Assimilating flow and level data into an urban drainage surrogate model for forecasting flows and overflows

Schou Vorndran Lund, Nadia; Madsen, Henrik; Mazzoleni, Maurizio; Solomatine, Dimitri; Borup, Morten

Published in:
Journal of Environmental Management

Link to article, DOI:
10.1016/j.jenvman.2019.05.110

Publication date:
2019

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Assimilating flow and level data into an urban drainage surrogate model for forecasting flows and overflows

Nadia Lund*, Henrik Madsen², Maurizio Mazzoleni³, Dimitri Solomatine³,⁴,⁵, and Morten Borup¹

* corresponding author: nalu@env.dtu.dk

¹ Department of Environmental Engineering (DTU Environment), Technical University of Denmark, Bygningstorvet, Building 115, 2800 Kgs. Lyngby, Denmark
² DHI, Agern Allé 5, 2970 Hørsholm, Denmark
³ IHE Delft Institute for Water Education, Delft, The Netherlands
⁴ Water Resources Section, Delft University of Technology, Delft, The Netherlands
⁵ Institute for Water Problems, Russian Academy of Science, Moscow, Russia

Abstract

It is crucial to be able to forecast flows and overflows in urban drainage systems to build good and effective real-time control and warning systems. Due to computational constraints, it may often be unfeasible to employ detailed 1D hydrodynamic models for real-time purposes, and surrogate models can be used instead. In rural hydrology, forecast models are usually built or calibrated using long historical time series of, for example, flow or level observations, but such series are typically not available for the ever-changing urban drainage systems. In the current study, we therefore used a fast, reservoir-based surrogate forecast model constructed from a distributed urban drainage model. Thus, we did not rely directly on historical time series data. Forecast models should preferably be able to update their internal states based on observations to ensure the best initial conditions for each forecast. We therefore used the Ensemble Kalman filter to update the surrogate model before each forecast. Water levels or flow observations were assimilated into the model either directly, or indirectly using rating curves. The model forecasts were validated against observed flows and overflows. The results showed that model updating improved the forecasts up to two hours ahead, but also that updating using water level observations resulted in better flow forecasts than assimilation based on flow data. Furthermore, updating with water level observations was insensitive to changes in the noise formulation used for the Ensemble Kalman filter, meaning that the method is suitable for operational settings where there is often little time and data for fine-tuning.

Keywords

CSO, data assimilation; Ensemble Kalman filter; flow forecasts; surrogate model; urban drainage
1 Introduction

Many cities around the world are currently facing issues with changing precipitation patterns and increased urbanization, which may lead to an increased amount of runoff (Angel et al., 2011; Kaspersen et al., 2017). The sewer systems in many older cities were constructed more than 100 years ago. They are therefore not designed to cope with the increased inflow, which leads to discharges to the environment through sewer overflows or bypasses at the wastewater treatment plants. We can use real-time urban drainage models to investigate and diminish these negative effects. The models span from simple linear reservoir models to detailed 1D hydrodynamic models (high-fidelity (HiFi) models) such as MIKE URBAN, SWMM and InfoWorks.

Many utilities already have HiFi models of their systems for planning purposes. The main focus of such models is to investigate the impact of system alterations, and to inspect if the system complies with the design criterion, such as frequency of full running pipes or flooding occurrences. They can be constructed almost entirely from the utility company’s asset database with information on pipe locations, diameters, etc. Sometimes, such models can even be used without any calibration and still provide satisfactory results if the asset database is well maintained and the hydrology is simple (Borup et al., 2016). When they are calibrated, this is often done as a lumped correction of the catchment properties based on a single (often downstream) in-sewer observation (see, for example, Barco et al. (2008); Rasmussen et al. (2008); and Tscheikner-Gratl et al. (2016)).

Forecast models are important in supporting urban drainage management, for example, in connection with model predictive control (MPC) of urban drainage systems (Lund et al., 2018a). HiFi models are often not applicable in real-time due to the computational demand, which is the main incentive for using simplified models; however, it is typically at the expense of loss of performance. In rural hydrology, forecast models are therefore often conceptual models derived from and/or calibrated on long time series of historical data. There are also examples of this type of models in urban drainage research (see, for example, Löwe et al. (2014); Pedersen et al. (2016); and Thorndahl et al. (2013)), but such models are usually not applied in operational settings due to a general lack of long historical time series data that can be used to build and calibrate the models. This lack of useful time-series data is partly due to the cost of maintaining sensors in the harsh environment in the sewer systems, but also because sewer systems are ever-changing. Hence, historical sensor data might not be useful for constructing and/or calibration of simpler and faster forecast models for current conditions (including both conceptual and data-driven models). The collection of new observations can be both expensive and time-consuming. Instead, a surrogate that emulates the response of an existing HiFi model can be constructed (see, for example, Carbajal et al. (2017); Thrysøe et al. (2019); and Wolfs et al. (2013)), which due to the reasonably accurate representation of the physical system components in the HiFi model may be applicable without further calibration. Compared to purely data-driven models (for example, neural networks) such emulators furthermore have the advantage of being able to
estimate conditions outside the range previously observed by sensors in the system, and to estimate
flows and levels in locations without sensors.

When there is sensor data from the urban drainage system, it is normally present as either level or
flow measurements. Level sensors are cheaper, more reliable and easier to maintain than flow sensors,
which makes level measurements the most common type of observation. It can be difficult, however,
to build forecast models based solely on level measurements, since these do not directly relate to the
flow.

