15 minute observations used in the DA procedure with a time lag of 100 minutes, for a section of the runoff time-series. The thin black lines show selected 60 minute forecasts following DA.

4.1 Case Two

In the second case study considered, the conceptual urban rainfall-runoff model used in the first case study (Figure 2) is modified with the addition of a storage tank to evaluate how well the data assimilation methodology works in the case where system storage and control influences threshold behaviour to the system (Figure 2). In this configuration, when LR1 in the sewer model reaches capacity, additional flow is routed to the storage tank, which has its own, slower runoff time constant of 2 hours (governing flow to LR2), and a maximum storage capacity of 400m^3. Once the storage capacity of the tank is exceeded, additional flow is routed to LR2.

The result of this change on system performance is to attenuate downstream discharge and reduce peak runoff (Figure 5). In addition, the tank introduces threshold behaviour into the system, producing constant discharge at the catchment outlet whilst the tank is filled, which also has the effect of delaying the timing of the hydrograph peak discharge once the tank capacity is exceeded.

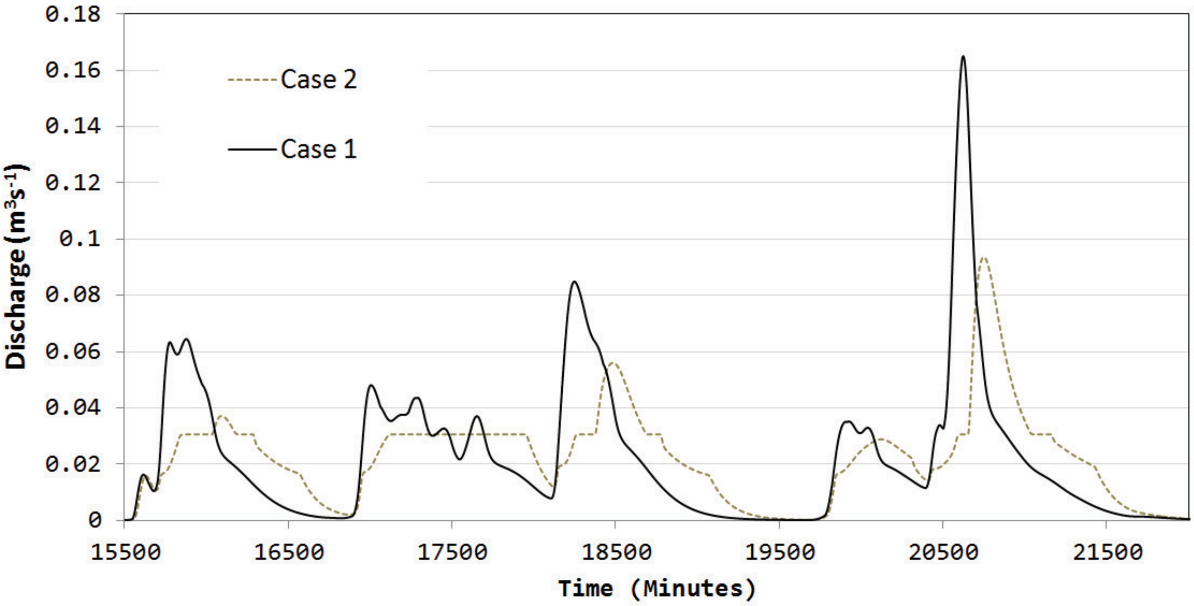


Figure 5. Comparison of downstream discharge produced from the conceptual rainfall-runoff model for Case 1, and for Case 2 where a storage tank is introduced into the network (see Figure 2 for structures).

Sensitivity analysis of the data assimilation method conducted for case one was repeated for the new system; however given the additional storage and resultant delay in runoff response, the time lag *n* was varied between 1 and 450 minutes. As in the previous case, assimilation performance improved with an increase in the frequency of observations, with an optimal assimilation performance MAE of 1.3×10^{-4}m^{3}s^{-1} at a time-lag of 338 minutes (Figure 6). This figure is approximately 5% of the uncorrected, open loop simulation error of 2.9×10^{-3}m^{3}s^{-1}. At observation frequencies of 10 minutes an optimal MAE of 6.6×10^{-4}m^{3}s^{-1} was produced with a lag of 313 minutes, and at 15 minute frequencies, an MAE of 9.0×10^{-4}m^{3}s^{-1} at a lag of 323 minutes. The increase in optimal time lag for system state correction results from the additional tank storage in the model, which increases the lag time between the state of the upstream surface store and its propagation downstream to the observation location.

