Data assimilation in hydrodynamic models for system-wide soft sensing and sensor validation for urban drainage tunnels

Rocco Palmitessa, Peter Steen Mikkelsen, Adrian W. K. Law and Morten Borup

ABSTRACT

Tunnels are increasingly used worldwide to expand the capacity of urban drainage systems, but they are difficult to monitor with sensors alone. This study enables soft sensing of urban drainage tunnels by assimilating water level observations into an ensemble of hydrodynamic models. Ensemble-based data assimilation is suitable for non-linear models and provides useful uncertainty estimates. To limit the computational cost, our proposed scheme restricts the assimilation and ensemble implementation to the tunnel and represents the surrounding drainage system deterministically. We applied the scheme to a combined sewer overflow tunnel in Copenhagen, Denmark, with two sensors 3.4 km apart. The downstream observations were assimilated, while those upstream were used for validation. The scheme was tuned using a high-intensity event and validated with a low-intensity one. In a third event, the scheme was able to provide soft sensing as well as identify errors in the upstream sensor with high confidence.

Key words | combined overflow tunnel, data assimilation, distributed urban drainage models, model-based sensor validation, soft sensing, urban drainage tunnel

HIGHLIGHTS

● We propose a data assimilation scheme tailor-made for urban drainage tunnels that can efficiently assimilate observations into an ensemble of 1D hydrodynamic models.
● We tested and validated our methodology with a real case study.
● The results support our hypothesis that the scheme is capable of promoting the hydrodynamic model to a soft sensing tool, which can be further used for validating physical sensors.

INTRODUCTION

Tunnels are increasingly integrated in urban drainage systems (UDSs). Notable examples include the pioneering Tunnel and Reservoir Plan in Chicago (USA), the Thames Tideway Tunnel in London (UK) for overflow interception and the Deep Tunnel Sewerage System in Singapore for the centralized collection of used water. Despite the variety of uses and designs, their defining feature is the ability to expand the capacity of UDSs while conveying water to final treatment and disposal/re-use (Palmitessa et al. 2018). The depth and scale of urban drainage tunnels also allow UDSs to minimize disruption at the surface level and achieve attractive cost-benefit ratios.

To take full advantage of their capacity, tunnels need to be optimally controlled. A control system includes...
monitoring equipment to measure the current state of the system, actuators to change the dynamics of the system and a control strategy aimed at reaching defined objectives. It is essential to have as much knowledge of the state of the system as possible to take the right control decisions. The monitoring and control of tunnel systems primarily relies on sensor data. Water level and flow sensors measure the hydraulic state in strategically chosen locations. Additional information can be indirectly derived knowing the state of the control actuators, e.g. if gates are open or pumps are activated. However, the sensor data is limited by the instrument accuracy and has mostly a point-wise validity. Models can augment the monitoring of drainage tunnels in real time by providing a system-wide estimate of the hydraulic variables as compared to the point-wise observations from the sensors. The model estimates could also ideally be used for validating observations.

Models incorporate our knowledge of the governing hydrological and hydraulic processes for the UDSs. The degree of detail and accuracy varies widely from conceptual models to Computation Fluid Dynamics (CFD) models, and each serves specific purposes. Distributed Urban Drainage Models (DUDMs) built in state-of-the-art software such as Mike Urban, SWMM and InfoWorks, balance a sufficient level of detail with a feasible computational cost and are commonly used for both design and optimization of UDSs. DUDMs rely on a large number of parameters, related to geometry, materials, etc., to be able to simulate the complex and non-linear behaviour of UDSs. Many of these parameters can be derived from the asset information, e.g. diameter and length of the pipes, others need to be calibrated. Deletic et al. (2012) identify the calibration process as a key source of uncertainty in urban drainage modelling together with the model input, e.g. rainfall data and dry weather inflows, and the model structure. Moreover, the calibration is sensitive to the quantity and quality of available measurements (Tscheikner-Gratl et al. 2016), which are typically limited in UDSs.

Data assimilation (DA) is capable of improving the accuracy of model predictions by correcting the model states and/or parameters dynamically based on information from the sensors, thus preventing the model from drifting far from reality (Hutton et al. 2021). The correction or ‘update’ can be either restricted to the locations where the observations are available, or propagated in space depending on the physical relationship among the states. DA algorithms can be broadly classified as deterministic, which correct the model variables based on some prior assumptions on their uncertainty, and statistical, which dynamically assess and compare the uncertainty of model predictions and observations, and update the model with a weighted average of the two, with more weight being given to the less uncertain (Borup 2014). In ensemble-based methods, the uncertainty of model predictions is computed by running an ensemble of the same model, where each ensemble member is forced with a differently perturbed input (e.g. initial conditions, boundary conditions and model parameters).

The use of DA in rural hydrologic forecasting is well-established (Liu et al. 2012) but has yet to reach the same level of adoption in the urban context. Borup et al. (2014) suggest that only recently the need for controlling UDSs in real time and the availability of communication and computing resources have prompted the application of DA in urban drainage modelling. Previous studies have explored the challenges and benefits of both deterministic and statistical approaches. Hansen et al. (2014) updated a DUDM by inserting or extracting water at every computational time step at the location with available observations. Direct correction requires limited computational resources but does not provide a measure of uncertainty and can only propagate the update downstream. Borup et al. (2011) assimilated downstream flow measurements to indirectly correct the (unknown) infiltrating inflow upstream, thus propagating the update across the entire system. Similarly, Hutton et al. (2014) used the downstream residual errors to update the upstream states. While the deterministic approaches have proven to perform well with gradually changing hydraulic conditions, they cannot produce uncertainty estimates and cannot easily assimilate ambiguous information from multiple sensors. For these reasons, ensemble-based methods are to be preferred over the deterministic methods if the required computational resources are available (Borup et al. 2014). Lund et al. (2019) limited the computational cost of running large ensembles by replacing the DUDM with a lumped surrogate model, at the expense of the spatial resolution of the model estimates.