Any model is only a partial representation of reality and cannot fully replicate the dynamics of the
real system. Thus, models are subject to different sources of uncertainty which negatively affect their
performance. In urban drainage systems, the deviations between observations and model can arise
from uncertainties in the model structure (as, for example, an improper description of non-linear
processes, such as backwater effects, or the exclusion of interactions between the drainage system and
the surroundings through infiltration/exfiltration), uncertain parameter values, uncertain initial
conditions (for example, the initial wetting of the soil affecting the amount of runoff), as well as
uncertainties in measurements of precipitation, flow, level, etc. (Thorndahl et al., 2008). Data
assimilation (also denoted ‘model updating’) can be used to prevent the model from drifting away
from reality and for obtaining a better estimate of the current system state (Houser et al., 2012; Hutton
et al., 2010). Data assimilation aims at improving model performances and forecasts by combining the
knowledge from both the model and the observations to dynamically update model states, input,
parameters and/or output (Refsgaard, 1997).

Data assimilation in rural hydrology is a relatively well-proven field of research (Liu et al., 2012).
The Kalman filter (KF) is a classic data assimilation scheme, which relies on assumptions of linear
models and where the input, model and observations noises are assumed to be normally distributed.
Using the KF to update non-linear models can lead to sub-optimality and instability issues (Abaza et
al., 2015). This has inspired the development of other variations of the filter, such as the Extended
Kalman filter and the Ensemble Kalman filter (EnKF), where the EnKF furthermore has the
advantage of being computational efficient for large models (Evensen, 1994).

For a filter to be relevant for urban drainage modelling, it needs to be able to handle non-
linearities, as sewer systems have a highly non-linear behaviour with both backwater effects and
internal and external overflows. Literature on data assimilation in urban drainage models is, contrary
to rural hydrology, relatively limited. Borup et al. (2014) hypothesize that this is because ambitions
with respect to real-time control and flow forecasting have historically been low, and that
computational power and data acquisition technology, which makes it possible to apply real-time
updating, has only developed sufficiently in recent years. The reported data assimilation techniques
applied within urban drainage span from simple methods, where uncertainty is disregarded, to quite
sophisticated methods. For example, Hansen et al. (2014) applied simple direct insertion to correct the
water level in a HiFi model. Here, the states were directly changed to the value of the observation, and
uncertainty estimations were thus not required. Direct insertion may lead to model shocks that cause
the simulated outflow at the updated location to oscillate. Furthermore, it can only be used to
propagate the information to downstream parts (Hansen et al., 2014; Houser et al., 2012). Borup et al.
(2011) indirectly updated the flows in a HiFi model by employing a dampened feedback from
downstream model residuals to the hydrological states of the model, while Hutton et al. (2014)
updated the states of a simplified model based on flow measurements, using a deterministic Kalman
filter-type update. Both Borup et al. (2011) and Hutton et al. (2014) took the time lag of the system
into account. Branisavljevic et al. (2014) applied the Extended Kalman filter to update the states in a
simplified model based on water level and flow observations, and further found that subsequent bias
reduction can improve the results. Borup et al. (2014, 2018) showed that the EnKF can be used to
update the water levels throughout a HiFi model based on flow as well as water level observations.
Even though the EnKF is much cheaper to run for large models than a traditional KF, the propagation
of the ensemble forward in time is still computationally expensive for large HiFi models. Borup et al.
(2014) used an ensemble size of 10; however, the extent of the study was limited and they did not
produce flow forecasts. The presented examples demonstrate the applicability of data assimilation to
urban drainage models, and at the same time show that this area allows for (and requires) much more
research.

This study proposes a methodology for forecasting both flows and overflows using a distributed
conceptual piecewise-linear reservoir model as surrogate for a HiFi model, without relying on long
historical time series for constructing or calibrating the model. While the surrogate model is
constructed purely from the parameters and results from the HiFi model, the surrogate model states
are updated every time a new water level or flow observation is available using the EnKF as data
assimilation method. The updated states are used as initial conditions for forecasting with the
surrogate model. Since level gauges are the preferred sensors in urban drainage systems, we
investigate the impact of using flow and level data, respectively, for the updating. The observations
are both used directly in the data assimilation scheme, or converted via rating curves. The focus on
observation type and independence of long historical time series has, to our knowledge, not been the
focus of previous data assimilation research within urban drainage. A small sub-catchment in
Copenhagen, Denmark is used as case study. The case area has a downstream CSO structure, and the
surrogate model is used to make forecasts of the throttle and CSO flow from this structure. The
surrogate model and data assimilation scheme are set up only using data from the HiFi model and the
technical drawings of the inspected CSO structure to develop a method that would not need long
historical time series.
2 Methodology

2.1 Surrogate model
The surrogate model used in the current study represents the drainage system as a system of connected virtual tanks. Each virtual tank represents a volume of water in a specific part of the pipe system, hereafter referred to as a ‘compartment’. The outflows from these compartments are calculated using a series of tabulated storage-discharge relationships, which are the main parameters of the surrogate model. These are extracted directly from the HiFi model by propagating a slowly increasing design rainfall through the model. During the surrogate model simulation, the flow at any given point in time is found by interpolating between the nearest tabulated storage-discharge values for the relevant storage volume. Each compartment can have multiple outlets, as can be seen in Figure 1a, for example, for the downstream compartment, which has both an outlet through ‘CSO’ and ‘out downstream’. A more detailed description of the surrogate model and of its calibration can be found in Borup et al. (2017) and Thrysøe et al. (2019). The HiFi model used for estimating the surrogate model is a MIKE URBAN model (DHI, 2016), which has the desired property of being able to produce volume time series for the selected areas of the pipe network. This makes it possible to almost automatically produce the storage-discharge data used as parameters for the surrogate model. The surrogate model can be used for flow forecasting but, due to its computational speed, may also be used in a real-time control setting, including both rule-based control and MPC.

For the current study, the parameters of the surface runoff compartments are calculated from the parameters of the HiFi surface model. The surface runoff in the HiFi model is produced by multiplying the impervious area with the rainfall intensity after subtracting a catchment-specific initial loss. It is routed to the network model using a simple time-area model based on the time of concentration, Tc. The surface runoff in the surrogate model is also based on the total impervious area, which is easily extracted from the HiFi model’s parameter file, and the rainfall intensity – but without considering the initial loss. Instead of using a time-area model for routing the runoff to the network compartments, a single linear reservoir is used where the time constant corresponds to the mean retention time when using the time-area model, which is 0.5Tc.

