Estimating inflow to a combined sewer overflow structure with storage tank in real time: evaluation of different approaches

G. Leonhardt, M. D'Oria, M. Kleidorfer and W. Rauch

ABSTRACT

The performance assessment of storage tanks and combined sewer overflow (CSO) structures in sewer systems requires knowledge of the total inflow from the catchment during rainfall events. Many structures are, however, only equipped with sensors to measure water level and/or outflows. Based on the geometry of the tank, expressed as a level-storage relationship, inflow can be calculated from these data using a simple conceptual storage model. This paper compares a deterministic and a Bayesian approach for estimating the inflow to a CSO structure from measurements of outflows and water level. The Bayesian approach clearly outperforms the deterministic estimation which is very sensitive to measurement errors. Although computationally more demanding, the use of a simple linear storage model allows the online application of the Bayesian approach to repeatedly estimate inflow in short time intervals of a few minutes. The method could thus be used as an online software sensor for inflow to storage structures in sewer systems.

Key words | Bayesian geostatistical approach, bgaPEST, combined sewer overflow, reverse level pool routing, storage structure

G. Leonhardt (corresponding author)
M. Kleidorfer
W. Rauch
Unit of Environmental Engineering,
Institute of Infrastructure Engineering,
University of Innsbruck,
Technikerstrasse 13,
A 6020 Innsbruck,
Austria
E-mail: *guenther.leonhardt@uibk.ac.at*

M. D'Oria
Department of Civil, Environmental and Land Engineering and Architecture,
University of Parma,
Parco Area delle Scienze 181/A,
43124 Parma,
Italy

INTRODUCTION

Storage tanks and combined sewer overflow (CSO) structures (or combinations of both) are important parts of sewer systems with regard to the control of emissions to receiving waters. Very often, these storage structures are equipped with one or more sensors, for example, to measure the water level in the tanks or any of the outflows to the receiving water or to a waste water treatment plant. Due to the difficult conditions, such as backwater effects, wide ranges of flows, large pipe diameters or difficult access for the installation of devices, the inflow is rarely measured. However, knowledge of inflow to a structure, which is at the same time the outflow of a connected catchment, would be valuable information, for example, to assess the efficiency of the structure to reduce CSO emissions or for the calibration of a rainfall-runoff model. If sufficient other data (such as measurements of water level or outflow) are available, these can be used to calculate the inflow using a simple conceptual storage model. Thus, the storage structure can be used as a flow meter for the total inflow from the connected catchment.

doi: 10.2166/wst.2014.331

In the case where the dynamics inside a storage structure can be neglected, outflow Q_{out} can be calculated based on inflow Q_{in} and stored volume V using the 'level pool routing' approach (Chow *et al.* 1988). This is based on the continuity equation, which relates inflow, outflow and the change of volume with respect to time t:

$$\frac{dV}{dt} = Q_{in}(t) - Q_{out}(t) \quad (1)$$

Water level h in the structure and stored volume are related by the level storage relation, that is, the function $V = f(h)$ which can be derived from the geometry of the tank. In the case of a CSO structure, Q_{out} comprises both excess flow Q_E to the receiving water and outflow to the treatment plant Q_{TP}. Q_E is usually discharged via an overflow weir and depends on the water level, and Q_{TP} might either be regulated by a throttle valve or also be dependent on h. For conceptual modelling of sewer systems, this is often further simplified. By assuming a maximum storage volume, unlimited overflow and ideally regulated outflow to the treatment plant, the water level h can be omitted (see e.g. Butler & Davies 2004).

The estimation of inflow Q_{in} to a storage structure, based on the knowledge of Q_{out} and $V = f(h)$, is referred to as 'reverse level pool routing'. Whereas the reformulation of Equation (1) and its numerical solution is usually straightforward, problems occur if the measured data and geometric information (h, Q_{out} and $V = f(h)$) are affected by measurement errors (e.g. due to measurement noise, drift of the sensor, poor resolution or waves) and errors in the geometric data. As a derivative is calculated from measured data, measurement errors are amplified and hence spurious oscillations are introduced to the estimated inflow. To cope with these problems, the use of different numerical schemes and filtering of measured data have been proposed.

Zoppou (1999) investigated different numerical schemes for the discretisation of Equation (1) with respect to 'reverse level pool routing' and proposes to use a centred explicit scheme. He also introduces a specific filter to be applied to model output, that is, the calculated Q_{in}. Filtering of input data is a commonly suggested approach when calculating derivatives from measured data (e.g. Olsson & Newell 1999). However, many low-pass filters induce peak smoothing of the highly dynamic measured time series (Bogner & Constantinides 1975). This is especially a concern in urban drainage modelling where the time of concentration and consequently the modelling time step is short (in the range of minutes).

A different approach is to use a forward model (e.g. Equation (1) or any other model), and optimize the inflow time series in order to achieve the best fit between forward model output and measured data. However, such a problem is highly parameterized and given the presence of errors in the observations, imprecisions in the geometric data and simplifications in the routing model, non-uniqueness and overfitting of the solution must be overcome. Recently D'Oria *et al.* (2012) proposed a Bayesian approach to estimate the inflow hydrograph to a reservoir based on the forward solution of the level pool routing model equation. The method is referred to as the Bayesian Geostatistical Approach (BGA); it was initially developed (Kitanidis 1995) and applied for parameter estimation in groundwater problems such as pumping test analysis (e.g. Snodgrass & Kitanidis 1998), hydraulic tomography tests (e.g. Li *et al.* 2007), tracer data inversion (e.g. Fienen *et al.* 2009) and contaminant source identification (e.g. Snodgrass & Kitanidis 1997). Applications outside the subsurface context mainly involve atmospheric modelling (e.g. Michalak *et al.* 2004) and reverse flow routing problems (D'Oria & Tanda 2012; D'Oria *et al.* 2012, 2014).