The quantity and quality of sensor data in UDSs is not always reliable enough for monitoring purposes. This is
even more so in large drainage tunnels where sensors are difficult to maintain, due to their location deep underground in a dangerous and harsh environment. If a detailed hydrodynamic model correctly updated in real time is capable of producing reliable, probabilistic estimates throughout the entire system, it can serve as a soft (software) sensing tool to complement the hard (hardware) physical sensors. Independent observations can then be automatically validated against the model estimate, given our degree of confidence in the sensor and the updated model.

The aim of this study is to enable soft sensing and sensor validation of urban drainage tunnels by developing and validating a DA scheme tailor-made for tunnel models that can efficiently assimilate observations into an ensemble of models. Our proposed scheme restricts the assimilation only to a subsystem of the UDS, which includes the tunnel and its connected structures. By doing so, we significantly decrease the computational cost of the assimilation, while retaining the spatial resolution and level of detail of the fully distributed model.

**METHODOLOGY**

**Assimilation scheme**

The application of ensemble-based DA to DUDMs requires running in parallel several instances of the same model. To limit the computational overhead, we develop a scheme that isolates an ‘inner’ model of the tunnel from the ‘outer’ DUDM, where the output from the outer model is used as the boundary condition for the inner model, and it is only this model that is updated. The inner model retains the same level of detail as the outer model and includes the tunnel itself and its inlet and outlet structures (e.g. overflows, dropshafts and pumping stations).

Setting the inner model boundaries is a critical step in the implementation of the scheme. It is a case-specific procedure but should satisfy some general criteria:

(i) the inner model should include the smallest number of structures besides the tunnel itself, to minimize the computational cost;
(ii) the inner model should include all locations with sensors whose data is relevant for the assimilation;
(iii) at the tunnel inlets, the boundaries should be set at locations where no backwater effects from the tunnel are expected, e.g. overflows;
(iv) the outlet should include only the structures necessary to simulate the emptying of the tunnel, e.g. pumping stations and safety overflows.

To optimize for the first and second criteria, the boundaries are ideally set at locations where the sensors are placed. In principle, the boundary conditions could be either water levels or discharges. However, level gauges are preferred because they are usually more common and accurate than flow meters.

The scheme is implemented in three stages, as seen in Figure 1. In an offline setting, the stages are sequential, so that one stage is completed for the entire simulation period before moving to the next. In an online scenario, the sequence should be repeated at each time step. In the first stage, a single instance of the drainage model receives as input rainfall observations. From its output, the simulated states at the tunnel boundaries are read. The boundary conditions are perturbed according to a user-defined perturbation model in the second stage, to account for the uncertainty of the drainage model. In the final stage, each set of perturbed boundary conditions is linked to an identical copy of the tunnel model. Therefore, each member of the tunnel model ensemble returns a different prediction of the tunnel states, to represent the uncertainty of the model. Given the spread of the predicted tunnel states and the uncertainty assigned to the corresponding observations, the DA algorithm computes the updated states. The process is repeated at each time step and the updated states serve as the initial condition for the prediction at the following time step. If the observations are missing at a given point in time, the assimilation step is bypassed, and the predicted tunnel states are used directly as the initial condition for the following prediction.

**1D Hydrodynamic models**

DUDMs in state-of-the-art software such as Mike Urban (DHI 2019) include a lumped-conceptual surface module, that converts precipitation data into stormwater inflows, and a hydrodynamic module, that simulates the storage and conveyance of water in the system. 1D hydrodynamic models
compute the change over time and space of the hydraulic variables by solving the continuity and momentum equations, also known as the St. Venant equations. The transition between the free-surface flow and pressurized flow is numerically dealt with by introducing a narrow virtual opening with insignificant volume longitudinally at the top of all pipes, called the Preissmann slot, which makes it possible to use the St. Venant equations, intended for free-surface conditions, for pressurized flows as well. Water levels $h$ and discharges $Q$ are computed at pre-defined time and space intervals. The computational grid is generated automatically by the software, or with a user-specified number of grid points (Figure 2).

**Ensemble-based data assimilation**

The Kalman Filter is a sequential assimilation method, which means that the model propagates in time its internal variables or state vector $x$ and is updated whenever measurements $d$ are available. At time $k$, the state vector $x_k$ is given by the model $M$ as:

$$x_k = M(x_{k-1}, u_{k-1}, e_{k-1})$$

where $u$ is the model input and $e$ is the system noise, which represents the uncertainty of the model forecast. For hydrodynamic models, the vector $x$ usually includes both water levels and discharges, i.e. all the $h$ and $Q$ variables shown in Figure 2.

If one or more observations are available, the analysis or updated states $x^a$ are computed as a weighted average of the model forecast or background states $x^f$ and the observations using the analysis equation:

$$x^a = x^f + K(d - Hx^f)$$

where $H$ is a matrix mapping the model states to the observations and $K$ is the Kalman gain matrix. The elements of $K$ act as weights between the vectors $x^f$ and $d$. The higher the value of the gain, the closer the state is updated to the observation. The Kalman gain is calculated as:

$$K = P^f H^T (HP^f H^T + R)^{-1}$$

where $P^f$ is the forecast error covariance matrix and $R$ is the observation error covariance matrix. If $n$ is the size of the state vector, then $P$ has the dimensions $n^2$. The Kalman Filter requires $P^f$ to be propagated in time analytically, which is too costly an operation for large models, and is only valid for linear models.