2.2 Ensemble Kalman filter
In this paper we use the EnKF, which is Monte Carlo implementation of the KF where the uncertainties are represented using an ensemble of models. The EnKF allows for non-linear models and non-Gaussian uncertainty, and is less computationally demanding than the KF for large models (Evensen, 2003).

In the EnKF, an ensemble of model instances, representing the probability distribution of the model states, is propagated forward in time.

\[
\mathbf{x}_{t+1}^{e_i} = M(\mathbf{x}_t^{a_i}, \mathbf{u}_t^{i}, \mathbf{w}_t^{i})
\]
In Eq. 1, \( x_t^{a,i} \) is a vector with the updated states from the previous time step for ensemble member \( i \). These states are propagated forward in time by the model \( M \), applying the input \( u_t^i \) (for example, the measured rainfall intensities) and the system noise, \( w_t^i \), which includes input and model uncertainty. Hereby, the vector with the forecasted states at time \( t+1 \), \( x_{t+1}^{f,i} \), is obtained. The states of the \( N \) individual ensemble members form the model state matrix (Eq. 2).

\[
X_{t+1}^f = \begin{bmatrix} x_{t+1}^{f,1} & x_{t+1}^{f,2} & \ldots & x_{t+1}^{f,N} \end{bmatrix}
\]  

The error covariance, \( P_t^f \), is estimated from the ensemble (Eqs. 3-5)

\[
\bar{x}_{t+1}^f = \frac{1}{N} \sum_{i=1}^{N} x_{t+1}^{f,i}
\]

\[
A_{t+1}^f = \begin{bmatrix} x_{t+1}^{f,1} - \bar{x}_{t+1}^f, x_{t+1}^{f,2} - \bar{x}_{t+1}^f, \ldots, x_{t+1}^{f,N} - \bar{x}_{t+1}^f \end{bmatrix}
\]

\[
P_{t+1}^f = \frac{1}{N-1} A_{t+1}^f A_{t+1}^f T
\]

Whenever observations, \( z \), become available, the state ensemble is updated by

\[
x_{t+1}^{a,i} = x_{t+1}^{f,i} + K_{t+1} (z_t - H(x_{t+1}^{f,i}))
\]

For simplicity, the time index, \( t \), is omitted in Eq. 6 and in the subsequent equations In Eq. 6, \( z_t^i \) is the \( i \)th perturbed observation and \( \nu_t^i \) is the \( i \)th observation noise, which is assumed to have zero mean and covariance matrix \( R \). If observations are uncorrelated, the observation error covariance will become a diagonal matrix. \( H \) is a function relating model states to the observations. In the standard KF implementation, \( H \) is a matrix and thus assumes a linear relation between observations and states. The EnKF allows non-linear relations between states and observations, and is therefore written instead as a vector function \( H \). The term \( z_t - H(x_{t+1}^{f,i}) \) is the deviation between the observations and the modelled equivalents of the observations. \( K \) is the Kalman gain matrix, which weights the trust in the model and observations. A Kalman gain of 0 reflects perfect trust in the model and, in this case, no update will take place. On the other hand, a Kalman gain of 1 will update the model such that it would exactly correspond to the measured value. Neither of these options is realistic, as both the model and the observations are uncertain. The Kalman gain is calculated as

\[
K = P_{t+1}^f H_T (H P_{t+1}^f H_T + R)^{-1}
\]

In the EnKF, \( P_{t+1}^f H_T \) is approximated by the ensemble covariance between the model states and the modelled observations (Eq. 8), while \( H P_{t+1}^f H_T \) can be approximated by the ensemble variance of the modelled observations (Eq. 9).
The EnKF is one of the most frequently used techniques for data assimilation in hydrological applications. The advantages of using the EnKF are: 1) the possibility of using non-linear models (Clark et al., 2008; Liu et al., 2012; McMillan et al., 2013); 2) the model does not have to be formulated in state-space as in some of the other updating schemes (Clark et al., 2008); 3) it is computationally efficient for large models compared to the covariance propagation in the standard KF (Clark et al., 2008; Rakovec et al., 2012); 4) it is robust (Liu et al., 2012); and 5) it fits well into the probabilistic approach related to forecasting schemes (Clark et al., 2008). Even though non-linear models can be applied, the model update itself is linear, which means that the filter is optimal only for strictly linear systems with Gaussian errors. This rarely holds in hydrological models, but the filter can still be successfully applied despite not being optimal (Clark et al., 2008; McMillan et al., 2013).

### 3 Experimental setup

#### 3.1 Case area

The case area is a 1.48-km² urban catchment located in the Damhusåen catchment in the western part of the Greater Copenhagen area, Denmark. This area is also subject to an ongoing study on model predictive control (Lund et al., 2017). The outflow from the downstream CSO structure is influenced by backwater effects at high water levels. Under normal conditions the flow runs through the CSO structure and on to the main collector through a throttle pipe (‘out downstream’, see Figure 1), but during heavy rainfall the capacity of the throttle pipe is superseded, resulting in water backing up into the pipe system. Once the weir level is exceeded in the CSO structure, it eventually causes overflow into the local creek. A simplified schematic of the CSO structure is shown in Figure 1c. A minor fraction of the upstream flow runs to the sewer system in the neighbouring catchment (through ‘out west’, see Figure 1) but otherwise there are no direct interactions with the drainage systems in the surrounding areas.