The performance of the data assimilation procedure is affected by the presence of the tank in the system (Figure 7). A constant discharge is observed at the downstream end of the system once the capacity of LR1 is reached. This results because excess discharge enters and fills the tank, which has a much larger response time. The effect of this threshold behavior on assimilation performance is to effectively de-couple the response of the downstream catchment to changes in upstream behavior. Thus, whilst the tank fills, there is no change in the observed downstream discharge, which reduces the ability of assimilation procedure to correct model states upstream of the tank. Only when the tank is full and discharge starts to increase again downstream can the assimilation procedure map the observations downstream to upstream state space, and update the surface storage. The phenomenon is most notably observed at approximately 18,400 minutes in Figure 7, where downstream discharge increases by approximately one third, as the observation is used to correct the surface store. As the model is re-propagated forwards in time from *t-n* to *t* the corrected discharge propagates downstream, filling the tank to the level consistent with the observations further downstream, and producing the observed excess discharge.

Despite the issue in assimilation performance that results from the threshold behavior introduced into the model, data assimilation leads to much better predictions than the open loop simulation, which fails to predict the hydrograph peak in Figure 7. The under-estimation of rainfall and modification to parameters means that the model without assimilation does not fill the tank, and therefore fails to adequately reproduce the hydrograph peak discharge.

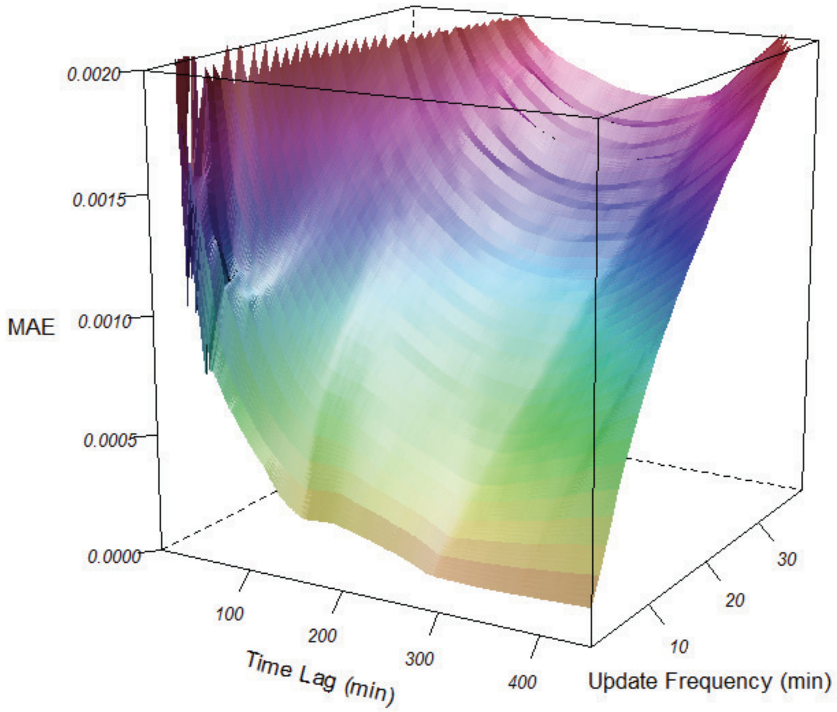


Figure 6. Mean Absolute Error (MAE) of corrected discharge predictions calculated for all corrected observations as a function of the time lag *n* (e.g. observation frequency) and the update frequency (frequency of observations) for Case 2 where a storage tank is introduced into the model.