To overcome these two limitations, Evensen (1994) introduced the Ensemble Kalman Filter (EnKF), which uses an ensemble of model instances to propagate in time the error statistics. In the EnKF, the error covariance
matrix is replaced by the ensemble covariance, which is calculated as:

\[
P = \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{X})(x_i - \bar{X})^T = \frac{1}{m-1} AAT
\]  

(4)

where \( X = [x_1, \ldots, x_m] \) is the ensemble of model states, \( m \) is the ensemble size, \( \bar{X} \) is the ensemble mean, and \( A = [A_1, \ldots, A_m] \) is the ensemble of anomalies \( A_i = (x_i - \bar{X}). \) Therefore, the dimensions of \( P \) are here reduced to \( m \times n. \) The gain computed with the error covariance is then used to update the ensemble members individually. To prevent the ensemble spread reducing too rapidly, the EnKF requires the use of perturbed observations, which means that each member is associated with a different realization of the observation noise consistent with the observation error covariance. However, this introduces sampling error, especially for small ensembles.

We choose to use the Deterministic Ensemble Kalman Filter (DEnKF), which eliminates the need for perturbing the observations (Sakov & Oke 2008). It does so by (i) applying the analysis to the mean ensemble forecast \( \bar{X} \) instead of each ensemble member, (ii) updating the ensemble anomalies using only half the Kalman gain:

\[
A^a = A^l - \frac{1}{2} KHA^l
\]  

(5)

and (iii) calculating the analysed ensemble by adding the updated ensemble mean to the updated anomalies:

\[
X^a = A^a + [\bar{X}^a, \ldots, \bar{X}^a]
\]  

(6)

**Prediction metrics**

The user-defined perturbation model and the chosen DA technique will likely require some tuning parameters. To compare the performance of the scheme for different sets of parameters and events and assess the efficacy of the DA, the quality of the ensemble predictions needs to be quantified. Quality refers to the correspondence between the predictions and the matching observations (e.g. Bennett et al. 2013) and can be broken down in several aspects or attributes, e.g. accuracy, reliability and sharpness (Murphy 1993). The accuracy is related to the difference between individual predictions and observations. In case of a probabilistic prediction, the accuracy can either be computed from the mean or the full distribution function. A prediction is deemed to be reliable if there is an agreement between the prediction probability and mean observed frequency, i.e. if on average \( x\% \) of the observations fall within the \( x\% \) prediction interval. Therefore, predictions that consistently over- or underestimate the frequency interval of the observations are considered unreliable. The sharpness is related to how the prediction probability approaches zero and one, so does not depend on the observation. A sharp prediction tends to have probabilities clustered around the mean, as opposed to values near 0 or 1. Prediction metrics typically focus on one or more attributes of the prediction quality. Therefore, the choice of metrics depends on the objective of the verification and generally a combination of metrics is needed for a thorough analysis (Anctil & Ramos 2019). We focus on the predicted water levels and identified six metrics suitable to compare the ensemble predictions against the sensor observations (Table 1). Some of these are designed for deterministic forecasts, so they are applied to the ensemble mean.

The mean absolute error (MAE) and the root mean square error (RMSE) are commonly used measures of

| Table 1 | Comparison of prediction metrics (the measuring unit refers to water level predictions) |
| Verification metric | Unit | Range | Best score | Quality attribute | Type of prediction |
|---------------------|------|-------|------------|------------------|-------------------|
| Mean absolute error (MAE) | m    | 0 to \( \infty \) | 0 | Accuracy | Deterministic |
| Root mean square error (RMSE) | m    | 0 to \( \infty \) | 0 |            |                   |
| Nash–Sutcliffe efficiency (NSE) | –    | \(-\infty \) to 1 | 1 |            |                   |
| Coverage bias (CB) | –    | \(-1 \) to 1 | 0 | Reliability | Probabilistic |
| Average band width (ABW) | m    | 0 to \( \infty \) | 0 | Sharpness |                   |
| Continuous Rank Probability Score (CRPS) | m    | 0 to \( \infty \) | 0 | Mix of the above | |


accuracy. They are calculated as:

\[
\text{MAE} = \frac{1}{N} \sum_{k=1}^{N} |X_k - d_k| 
\]

(7)

\[
\text{RMSE} = \left[ \frac{1}{N} \sum_{k=1}^{N} (X_k - d_k)^2 \right]^{1/2} 
\]

(8)

where \(X_k\) is the mean across ensemble members at the \(k\) time step, \(d_k\) is the corresponding observation and \(N\) is the total number of time steps in the observation period. Both metrics retain the same unit of the prediction, but the RMSE penalizes large errors, which in urban drainage are known to occur especially for peak flows and water levels that can be significantly affected by timing errors due to uncertain rain inputs.

The Nash–Sutcliffe efficiency (NSE) is calculated as:

\[
\text{NSE} = 1 - \frac{\sum_{k=1}^{N} (X_k - d_k)^2}{\sum_{k=1}^{N} (\bar{d} - d_k)^2} 
\]

(9)

where \(\bar{d}\) is the mean value of the observations over the observation period. NSE ranges between 1 and \(-\infty\), with the highest value being a perfect fit. Negative values indicate the ensemble mean is a worse predictor than the mean of the observations. NSE is a normalized metric, so it is useful to compare predictions across different events.

The coverage bias (CB) can be used to quantify the reliability of a probabilistic prediction. For a given quantile level \(\beta\), the prediction interval is defined as:

\[
I_k(\beta) = [X^l_k; X^u_k] 
\]

(10)

where \(X^l_k\) and \(X^u_k\) are, respectively, the lower and upper limits of the ensemble prediction at levels \(\beta/2\) and \(1 - \beta/2\). The observed coverage \(n(\beta)\) is the fraction of observations falling within \(I_k(\beta)\) during the observation period and is expected to match the nominal coverage \(1 - \beta\). CB for a given quantile \(\beta\) is defined as the discrepancy between the nominal and observed coverage

\[
\text{CB}(\beta) = 1 - \beta - n(\beta) 
\]

(11)

and ranges between \(-1\) and 1 (Thordarson et al. 2012). If the bias is negative, the observed coverage is larger than the nominal one and the prediction is overestimating the uncertainty. The opposite is true if the bias is positive. Therefore, the CB can be used to quantify the reliability of a probabilistic prediction, since the ideal reliability corresponds to \(\text{CB}(\beta) = 0\) for any \(\beta \in [0, 1]\).