The case area is one of many comparable catchments from which the sewage can be led to a newly constructed 3.2-km tunnel via centralized control schemes that aim at reducing overflows from...
the system as well as ensuring optimal operation of the downstream wastewater treatment plant. When there is CSO from one of the sub-catchments connected to the tunnel, the water enters the channelized local creek from which it quickly ends up close to a planned bathing area at the coast. This means that a CSO from this catchment can influence the objective of the control of the larger system. It is therefore important to be able to model and forecast both flows and CSOs from these catchments.

3.2 Models

The MIKE URBAN model used as the HiFi model has been provided by the utility company (HOFOR), and the surrogate model has been based on this model. We have subsequently reduced the extent of the HiFi model to include only the part important for the investigated CSO structure. This part consists of the area upstream of the CSO structure as well as a smaller downstream part, which is important in accounting for the backwater effects into the CSO structure (Figure 1a). The reduced HiFi model contains 233 nodes, 97 links, 67 sub-catchments and the CSO structure, and will from here on simply be denoted ‘the HiFi model’.

The surrogate model representing the network part of the HiFi model consists of five compartments of comparable spatial extent (Figure 1). The surrogate model does not cover the full extent of the HiFi model because the surrogate model cannot take backwater effects into account, which is why it is not needed to model the flow downstream of the CSO structure. Each network compartment receives surface inflow from a single runoff compartment (Figure 1b). Figure 2 shows the piecewise-linear storage-discharge curves for each of the compartments. The surrogate model is run with 60-second time steps. This study has shown that the surrogate model is 3500 times faster than the HiFi model when the latter is run with adaptive time steps between 10 and 60 seconds. This speed of the surrogate model is a huge benefit when using the model in combination with EnKF.

The exclusion of dry weather flow (DWF) can lead to non-optimal updating of the model, and a DWF component is therefore added to the downstream flow. This component is constructed using a double sine curve and calibrated on the DWF pattern in the HiFi model.
Figure 1. (a) HiFi model, the approximate division of pipes into compartments and rain gauge (SVK 5710); (b) conceptual drawing of the compartment model; (c) Simplified schematic of the CSO structure - the two yellow shapes mark the level measurement (L) inside the CSO structure and the flow measurement (Q) in the throttle outlet from the CSO structure.
3.3 Input and observations

Time displacement caused by the time it takes for a rain cell to travel from the rain gauge to the catchment can severely reduce the quality of the forecast from updated models (Borup et al., 2013). In this study, rainfall observations come from the rain gauge SVK 5710 (Jørgensen et al., 1998) placed about 1.2 km south-west of the considered sub-catchment. This is sufficiently close to the case area to expect no large impacts of the time displacement. The flow data (Q) comes from an electromagnetic sensor placed inside the throttle pipe leading out from the CSO structure, while the level sensor (L) is an ultrasonic sensor that directly observes the water level in the CSO chamber, Figure 1c. All observations have a frequency of 1 minute.

Figure 3a shows the relationship between the measured flow and level data for seven rain events from August 1 to October 30, 2016. There is a clear, though not unambiguous, relationship. Figure 3b shows the level and flow modelled by the HiFi model for the same seven rain events. It is clear that the variability in the measurements is larger in reality than in the model, and that the bottom levels and crest levels are different; however, the overall shape fits well.
3.4 Observations used for updating

The data assimilation is carried out for a period of approximately one month from June 16 to July 12, 2016, during which several overflows (water level above crest level) occurred, and during which the data was visually assessed as being trustworthy, see Figure 4. Missing and erroneous data in the flow and level measurements have been replaced with the last occurring flow and level, respectively.

Both the flow and level observations are used for updating in this study. The compartment model uses volume as the internal state and describes the flow out of compartments, meaning that the model can be updated either with the flow out of the downstream compartment or via the volume in the same compartment. We examine four experimental setups related to the handling of the observation in the EnKF scheme:

1) EnKF-Q Updating with the flow measurement directly.
2) EnKF-QV Converting the flow measurement to volume, and using this volume for updating.
3) EnKF-LV Converting the level measurement to volume, and using this volume for updating.
4) EnKF-LQ Converting the level measurement to flow, and using this flow for updating.
A conversion between the measured quantity and the quantity used for updating is done using rating curves for the experimental setups EnKF-QV, EnKF-LV and EnKF-LQ. These curves are created using the HiFi model, which is run from August 1 to October 30 (outside the time period later used for data assimilation). We are only interested in rainy situations and therefore choose seven rain events in this period (also used to create Figure 3). The levels of the CSO structure in the HiFi model differ from the actual levels of the CSO structure on the technical drawing (the level of the bottom is 21 cm lower in the HiFi model than on the drawing and the crest level is situated 4 cm lower), as seen in Figure 3. We use this information to correct the levels in the HiFi model before establishing the two level-dependent rating curves (Figure 5). In Figure 5, the grey circles indicate the results from the HiFi model (after correction of the levels). The quantity being converted (x axis) is divided into 14 equally large bins and a straight line is fitted to the data within each bin. Subsequently, the centre of each line (one for each bin, the red circles) are connected with the minimum and maximum values, creating 15 regression line segments that together constitute the rating curve (red line).

In KF, the measurement operator, $H$, would normally be used to map between states and observations. We choose to do the conversion outside the filter implementation in order to use the rating curves to estimate the observation error. Thus, for EnKF-Q and EnKF-LQ, the measurement operator, $H$, which converts the state to the modelled measurement, is the piecewise-linear compartment model that converts the downstream network compartment state to the throttle flow out of the CSO structure. For EnKF-QV and EnKF-LV, the measurement operator, $H$, is a vector of zeros.

![Figure 5. Rating curves used to convert level and flow measurements in the downstream compartment.](image)
with a single value of one that picks the volume for the downstream compartment from the state vector.

### 3.5 Implementation of the Ensemble Kalman filter

The data assimilation is performed in MATLAB by connecting the surrogate model with the EnKF-MATLAB toolbox developed by Sakov (2013). The state vector for the updating consists solely of the volumes in the five network compartments, since a preliminary study showed that additional updating of the volume in the runoff compartments did not improve the results (Lund et al., 2018b).