4. Discussion and Conclusions

The Data assimilation procedure presented in this paper is based on the general concept that residual errors in downstream model predictions at observation locations result from errors in upstream model components. In the case study considered, the update of the catchment surface store in the rainfall-runoff model led to improved state estimates, and subsequently to improved downstream discharge forecasts. Update performance was sensitive to the lag time used in the state update, which resulted from the need to balance the response times of both the fast and slow runoff responses of the catchment when updating upstream model states. The method may be improved by updating multiple system states, depending on their lag response relative to the observation location; however, multiple updates may lead to instability in the model response (Borup et al., 2011).

In the second case study considered, the presence of threshold system behavior affects the ability of data assimilation procedures to map observation information to state space. Such a problem is of more concern to urban rainfall-runoff models, in comparison to natural runoff systems, where data assimilation methods have been more widely applied, given the former typically contain control structures that can lead to stronger nonlinearities and threshold performance in system behavior. However, despite this problem, the simulations conducted in this study demonstrate that assimilation is required to avoid, or rather correct the divergence in system prediction that can result from such system behavior. Further work is required to incorporate uncertainties in downstream observations into the assimilation procedure, and also to consider how such data assimilation procedures may be used with models to improve real-time control optimization, where the optimized control settings will subsequently feed back to affect how observation information is mapped to state space.

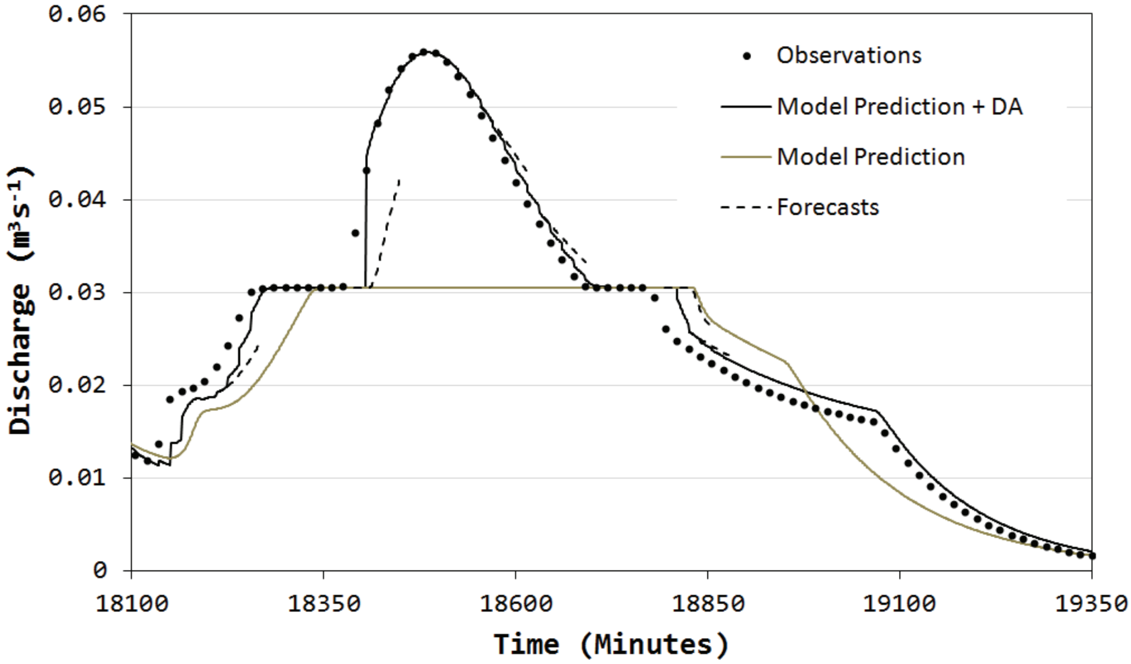


Figure 7. Modelled downstream runoff without (Prediction) and with Data Assimilation (Prediction + DA) in comparison to the 15 minute observations used in the DA procedure with a time lag of 323 minutes, for a section of the runoff time-series, using the conceptual model depicted in Figure 1 with a tank (Case two). The thin black lines show selected 60 minute forecasts following DA.