The average band width (ABW) for a given quantile level \(\beta\) is calculated as:

\[
\text{ABW}(\beta) = \sum_{k=1}^{N} I_k(\beta) 
\]

(12)

and is an indicator of the prediction sharpness, although not sufficient to characterize it (Gneiting et al. 2007). In this study, it is computed to assess how the perturbation and assimilation affect the spread of the ensemble prediction.

The continuous ranked probability score (CRPS) for a given \(k\)th pair of observation-ensemble prediction is calculated as:

\[
\text{CRPS}(k) = \int_{-\infty}^{\infty} |F_k(x) - F^P_k(x)|^2 dx 
\]

(13)

where \(x\) is the predicted variable, \(F_k(x)\) and \(F^P_k(x)\) are, respectively, the cumulative distribution function of the \(k\)th probabilistic prediction and of the corresponding observation (Hersbach 2000). When the observation is a unique value, its distribution function is given by:

\[
F^P_k(x) = \begin{cases} 
0 & \text{for } x < d_k \\
1 & \text{for } x \geq d_k 
\end{cases} 
\]

(14)

which is also known as the Heaviside function. To remove the time dependency, the continuous ranked probability score is averaged over the observation period in this study (simply referred to as CRPS). For a deterministic forecast, \(F_k(x)\) can also be approximated with a step-wise function and the CRPS is reduced to the mean absolute error. Similar to the MAE, the CRPS retains the measuring unit of the predicted variable and ranges between 0 and \(\infty\), with 0 equal to a perfect score. Since the CRPS measures the quadratic distance between the cumulative distribution of the prediction and the observation, it mainly quantifies the accuracy of the
prediction while also carrying information about its reliability and sharpness.

**Model-based sensor validation**

One of the potential benefits of DA is the model-based cross-validation of sensors, i.e. using the updated model to identify malfunctioning sensors. This requires that the sensor under analysis is not used for the model updating, in order to ensure that the model prediction and the sensor observation are independent. All sensor observations carry a degree of uncertainty. Assuming for simplicity that (i) the observation error is normally distributed around the measured quantity with variance \( \sigma_{\text{obs}}^2 \); (ii) the ensemble is regarded as Gaussian with variance \( \sigma_{\text{ens}}^2 \) and (iii) the model ensemble provides a correct representation of the distribution of the measured quantity, then the sensor values should be distributed around the ensemble mean with a variance of \( \sigma_{\text{obs, ens}}^2 = \sigma_{\text{obs}}^2 + \sigma_{\text{ens}}^2 \). This implies that the standardized residuals

\[
r_{\text{std}} = \frac{(d - X)}{\sigma_{\text{obs, ens}}}
\]

should be normally distributed with zero mean and a variance of \( \sigma_{\text{obs, ens}}^2 \), and that about 95% of the values should fall within the \([-2, 2]\) interval. Therefore, the standardized residuals could be used to test the validity of a sensor by checking if the mean of \( r_{\text{std}} \) over a given time period is statistically different from zero and if its value exceeds a confidence interval of \([-2, 2]\) for prolonged periods in time. In the current article, we are only using \( r_{\text{std}} \) for visual identification of suspicious sensor data. However, the validation process could be automated and the user could in a real-time setting be alerted when the sensor data does not pass the quality check.

**EXPERIMENTAL SETUP**

**Case study**

The Damhus combined sewer overflow tunnel in Copenhagen, Denmark, was used as a case study. The tunnel is integrated within a large combined sewer system that drains to the local wastewater treatment plant. We were provided with an up-to-date and highly detailed MIKE URBAN model of the urban drainage system (Figure 5(a)). This DUDM was forced with rainfall intensity series from 10 gauges distributed in and around the catchment perimeter and included all dry weather inflow patterns (both domestic and industrial). MIKE URBAN automatically assigns to each sub-catchment the rainfall intensity from the nearest gauge to simulate the spatial distribution of the precipitation. The large size of the model, which counts more than 15,000 links (equivalent to 800 km of pipes) and a similar number of nodes, translates into a high computational cost, making this an interesting case study for testing the proposed DA scheme.

Several overflow structures were originally constructed in the catchment to discharge excess combined sewage directly into the local stream in case of intense rain events. Recent environmental regulations have limited the permitted number of overflows per year, with the aim of protecting the water quality at the stream and the nearby harbour. In response, the Damhus tunnel was built along the path of the stream to intercept the outflow of the overflow structures connected to the eight tunnel dropshafts (Figure 3(b)). These range in diameter between 1.5 and 12.5 m and part of their volume contributes to the tunnel storage capacity. Ultrasonic water level sensors placed at the most upstream and downstream dropshafts are used to monitor the filling degree of the tunnel. With a length of 3.4 km and a constant diameter of 3 m, the tunnel system can store up to 29,000 m³ of water. The invert level of the tunnel varies between 13.7 and 15.7 m below the ground level. Main dimensions are reported in Figure 3(e) as relative to the Danish Vertical Reference 1990 (DVR90).

Each dropshaft receives inflow from one or more inlet pipes connecting the pre-existing overflow structures. Most of the overflows are passive, so they are activated when the water level upstream exceeds a fixed crest level. Others are regulated with gates and valves to control the timing and magnitude of the discharge. As shown in Figure 3(c) for the most upstream dropshaft, some inlet pipes receive inflow from a single overflow, others branch off to connect multiple overflows. The 17 boundaries of the tunnel model are set in agreement with the assimilation scheme criteria, and mostly coincide with the overflow structures. Three of the four boundary locations upstream are also equipped.
with water level sensors, which provide valuable information on the filling of the tunnel, given that in most cases more than 50% of the tunnel inflow comes from the upstream inlets.