Choosing the ensemble size is a trade-off between the performance of the updating and the computational cost. We performed initial tests with ensemble sizes of 10, 20, 50, 100, 200 and 500 members. Using 10 and 20 members provided poor results while 50-100 members and above provided similar results in terms of the squared error on forecast flows. To avoid negative impacts on the performance due to an insufficient ensemble size, we choose an ensemble size of 100. This is still feasible in terms of computational cost due to the efficiency of the surrogate model.

In EnKF applications, all uncertainties have to be represented as realizations of random noise. This means that model forcing, observations and state variables will be perturbed according to the specified noise properties. The specification of these can have a big impact on the filter performance (Zhang et al., 2015). When long historical observations are available, the various noises can to some extent be estimated or optimized from the data, but in the current setup we constrain ourselves either to derive the noise terms from the HiFi model or to simply estimate a range of reasonable noise terms. We assume that the uncertainty mainly originates from the rainfall observations as well as the runoff and network parts of the model. There is also uncertainty related to the observations. Therefore, we perturb the rainfall input, the model compartment states and the observations. The noise formulation for the observations differs, depending on the experimental setup. Table 1 gives an overview of the choices made for the noise formulations, while the subsequent sections describe each of the three error sources in detail.

#### Table 1: Noise formulations for rainfall, states and observations.

<table>
<thead>
<tr>
<th>Perturbation</th>
<th>Rainfall</th>
<th>States</th>
<th>Observations in each experimental setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>EnKF-Q</td>
<td>$q_{\text{env}}(t) = q_{\text{env}}(t) + \epsilon q(t)$, $\epsilon q(t) \sim \mathcal{N}(0, \sigma_q^2)$, $\sigma_q = 0.1$; $q_{\text{obs}}$=observed flow.</td>
<td>$v_{\text{env}}(t) = v_{\text{mod}}(t) + \epsilon v(t)$, $\epsilon v(t) \sim \mathcal{N}(0, \sigma_v^2)$, $\sigma_v = \sqrt{\alpha_{v,\text{netw}}}, \alpha_{v,\text{runof}} \in [0.05, 0.1]$; $v_{\text{mod}}$=modelled volume.</td>
<td>- EnKF-Q: $Q_{\text{env}}(t) = Q_{\text{env}}(t) + \epsilon Q(t)$, $\epsilon Q(t) \sim \mathcal{N}(0, \beta \sigma_Q^2)$, $\beta = 1$; $Q_{\text{obs}}$=from HiFi model results.</td>
</tr>
<tr>
<td>EnKF-QV</td>
<td>$v_{\text{QV}}(t) = v_{\text{mod}}(t) + \epsilon v(t)$, $\epsilon v(t) \sim \mathcal{N}(0, \beta \sigma_v^2)$, $\beta = 1$; $\sigma_v$=from HiFi model results.</td>
<td>$\epsilon v(t) \sim \mathcal{N}(0, \beta \sigma_v^2)$, $\beta = 1$; $\sigma_v$=from HiFi model results.</td>
<td>- EnKF-LQ: $Q_{\text{QV}}(t) = Q_{\text{QV}}(t) + \epsilon Q(t)$, $\epsilon Q(t) \sim \mathcal{N}(0, \beta \sigma_Q^2)$, $\beta = 1$; $\sigma_Q$=from HiFi model results.</td>
</tr>
<tr>
<td>EnKF-LQ</td>
<td>$v_{\text{LV}}(t) = v_{\text{mod}}(t) + \epsilon v(t)$, $\epsilon v(t) \sim \mathcal{N}(0, \beta \sigma_v^2)$, $\beta = 1$; $\sigma_v$=from HiFi model results.</td>
<td>$\epsilon v(t) \sim \mathcal{N}(0, \beta \sigma_v^2)$, $\beta = 1$; $\sigma_v$=from HiFi model results.</td>
<td>- EnKF-LQ: $Q_{\text{QV}}(t) = Q_{\text{QV}}(t) + \epsilon Q(t)$, $\epsilon Q(t) \sim \mathcal{N}(0, \beta \sigma_Q^2)$, $\beta = 1$; $\sigma_Q$=from HiFi model results.</td>
</tr>
</tbody>
</table>
3.5.1 Perturbation of rainfall data

The measured rainfall is perturbed by multiplying it with a random factor $\epsilon_i$ that is drawn from a uniform distribution between 0.5 and 1.5 every $\Delta t$ minutes, where $\Delta t$ itself is a random variable that is drawn from a uniform distribution between 1 and 10 minutes at the same time as $\epsilon_i$. Hereby, we obtain the rainfall ensemble, $I_{\text{ens}}$. The limits of the uniform distribution are chosen based on a sensitivity study carried out by Lund et al. (2018b).

$$I_{\text{ens}}(t) = I_{\text{obs}}(t) \cdot \epsilon_i(\Delta t), \quad \epsilon_i(\Delta t) \sim U[0.5,1.5], \quad \Delta t \sim U[1,10]$$  \hspace{1cm} (11)

Temporal correlation in the error is important, as rapid fluctuations around the mean can otherwise cancel themselves out as the rainfall errors are aggregated in the model, which can lead to insufficient ensemble spread. The temporal persistence in the rainfall observation error can be explained by the nature of a tipping bucket rain gauge that integrates the rain since the last tip, and by the fact that rainfall observations are affected by sudden changes in wind direction, etc. This way of making ensemble representation of the rainfall error has previously been used in an urban runoff data assimilation study (Borup et al., 2014). The method makes it possible to obtain a larger spread in the resulting runoff ensemble without having negative factors or affecting the ensemble mean, compared to using, for example, a Gaussian random walk, which can be a desirable property in data assimilation studies.