Acknowledgements

The work presented in this paper is part of 'PREPARED, Enabling Change', an ongoing European Commission Seventh Framework funded large scale integrating interdisciplinary project (Grant agreement no.: 244232, 2010-2014). PREPARED targets the urban waste water, drinking water and storm water infrastructure, aiming to improve the ability of cities to cope with new climate change scenarios, in an optimal way.

References

Borup, M., Grum, M., and Mikkelsen, P. S. 2011. Real time adjustment of slow changing flow components in distributed urban runoff models. 12th International Conference on Urban Drainage, Porto Alegre, Brasil.

Butler, D., and Schutze, M. 2005. Integrating simulation models with a view to optimal control of urban wastewater systems. Environmental Modelling & Software, 20(4), 415-426.

Deletic, A., Dotto, C. B. S., McCarthy, D. T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T. D., Rauch, W., Bertrand-Krajewski, J. L., and Tait, S. 2011. Assessing uncertainties in urban drainage models. Physical and Chemistry of the Earth, doi:10.1016/j.pce.2011.04.007.

Dotto, C. B. S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., Freni, G., Rauch, W., and Deletic, A. 2012. Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. Water Research, 46(8), 2545-2558

Hansen, L. S., Borup, M., Moller, A., and Mikkelsen, P. S. 2011. Flow Forecasting in Urban Drainage Systems using Deterministic Updating of Water Levels in Distributed Hydraulic Models. 12th International Conference on Urban Drainage, Porto Alegre, Brasil.

Hutton, C., Kapelan, Z., Vamakeridou-Lyroudia, L. S., and Savic, A. D. Submitted. The Application of Formal and Informal Bayesian Methods for Water Distribution Hydraulic Model Calibration. Journal of Water Resources Planning and Management-Asce.

Hutton, C. J., Kapelan, Z., Vamakeridou-Lyroudia, L. S., and Savic, D. 2012a. Dealing With Uncertainty in Water Distribution Systems' Models: A Framework for Real-time Modelling and Data Assimilation. Journal of Water Resources Planning and Management-Asce.

Hutton, C. J., Vamakeridou-Lyroudia, L. S., Kapelan, Z., and Savic, D. 2012b. Quantifying and Reducing Urban Water System Model Prediction Uncertainty Through Sequential Monte Carlo Sampling. HIC 2012- 10th International Conference on Hydroinformatics, Hamberg, Germany.

Kalman, R. E. 1960. A New Approach to Linear Filtering and Prediction Problems, Transaction of the ASME—Journal of Basic Engineering 82, 35-45.

Leonhardt, G., Schopf, M., Kleidorfer, M., Diao, K., and Rauch, W. 2012. Updating of a concpetual uban drainage model for online operation. 9th International Conference on Urban Drainage Modelling, Belgrade, Serbia.

Madsen, H. 2000. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. Journal of Hydrology, 235, 276-288

Molini A, Lanza LG, Barbera PL. 2005. The impact of tipping-bucket rain gauge measurement errors on design rainfall for urban-scale applications. Hydrological processes 19, 1073–1088.

Pauwels, V. R. N., and De Lannoy, G. J. M. 2009. Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. Water Resources Research, 45. W08428, doi:10.1029/2008WR007590

Puig, V., Cembrano, G., Romera, J., Quevedo, J., Aznar, B., Ramon, G., and Cabot, J. 2009. Predictive optimal control of sewer networks using CORAL tool: application to Riera Blanca catchment in Barcelona. Water Science and Technology, 60(4), 869-878.

van Leeuwen, P. J. (2009). Particle Filtering in Geophysical Systems. Monthly Weather Review, 137(12), 4089-4114.

Schutze, M., Campisano, A., Colas, H., Schilling, W., and Vanrolleghem, P. A. 2004. Real time control of urban wastewater systems: where do we stand today? Journal of Hydrology, 299(3-4), 335-348.

Vanrolleghem, P. A., Benedetti, L., and Meirlaen, J. 2005. Modelling and real-time control of the integrated urban wastewater system. Environmental Modelling & Software, 20(4), 427-442.

Wolfs, V., Villazon, M. F., and Willems, P. 2012. Development of a semi-distributed model identification and calibration tool for conceptual modelling of sewer systems. 9th International Conference on Urban Drainage Modelling, Belgrade, Serbia.