At the downstream dropshaft, the water level builds up gradually as the tunnel fills up until there is no more overflow into the tunnel or the max storage capacity is met. To prevent water from backing up upstream, a safety overflow releases excess water to the stream when the tunnel has reached its full capacity. Five large pumps lift the stored water from a cavity at the bottom of the downstream dropshaft back to the combined sewer. The pumps are controlled by the water level measured in the tunnel and at the pumps outlet. To ensure a realistic simulation of the tunnel emptying, the pumps and their outlet are retained in the tunnel model, with the outlet serving as a boundary (Figure 3(d)).

Observations

Since the tunnel receives overflows from the combined sewer, it is only employed during high-intensity rainfall events. Once the tunnel content has been pumped back to the sewer, sediments are removed by flushing. An ‘event’ here refers to a cycle of partial/complete filling, emptying and flushing. Eight events occurred in the period between August and December 2018. Among them, three characteristic events (A, B and C) were selected to cover various seasons (respectively, late summer, autumn and early winter) and a wide range of tunnel filling degrees (Figure 4). The sensor data from these three independent events were used in this study.

Rainfall intensity observations from the 10 gauges linked to the drainage model were obtained at 1-min resolution from a national archive hosted by the Danish
Meteorological Institute (Jørgensen et al. 1998). For Event A, the average of the 10 rainfall series showed a peak intensity of 30.2 mm/h (8.4 μm/s) and a total duration of about 2.5 h. This first event was recorded at the beginning of August and had the characteristics of a summer storm, or cloudburst. On the other hand, Events B and C, which happened in October and December, showed a moderate peak intensity of 6.8 and 2.5 mm/h (1.9 and 0.7 μm/s), respectively, but extended for durations of 5.5 and 14 h, respectively.

Water level observations from six ultrasonic sensors were obtained at 1-min resolution from the local utility company. They covered the most upstream and downstream dropshafts, three of the upstream boundaries and the pumps outlet for the three events. During Event A, large volumes of combined sewage were rapidly discharged into the tunnel, which filled up entirely in about 2.5 h. At this point, the water level rose above the crest of the safety overflow at 0.2 m and the excess water was discharged into the receiving water body (see a flat plateau in the middle, bottom-left panel of Figure 4). Once the water level at the downstream outlet dropped below the control set-point, the pumps were activated to return the water to the combined sewer for disposal at the nearby treatment plant. When the tunnel pipes were completely filled, which occurs at −6.6 m, the observed water level upstream almost matched the level downstream, with differences ranging between a few centimetres and 0.5 m. The lower rain intensity of Event B led only to the partial filling of the tunnel. All the combined sewer overflow was stored in the tunnel and pumped back without significantly building up at the upstream dropshaft, where the maximum recorded water level was about 0.4 m above the invert as opposed to about 10 m in the first event. Given its peak intensity, Event C was expected to behave similarly to Event B. While this was mostly the case for the downstream water level, sudden high water levels were recorded upstream, thus suggesting a possible malfunctioning of the upstream sensor. The incident was later explained by the tunnel operator. The upstream dropshaft includes a separate flushing chamber as tall as the dropshaft itself. At the end of each event, the tunnel is flushed by opening a valve at the bottom of the chamber. During Event C, the flushing chamber was accidentally filled beyond its capacity and the overflowing water spilled into the detection zone of the sensor located in the main chamber, thus causing a false echo.
DA framework

The assimilation of water level observations in the ensemble of tunnel models was carried out with the DA framework developed by the Danish Hydraulic Institute (DHI) and embedded in the beta version of MIKE URBAN 2019. Earlier versions of this framework were previously used with ground water models (Zhang et al. 2015) and urban drainage models (Borup et al. 2018), the advantage being the complete automation of the assimilation process, which removes the need to directly access the model states at each time step. Setting up the framework still includes, though, some manual steps: (i) generate an ensemble of size \( m \) by cloning the MIKE URBAN model \( m \) times; (ii) introduce the model error by perturbing the input (in our case, the boundary conditions); (iii) link the observation series to the corresponding structures in the model; (iv) assign the uncertainty of the observations; (v) define the spatial extent of the assimilation for each observation location, a process known as localization; and (vi) choose the assimilation algorithm for the update (with the option of running the framework as an open-loop simulation without assimilation). The software then automates the parallel running of the ensemble members and, at each time step, propagates the model states forward, updates them based on the available observations and replaces the predicted states with the updated ones. The process is repeated until the end of the simulation period.

Implementation

The inner model was set to consist of the tunnel and the nearest connecting structures only (the dark-blue and red parts in Figure 3(b)–3(d)) with a total of 28 links. The inner model can simulate a 2-day event in 10 s that takes the model of the full system 2 h to simulate, corresponding to a ~700 times reduction in the computational time. This implied that the main computational effort of the DA scheme in the specific implementation was to run the outer model, when using an ensemble size of 100 and updating the ensemble to observations once per minute.

The construction of the perturbation model is crucial for the success of the assimilation scheme and is case-specific. For our case study, the uncertainty of the simulated water levels at the boundaries of the tunnel model can be formulated in terms of time delay and peak value. To capture these two features, the boundary conditions were perturbed with a time displacement uniformly distributed within the interval \([- \tau_{\text{max}}, \tau_{\text{max}}]\) and a normal distributed multiplication factor with zero mean and standard deviation \(\sigma_p\). Both parameters were assumed constant in time but varied across locations and ensemble members. This is a rather simplistic approach that ignores the spatial and temporal correlation of the model errors at the tunnel boundaries. However, we find this sufficient for the current study, and our choice of perturbation scheme facilitates the tuning process and allows us to regulate independently the time delay and the peak values of the model boundaries.