3.5.2 Perturbation of states

We perturb the states (volumes) of the compartments as

$$V_{\text{ens}}(t) = V_{\text{mod}}(t) \cdot \epsilon_V(t), \quad \epsilon_V(t) \sim N\left(1, \left(\alpha_V V_{\text{mod}}\right)^2\right)$$  \hspace{1cm} (12)

where $V_{\text{mod}}$ is the modelled volume in the runoff compartment at time $t$. The proportional noise term, $\epsilon_V$, is normally distributed with a standard deviation of $\alpha_V V_{\text{mod}}$, with $\alpha_V = [\alpha_{V,\text{netw}}, \alpha_{V,\text{runoff}}]$. In this study, values of $\alpha_{V,\text{netw}}$ for network compartments and $\alpha_{V,\text{runoff}}$ for runoff compartments equal 0.05 and 0.1, respectively. These are used as default estimates of the state noise. Values of 0 and 0.2 are also applied (only noise on runoff compartments) to explore how sensitive the model is to the state perturbation.

To make the updating more robust and limit the dispersion of the ensemble within a reasonable range, we use the upper volume limit of the storage-discharge curves as the maximum possible compartment state value. If an ensemble member has a volume larger than this value, its volume is set to be equal to the maximum value. This can potentially lead to an ensemble collapse if all the ensemble members are set to the maximum value; however, this potential problem is mitigated by the fact that the ensemble will be dispersed again when water flows out of the compartment and new noisy input is added at the next time step.
3.5.3 Perturbation of observations

For EnKF-Q, we perturb the measured flow, $Q_{\text{obs}}$, by assuming a normally distributed, proportional error, $\varepsilon_Q$, with a standard deviation of $\alpha Q_{\text{obs}}$. The standard deviation is proportional to the flow, meaning that the larger flow will also have the larger error. In this paper, $\alpha$ is set to 0.1 (default) based on Bertrand-Krajewski et al. (2003). However, values of 0.05 and 0.2 are also tested to assess the sensitivity of this parameter.

$$Q_{\text{obs}}(t) = Q_{\text{obs}}(t) \cdot \varepsilon_Q(t), \quad \varepsilon_Q(t) \sim N\left(1, (\alpha Q_{\text{obs}})^2\right)$$  \hspace{1cm} (13)

For EnKF-QV, EnKF-LV and EnKF-LQ, there are errors arising both from the measurement itself and from the conversion via the rating curve. We assume that the error in the measurement per se is much smaller than the one occurring from using the rating curve and therefore we are neglecting the measuring error. The data from the HiFi model simulation shown in Figure 5 is used to establish the observation noise description for the EnKF. The standard deviation is determined for the data around the rating curve for each of the 15 line segments (see the 95% confidence level in Figure 5 – dashed black line). Proportional noise terms, as used in EnKF-Q, are applied to indicate that the noise becomes larger for higher observed values. The change in noise depending on the measured value is, however, already embedded in the noise formulation obtained from the HiFi model (visible as narrow or broad confidence intervals in Figure 5); thus, we use additive noise terms in EnKF-QV, EnKF-LV and EnKF-LQ:

$$V_{QV}(t) = V_{QV}(t) + \varepsilon_{QV}(t), \quad \varepsilon_{QV}(t) \sim N\left(0, (\beta \sigma_Q)^2\right)$$

$$V_{LV}(t) = V_{LV}(t) + \varepsilon_{LV}(t), \quad \varepsilon_{LV}(t) \sim N\left(0, (\beta \sigma_V)^2\right)$$ \hspace{1cm} (14)

$$Q_{LQ}(t) = Q_{LQ}(t) + \varepsilon_{LQ}(t), \quad \varepsilon_{LQ}(t) \sim N\left(0, (\beta \sigma_Q)^2\right)$$

When using the noise estimation directly from the HiFi model, $\beta$ is set to 1 (referred to as default settings). In order to explore the sensitivity of the measurement error, we scale the standard deviation by changing $\beta$ to 0.001, 0.01 and 0.1. Values of 10 and 100 have also been applied but these gave extremely poor results, since the standard deviations extracted from Figure 5 are already quite large, especially for the conversions into volumes. A larger standard deviation may therefore result in negative volumes, which the model converts to zeros, thus resulting in a biased ensemble. These results are therefore not shown.

3.6 Evaluation measures

The objective of the study is to establish whether, and by how much, we can improve the flow and overflow forecast performance of a surrogate model by using EnKF. We use flow forecasts up to 180 minutes into the future to quantify the performance, both to give an indication of the quality of the update of the internal states and to examine the potential of using data assimilation in a forecast
setting. In this study, we assume that the forecast rainfall equals the measured one. This, however, still includes some uncertainty due to representation error of the rain gauge. We apply deterministic forecasts, meaning that we use the mean of the updated states as initial conditions for a single forecast model run. We make deterministic forecasts because this is what many model-based control schemes, such as MPC, can currently manage. The four experimental setups are compared with the surrogate model baseline (the surrogate model without data assimilation) and with the HiFi model without data assimilation. We quantify the performance of throttle flow forecasting using the mean absolute error (MAE), Nash-Sutcliffe efficiency (NSE) and persistence index (PI). Accurate forecasts are most important immediately before and during rain events and we therefore quantify the performance of the throttle flow only for time periods with an observed flow higher than 0.05 m³/s. The overflow forecasting skill is evaluated using the true positive rate (TPR) and positive predictive value (PPV). These evaluation measures are presented in the next sections.

3.6.1 Mean absolute error

The MAE is calculated as

\[
\text{MAE} = \frac{\sum |Y_{\text{model}}(t) - Y_{\text{obs}}(t)|}{n}
\]

where \(Y_{\text{model}}(t)\) is the modelled value and \(Y_{\text{obs}}(t)\) is the observed value at time \(t\) while \(n\) is the number of time steps.