A total of five series of water level observations were assimilated in the tunnel model ensemble, 1 from the downstream dropshaft and 4 from the surrounding boundaries. The sensor at the downstream dropshaft also measures water levels in the pump sump, where large variations in level correspond to small changes in volume. For simplicity, the cavity was not included in the model and all observations lower than the lowest tunnel invert were set equal to it. The sensor data from the upstream dropshaft was only used for cross-validation and to test the ability of the update to propagate across the entire tunnel. To account for the different ranges of water level at the tunnel dropshafts and in the combined sewer, and how they can affect the uncertainty of the measurement, we set separate values of the standard deviation at the downstream tunnel dropshaft \(\sigma_{a,1}\) and at the boundary locations \(\sigma_{a,b}\). Given the different hydraulic conditions inside and outside the tunnel, we also adjusted the localization settings separately, limiting the assimilation of the observation at the downstream dropshaft only to the states inside the tunnel and performing a point-wise update at the boundary location. The latter effectively reduces the uncertainty of the simulated states at some of the tunnel boundaries based on the information provided by the sensors, with the benefit of limiting unrealistically perturbed states.

Due to the arbitrary nature and correlation of the perturbations, some of the model realizations are likely to simulate extreme hydrodynamic conditions, which can cause the failure of the computational engine. To alleviate this phenomenon, we enlarged the Preissmann slot to 10% of the pipe diameter. The enlarged width slightly changed the volume of the system and therefore contributed to the
overall uncertainty of the model, but helped in simulating the pressure waves induced in the tunnel by extreme inflows, thus stabilizing the simulations.

**Tuning of ensemble and DA parameters**

In the proposed implementation, a total of five parameters ($\tau_{\text{max}}$ and $\sigma_p$ in the perturbation model, $\sigma_{o,t}$, $\sigma_{o,b}$ and $m$ in the DA framework) need to be tuned by the user to optimize the performance of the scheme. To better understand the individual effect of the parameters, each was treated as an independent variable and tested with an initial guess and two alternative options, while the remaining parameters were fixed. The preferred value among the three options was selected based on the response of the scheme quantified with the proposed prediction metrics and left unchanged in the following tests. The scheme was tuned with Event A, and the chosen parameters were subsequently applied to Events B and C.

**RESULTS AND DISCUSSION**

**Tuning (Event A)**

The scheme parameters were individually tuned by trial and error for Event A. In each trial, the error of the updated ensemble at the downstream dropshaft was quantified using the selected prediction metrics (Table 1). CRPS was used as reference metric, given its ability to capture various attributes of the prediction quality. MAE, RMSE and NSE were used for comparison and added interpretability but are expected to behave consistently with the reference metric. CB and ABW were both computed for the 90% quantile interval and used to gain insight into the reliability and sharpness of the ensemble prediction. The tuning started from the parameters of the perturbation model ($\tau_{\text{max}}$ and $\sigma_p$), then moved to the parameters of the assimilation scheme ($\sigma_{o,t}$ and $\sigma_{o,b}$) and ended with the ensemble size $m$ (Table 2).

The results of the tuning are discussed in detail in the Supplementary material. The process is highly specific to the case study and the event analysed, but leads to some general insights: (i) the value of the perturbation parameters needs to be large enough to represent the uncertainty of the model while avoiding the simulation of unrealistic flow conditions; (ii) the standard deviation of the observations is the single most sensitive parameter, given that it directly affects the extent of the update; (iii) the size of the ensemble only marginally affects the result, but a larger ensemble will return a better description of the states distribution. The final selection of parameters (corresponding to the 5th trial) returns the lowest CRPS of all trials but cannot be

| Parameters | Verification metrics |
|------------|----------------------|
| $\tau_{\text{max}}$ (min) | $\sigma_p$ | $\sigma_{o,t}$ (m) | $\sigma_{o,b}$ (m) | $m$ (–) | MAE (m) | RMSE (m) | NSE (–) | CB$_{90}$ (–) | ABW$_{90}$ (m) | CRPS (m) |
| 1 | 5 | 0.01 | 0.5 | 0.05 | 100 | 0.55 | 0.90 | 0.96 | 0.62 | 0.23 | 0.43 |
| 2 | 10 | – | – | – | – | 0.58 | 0.99 | 0.95 | 0.56 | 0.37 | 0.45 |
| 3 | 30 | – | – | – | – | 0.42 | 0.72 | 0.97 | 0.45 | 0.39 | 0.30 |
| 4 | – | 0.02 | – | – | – | 0.49 | 0.79 | 0.97 | 0.33 | 0.88 | 0.31 |
| 5 | – | 0.04 | – | – | – | 0.14 | 0.20 | 1.00 | 0.20 | 0.51 | 0.09 |
| 6 | – | – | 1 | – | – | 0.24 | 0.34 | 0.99 | 0.17 | 0.81 | 0.18 |
| 7 | – | – | 3 | – | – | 0.48 | 0.72 | 0.97 | 0.08 | 1.76 | 0.56 |
| 8 | – | – | 0.5 | 0.1 | – | 0.16 | 0.23 | 1.00 | 0.24 | 0.45 | 0.10 |
| 9 | – | – | – | 0.3 | – | 0.20 | 0.29 | 1.00 | 0.25 | 0.61 | 0.11 |
| 10 | – | – | – | 0.05 | 50 | 0.19 | 0.28 | 1.00 | 0.34 | 0.45 | 0.11 |
| 11 | – | – | – | – | 25 | 0.20 | 0.30 | 1.00 | 0.38 | 0.44 | 0.12 |
| Selection (5) | 30 | 0.04 | 0.5 | 0.05 | 100 | 0.14 | 0.20 | 1.00 | 0.20 | 0.51 | 0.09 |

Table 2 | Prediction metrics of assimilated, ensemble water level prediction at the downstream dropshaft during Event A given by different sets of scheme parameters

R. Palmitessa et al. | Updating models of urban drainage tunnels for soft sensing and sensor validation
Journal of Hydroinformatics | 23.3 | 2021
Downloaded from http://iwaponline.com/jh/article-pdf/23/3/438/892821/jh0230438.pdf by guest
assumed to be the global optimum. Figure 5 (left column) shows the observed, perturbed and updated water levels at the downstream and upstream locations (upper and middle row).

**Scheme verification (Events B and C)**

The tuned set of parameters was tested with two independent events (B and C) of lower intensity (Figure 5, middle and right columns). Despite the heterogeneity of the hydraulic conditions, the updated ensemble predictions at the downstream dropshaft showed a similar error in all three events, with CRPS ranging between 0.09 and 0.13 m and NSE always higher than 0.97 (Table 3). These values represent a significant improvement compared to the open-loop prediction without assimilation, where the CRPS ranges between 1.06 and 2.06 m and the NSE assumes negative values. The large discrepancy between the observed water level and the mean ensemble prediction prior to the assimilation can be attributed to the error in the simulated inflow and outflow to and from the tunnel. While the tunnel inflow is governed by the water level at the boundaries read from the output of the outer model, the tunnel outflow depends on the control rules of the pumps, which

![Figure 5](https://doi.org/10.2166/hydro.2020.074)
are activated based on a comparison between the observed water levels and pre-defined set-points. When the control scheme is replicated in the hydrodynamic model, the observed levels are replaced with the simulated ones. Therefore, an error in the simulated states can cause the pumps to behave differently from reality and the model drifts further from the observations. Meanwhile, the assimilation reduces the discrepancy between the observed and simulated water levels, and in doing so indirectly corrects the timing and magnitude of the pumped outflow.

Sensor validation

Once the quality of the updated predictions has been verified, they can be used to validate independent observations that are not assimilated in the model ensemble. This was the case for the observed water level at the upstream dropshaft. To quantify the discrepancy between the observed and expected value at the upstream location, we computed the standardized residual using Equation (15) assuming a standard deviation of observation error of $\sigma_{obs} = 0.1$ m. During Event A, the tunnel was filled to its maximum capacity, at which point the difference in water level at the upstream and downstream dropshaft is expected to be minimal. The observed values are in agreement with this assumption. Also, for most of Event A, the updated ensemble prediction upstream shows a similar accuracy to the downstream prediction, with CRPS of 0.15 m and NSE of 0.99 (Table 3), which is a significant improvement over the open-loop prediction.

The filling degree was much lower during Event B, and the water level at the upstream dropshaft raised little more than 0.3 m above the invert as opposed to almost 3 m at the downstream dropshaft. It can be assumed that the inflow moved rapidly downstream and accumulated only in the downstream section of the tunnel. As observed, the water levels upstream and downstream are expected to exhibit different behaviours and be mostly uncorrelated. This is further confirmed by the observed delay between the peak values. This circumstance likely explains the only marginal improvement in the ensemble prediction after the assimilation of the downstream observation. Since the update is driven by the Kalman gain, which in turn is a function of the covariance of the model states, only the states correlated with the downstream water level are expected to be affected by the update. Nonetheless, both open-loop and updated ensemble prediction upstream showed similar accuracy to the downstream water level with CRPS equal to 0.09 m.

Similar to Event B, Event C was driven by a low-intensity precipitation event of long duration, resulting in a partial filling of the tunnel as shown by the downstream observation. This is in agreement with the ensemble prediction upstream, both before and after the assimilation. However, the observation from the upstream sensor suggests that a large volume of water entered the tunnel and filled it up completely. However, while $r_{std}$ mostly stayed within the confidence interval $[-2, 2]$ during Events A and B, it raised to above $50$ for several hours during Event C, clearly indicating that the sensor was malfunctioning for this period in time (Figure 5, bottom row).

**CONCLUSIONS**

We developed a DA scheme tailor-made for system-wide soft sensing and sensor validation for urban drainage
tunnels. DA has the potential to equip hydrodynamic models with soft sensing capabilities, by improving their accuracy and expressing their uncertainty. Ensemble-based DA methods are particularly suitable for DUDMs but have high computational costs. By separating the ‘inner’ tunnel model from the ‘outer’ drainage model and updating only the inner model, our proposed scheme retains the benefits of distributed models and ensemble-based DA, while restraining the computational overhead. We applied the assimilation scheme to a 1D hydrodynamic model of a 3.4 km-long combined overflow tunnel, whose outer model counts over 15,000 links (equivalent to 800 km of pipes) and a similar number of nodes. The inner model covered only the tunnel and connecting structures with a total of 28 links and was roughly 700 times faster than the outer model. When using an ensemble size of 100 and updating the ensemble to observations once per minute, the main computational effort was related to running the outer model. The performance of the scheme was observed to be sensitive to the choice and tuning of the perturbation model that simulated the uncertainty of the tunnel input. A similar sensitivity was observed in the tuning of the standard deviation of the observations. We note that the tuning process is highly case-specific and requires balancing the degrees of confidence in the model and the sensors, while staying within realistic flow conditions.

In all three events tested, the updated water levels at the downstream dropshaft scored better when compared to observations than the ensemble predictions without assimilation. The update was also correctly propagated across the tunnel as confirmed by an independent observation at the most upstream dropshaft. Thus, the model generated an accurate, system-wide estimate of the hydraulic conditions. In the third event, the large discrepancy between updated and observed states enabled the automatic detection of a false echo in the upstream sensor. In this case, the updated ensemble was capable of reconstructing the likely water level upstream, acting as a soft sensing tool. Further research is needed to promote the proposed scheme into a reliable, operational tool, by for example introducing an automatic assessment of both observation and model error.