3.6.2 Nash Sutcliffe Efficiency

The NSE is calculated as

\[
\text{NSE} = 1 - \frac{\sum (Y_{\text{model}}(t) - Y_{\text{obs}}(t))^2}{\sum (Y_{\text{model}}(t) - \overline{Y_{\text{obs}}})^2}
\]

where \(\overline{Y_{\text{obs}}}\) is the mean value of the observations. NSE is a performance measure, ranging between 0 and 1. Due to the squaring of the terms, it especially punishes large deviations from the model to the observation. An NSE of 1 means a perfect fit between the model and the observation.

3.6.3 Persistence index

The persistence index (PI) is calculated as

\[
\text{PI} = 1 - \frac{\sum (Y_{\text{model}}(t) - Y_{\text{obs}}(t))^2}{\sum (Y_{\text{obs}}(t) - Y_{\text{obs}}(t-j))^2}
\]

where \(j\) represents the horizon of the prediction. \(Y_{\text{obs}}(t-j)\) is thus the observation at the time of the prediction. PI compares the performance of the model forecast with the performance of the last known observation (Kitanidis and Bras, 1980), often denoted as persistent forecast. If \(\text{PI} < 0\), the persistent forecast (last known observation) is better than the model forecast, and vice versa if \(\text{PI} > 0\).
We use the true positive rate (TPR) and positive predictive value (PPV) to determine how well the data assimilated model forecasts the observed CSOs. There is no flow meter installed at the CSO outlet and we can therefore only base the CSO evaluation on the measured water level in the overflow structure. Observations above crest level are categorized as ‘CSO’ and measurements below crest level are categorized as ‘no CSO’. The CSO predicted by the surrogate model (updated and baseline) and by the HiFi model are also categorized as CSO/no CSO. This information is used to make a contingency table (Figure 6).

The true positive rate (TPR) and positive predictive value (PPV) can be calculated based on the contingency table (Saito and Rehmsmeier, 2015):

\[
TPR = \frac{TP}{TP + FN}
\]

\[
PPV = \frac{TP}{TP + FP}
\]

The TPR shows how many of the observed CSOs are captured by the model prediction. PPV contrarily indicates how much you can trust that a predicted CSO will actually occur. Both TPR and PPV have 0 as the worst outcome and 1 as a perfect score. The two measures will in some cases be counteractive. If the model too often predicts that a CSO will occur it will likely capture all observed CSOs (high TPR) but at the same time the prediction cannot be trusted very much (low PPV). This implies one should look at both measures when evaluating forecast performance.

### 4 Results and discussion

#### 4.1 Evaluation of throttle and CSO flow forecast performance

Figure 7 shows the five evaluation measures for each of the four experimental setups (EnKF-Q, EnKF-QV, EnKF-LV and EnKF-LQ). Figure 8 shows the volume of water subtracted or added to each compartment for each time step for the default setups (Table 1). All schemes updated all compartments to some extent. The largest update happened in the downstream compartment for all four setups. For all four default setups there was almost no forecast skill left after 120 minutes, and all updates were “washed out” after 180 minutes. Of the four default setups, EnKF-LV performed the
best in all five evaluation measures for forecast lengths from 15 minutes and above, except for PPV where EnKF-Q and EnKF-LQ were superior. If the goal is to predict CSOs, EnKF-LQ should not be used due to the poor TPR. Also, this experimental setup excelled neither in MAE nor in NSE. The EnKF-LQ setup is seen to have the largest update, but gave the worst forecast skill. Constraining the update may thus increase the performance of this experimental setup.

By examining PI in Figure 7 it is seen that from 15 minutes and above it was better to use both the updated surrogate model and baseline surrogate model for forecasting than using the last observation. This was not the case for the one-minute prediction because the flow generally does not vary much from one time step to the next. The PI for the HiFi model shows that it would be better to use the last measurement for forecasting up to 60 minutes ahead, which indicates that the HiFi model has not been properly calibrated for this upstream area. This poor performance of the HiFi model is also seen with respect to MAE, NSE and TPR. The surrogate model was built on the basis of the HiFi model and we believe that this increase in performance must stem from a coincidental favourable simplification of inaccurate descriptions in the HiFi model.

The noise specifications of the filter were changed to evaluate the sensitivity of each experimental setup. The worst and best performances for each experimental setup (denoted ‘lower and upper bounds’) are shown in Figure 7. The values used to make these upper and lower bounds do not necessarily stem from one specific combination of noise formulations for all forecast horizons. From here it is clear that EnKF-QV was very sensitive to the choice of noise parameters, as the upper and lower bounds are far apart. Contrarily, the EnKF-LV bounds are very close together, meaning that this scheme was insensitive to the noise description. This is an advantage in an operational setting, where the time and data are often not available for fine-tuning of the data assimilation. It can also be seen that the results based on the default values were often very close to the upper bounds, meaning that tuning of the noise formulation had very little added benefit.
Figure 7. Performance comparison between the four experimental setups. MAE, NSE, PPV and TPR takes values at 0, 1, 15, 30, 120 and 180 minutes, whereas PI takes values at 1 minute and above. The default noise formulation represents the noises estimated or calculated from the HiFi model, whereas the upper and lower bound is the worst and best performance for each experimental setup in the sensitivity study. The axes have been scaled for better visualization.
Figure 8. The updated volume in each time step for the four experimental setups. Some values are removed by scaling the y axis to make a better visualization.

4.2 Throttle and CSO flow forecasts using EnKF-LV for selected events

Two different events are used to visualize the performance of the update for the best performing (default) experimental setup for both the throttle flow predictions (Figure 9a) and the CSO flow predictions (Figure 9b). The use of five different evaluation measures on six different forecast horizons (1-180 minutes) makes it difficult to give an unambiguous conclusion about which setup is the best, since this will depend on the purpose of the modelling. Do we want to be good overall at capturing the flow (low MAE), are we interested in the peaks (high NSE), do we want to capture the observed CSO events (high TPR), or do we want to be able to have a high confidence in our CSO prediction (high PPV)? Based on an overall assessment, EnKF-LV was chosen as the setup with the best performance.