ACKNOWLEDGEMENTS

We thank Lone Bo Jørgensen and Anders Breinholt from the Greater Copenhagen Utility (HOFOR) and the Danish Meteorological Institute (DMI) for providing the model and data for the case study. We also thank DHI for granting access to an early version of MIKE URBAN’s DA framework. This work is part of a joint PhD programme between the Technical University of Denmark and the Nanyang Technological University, Singapore.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Anctil, F. & Ramos, M. 2009 Verification metrics for hydrological ensemble forecasts. In: Handbook of Hydrometeorological Ensemble Forecasting. Springer Berlin Heidelberg, Berlin, Germany, pp. 893–922. doi:10.1007/978-3-642-39925-1.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D. & Andreassian, V. 2013 Characterising performance of environmental models. Environmental Modelling and Software 40 (February 2017), 1–20. http://dx.doi.org/10.1016/j.envsoft.2012.09.011.

Borup, M. 2014 Real Time Updating in Distributed Urban Rainfall Runoff Modelling. PhD Thesis, Department of Environmental Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark.

Borup, M., Grum, M. & Mikkelsen, P. S. 2011 Real time adjustment of slow changing flow components in distributed urban runoff models. In: Proceedings of 12th International Conference on Urban Drainage, pp. 11–16. Available from: http://orbit.dtu.dk/ws/files/6373440/prod1132162523249.PAP005261.pdf (accessed 14 January 2019).

Borup, M., Grum, M., Madsen, H. & Mikkelsen, P. S. 2014 Updating distributed hydrodynamic urban drainage models. In Proceedings of the 13th International Conference on Urban Drainage, pp. 1–8. Available from: http://orbit.dtu.dk/ws/files/101102837/BorupArticle.pdf (accessed 14 January 2019). doi:10.13140/2.1.1682.1766.

Borup, M., Madsen, H., Grum, M. & Mikkelsen, P. 2018 Technical note on the dynamic changes in Kalman gain when updating
hydrodynamic urban drainage models. *Geosciences* 8 (11), 416. doi:10.3390/geosciences8110416.

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. & Tait, S. 2012 Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth* 42–44, 3–10. doi:10.1016/j.pce.2011.04.007.

DHI 2009 MIKE URBAN. Collection System, Horsholm, Denmark.

Evensen, G. 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99 (C5). doi:10.1029/94jc00572.

Gneiting, T., Balabdaoui, F. & Raftery, A. E. 2007 Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69 (2), 243–268. http://doi.wiley.com/10.1111/j.1467-9868.2007.00587.x.

Hansen, L. S., Borup, M., Møller, A. & Mikkelsen, P. S. 2014 Flow forecasting using deterministic updating of water levels in distributed hydrodynamic urban drainage models. *Water (Switzerland)* 6 (8), 2195–2211. doi:10.3390/w6082195.

Hersbach, H. 2000 Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* 15 (5), 559–570. doi:10.1175/1520-0434(2000)015<0559:DCRPSE.2.0.CO;2.

Hutton, C. J., Vamvakерidou-Lyroudia, C., Kapelanz, L. & Savić, D. 2011 Real-time modelling and data assimilation techniques for improving the accuracy of model predictions. In *PREPARED 2011 Deliverable D3.6.2*, p. 83.

Hutton, C. J., Kapelanz, L., Vamvakeridou-Lyroudia, C. & Savić, D. 2014 Real-time data assimilation in urban rainfall-runoff models. *Procedia Engineering* 70, 843–852. doi:10.1016/j.proeng.2014.02.092.

Jørgensen, H. K., Rosenørn, S., Madsen, H. & Mikkelsen, P. S. 1998 Quality control of rain data used for urban runoff systems. *Water Science and Technology* 37 (11), 113–120. doi:10.2166/wst.1998.0448.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H-J., Kumar, S., Moradkhani, H., Seo, D-J., Schwanenberg, D., Smith, P., van Dijk, A. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O. & Restrepo, P. 2012 Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences* 16 (10), 3863–3887. doi:10.5194/hess-16-3863-2012.

Lund, N., Madsen, H., Mazzoleni, M., Solomatine, D. & Borup, M. 2019 Assimilating flow and level data into an urban drainage surrogate model for forecasting flows and overflows. *Journal of Environmental Management* 248 (December 2018), 109052. doi:10.1016/j.jenvman.2019.05.110.

Murphy, A. H. 1995 What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting* 8 (2), 281–293. doi:10.1175/1520-0434(1995)008<0281:WIAGFA>2.0.CO;2.

Palmitessa, R., Borup, M. & Mikkelsen, P. S. 2018 Urban tunnel systems for conveyance and storage of storm- and wastewater: features, classification, and modelling. In: *Proceedings of the 11th International Conference on Urban Drainage Modelling* (G. Mannina, ed.). Palermo, Italy, pp. 251–254. Available from: http://orbit.dtu.dk/ws/portal/dspace/154733729/UDM18_abstract_Palmitessa_2906.pdf.

Sakov, P. & Oke, P. R. 2008 A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus, Series A: Dynamic Meteorology and Oceanography* 60A (2), 361–371. doi:10.1111/j.1600-0870.2007.00299.x.

Thordarson, F. Ö., Breinholt, A., Møller, J. K., Mikkelsen, P. S., Grum, M. & Madsen, H. 2012 Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion. *Stochastic Environmental Research and Risk Assessment* 26 (8), 1151–1162. doi:10.1007/s00477-012-0563-3.

Tscheknert-Gratl, F., Zeisl, P., Kinzel, C., Rauch, W., Kleidorfer, M., Leimgruber, J. & Ertl, T. 2016 Lost in calibration: why people still do not calibrate their models, and why they still should – a case study from urban drainage modelling. *Water Science and Technology* 74 (10), 2337–2348. doi:10.2166/wst.2016.395.

Zhang, D., Madsen, H., Ridler, M. E., Refsgaard, J. C. & Jensen, K. H. 2015 Impact of uncertainty description on assimilating hydraulic head in the MIKE SHE distributed hydrological model. *Advances in Water Resources* 86, 400–413. http://dx.doi.org/10.1016/j.advwatres.2015.07.018.