For the 15-minute forecast, the updated surrogate model captured both the peaks and lows of the throttle flow really well compared to the surrogate baseline and HiFi models. We also see that the
baseline surrogate and HiFi models in general predicted that the CSO would happen later than it actually did based on the measurements (visualized as CSO/no CSO). The updated surrogate model managed to predict the CSO earlier but was nevertheless still delayed compared to the measurement. This was a general tendency for the entire time series. On four occasions, the updated surrogate model predicted a CSO that was not captured by the baseline surrogate model; one of these is shown in Figure 9b. Neither did the HiFi model predict the CSO in three of these four instances. Only on one occasion did the updated surrogate model completely miss a CSO event.

When looking at the 60-minute forecasts, it is clear that an increased forecast horizon decreased the performance significantly. There was still some forecast skill left for the throttle flow prediction, but the CSO prediction was almost identical for the updated and baseline surrogate models.

The updated surrogate model had a too high DWF compared to the measurements. A potential explanation is that it is difficult to obtain the “true” level-flow relationship in this range, since the uncertainty on the sensors increases for small amounts of water; furthermore, the level-flow relationship can be highly influenced by, for example, sedimentation.
4.3 Limitations of the study and future perspectives

The HiFi model performed worse than the baseline surrogate model. To be able to make a fairer comparison between the HiFi model and the updated surrogate model, the HiFi model could be calibrated. Many filter specifications have to be defined when setting up the EnKF, and this paper investigated the sensitivity of some of these. Other filter specifications, which were not tested in this paper, include the length of the temporal correlation in the rainfall perturbation; error distribution types; the choice between additive or proportional error terms; heuristically estimated noise descriptions for the observations; and the subdivision of the rating curves. Other data assimilation filters than the EnKF could also have been used in the study. Due to the small number of elements in the state vector of the surrogate model, even non-ensemble-based methods, such as the Extended Kalman filter, might have worked well. The limited size of the HiFi model implies that we could also have updated this with the EnKF within the given time constraints. This would, however, have entailed numerous other challenges, such as stability issues and how to specify the noise for the many hydraulic variables. We do foresee that typical applications of the surrogate models as forecast models will be for larger systems that would prohibit the use of a HiFi model. The state vector of the surrogate model will here have several hundred elements and the updating will be based on sensor data from multiple locations, in which case the EnKF is computationally very efficient. The main challenge will in this case be the specification of the temporal and spatial errors of the rainfall estimates, since this will determine the correlations throughout the model ensemble that the EnKF uses for the updating. Specified well, the correlations in the error of the rainfall estimates should make it possible to improve the forecasting also in ungauged parts of the model using data from nearby catchments. Nevertheless, it is to be expected that a localization scheme needs to be introduced to dampen long-range corrections arising from, for example, misspecification of the rainfall error.
Furthermore, the effect of a distribution of DWF to the network compartments and the use of distributed rainfall on the updating performance was not investigated due to the limited spatial size of the catchment area.

5 Conclusions

The low computational speed of HiFi models impedes their use for real-time forecasting in urban drainage systems. Thus, the HiFi model does not pose an alternative to using a surrogate model of a HiFi model for forecasting. Contrarily, conceptual and data-driven models, dependent on data for construction and/or calibration, may be used. However, in the usual situation, where only level data is available, such models will be difficult to construct to produce flow forecasts, and surrogate models are in these cases the only viable solution. Their simplified system description, however, prompts for carrying out data assimilation before forecasting. We therefore constructed a forecast scheme by combining a surrogate model of a distributed urban drainage model with the Ensemble Kalman filter, without relying on historical time series data for model construction and/or calibration. We tested the approach on a small catchment in Copenhagen, Denmark. The model updating was performed with real data by assimilating either water level or flow observations. The predictive skill of the updated surrogate model, the baseline surrogate model and the HiFi model was established by comparing the forecast flows and overflows from a downstream CSO structure to real flow and level observations.

The results showed that the HiFi model would benefit from calibration to the observations used in this study. Updating the surrogate model provides significantly better results than both the baseline surrogate model and the HiFi model. This was valid for forecasts up to two hours ahead when evaluating the models’ ability to forecast the flow through the CSO structure’s throttle pipe and the CSO discharge, showing that updating can counteract inaccuracies stemming from the HiFi model. Furthermore, updating using water level observations resulted in better flow forecasts than basing the assimilation on flow data – even though the forecasts were validated against flow data from the same flow sensor. Level sensors are the predominant sensor type in urban drainage systems, since these are cheaper and more reliable than flow sensors. It is therefore highly beneficial that the use of level measurements improved the forecast skill more than flow measurements. Another benefit of using water level data for the updating was that the forecast skills were rather insensitive to the noise formulations in the filter, which means that the experimental setup is robust and requires less tuning. This is especially valuable for operational implementations where there may be limited time and data for fine-tuning of the model.

This paper shows that updating of surrogate models may constitute a robust and efficient way of creating both flow and overflow forecasts. The robustness of the method and the use of cheap level sensors are important for the operationalization of data assimilation. Our hope is that this, along with more research on the topic, will make it the norm rather than the exception to update operational, online urban drainage models with measurements in real time.
6 Acknowledgements

We would like to thank Francesco Righetti, who constructed the surrogate model for the case area, Peter Steen Mikkelsen and Ole Mark for contributing to the initial concept of the paper and useful discussions, and HOFOR for providing us with the HiFi model and sensor data. This project was supported by Innovation Fund Denmark through the Water Smart Cities Project.

7 Declarations of interest

None.

8 References


Refsgaard, J.C., 1997. Validation and intercomparison of different updating procedures for real-time
Saito, T., Rehmsmeier, M., 2015. The Precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10.