If the calculation of inflow is performed in real time, further restrictions must be considered. As an instantaneous calculation of inflow at any time during an event can be based on past measurements only, not all numerical schemes can be used. Furthermore, the application of filters suitable for use in real time introduces a lag to the time series (Olsson & Newell 1999). If inflow is estimated based on a forward model and optimization, a repeated application in short time intervals requires a short computation time.

In this paper, we present an application of the Bayesian approach according to D'Oria *et al.* (2012) for online estimation of inflow to a CSO structure in short time intervals (i.e. a few minutes). The results are compared to a deterministic approach based on a numerical solution of the reverse level pool routing equation and filtering of measured data. Results are checked for plausibility and discussed with respect to their uncertainties.

MATERIALS AND METHODS

Case study

The approach is demonstrated using data from a CSO structure in Austria. The structure is situated at the outlet of a small catchment (9.6 ha paved area) and connected to an offline tank with a maximum storage capacity of 286 m^3. Sensors for independent measurements of excess flow Q_E, outflow to the treatment plant Q_{TP} and water level h were installed by the operator. The data are retrieved and stored in 2-minute time steps.

The storage tank has a circular layout and consists of a conical and a cylindrical part. It is separated from the sewer by a side weir, upstream of a throttle valve. The water level is measured in the tank with an ultrasonic sensor and transmitted as relative data (integer values in %) with a resolution corresponding to 2.6 cm. The stored volume V is calculated from the water level. Based on the resolution of the water level data, the tank geometry and accounting for further imprecisions, a standard uncertainty $u(V) = 1.25$ m^3 was estimated for the cylindrical part using the law of propagation of uncertainties. Outflow Q_{TP} is measured with a magnetic flow meter. The maximum design outflow $Q_{TP,max}$ of 0.02 m^3/s is controlled by a throttle valve. Excess flow is discharged to the receiving water over another side weir (with higher crest) from the

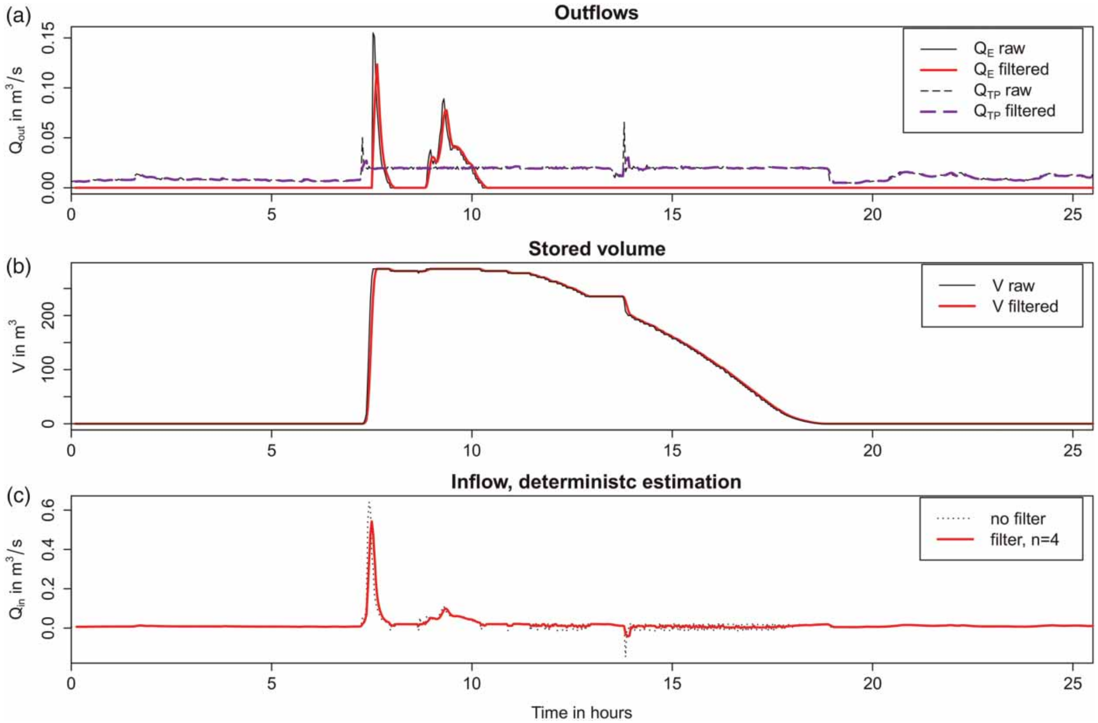

Figure 1 | (a) Measured outflows, (b) stored volume and (c) deterministically estimated inflow to the CSO structure using raw and filtered data.

sewer. The excess flow rate is computed from water level measurements from another ultrasonic sensor at the overflow weir using a simple overflow formula. Consequently, the storage volume in the sewer, between the two weirs, is not captured by measurements. Measured time series of outflows Q_E and Q_{TP} and stored volume V from an overflow event used for demonstration are shown in Figure 1.

Conceptual model of the storage structure

With regard to an application of the reverse model in real time, the following simple numerical scheme was chosen for the solution of Equation (1)

$$\frac{V(t) - V(t - \Delta t)}{\Delta t} = Q_{in}(t) - Q_E(t) - Q_{TP}(t) \quad (2)$$

In Equation (2), Δt refers to the time step, and Q_{out} has been replaced by the sum of Q_{TP} and Q_E. As both outflows in the case study are measured by independent sensors, their dependence on the tank water level must not be considered in the model. The formulation in Equation (2) allows the calculation of $V(t)$ based on data from the current (t) and the previous time step ($t-\Delta t$), as required in a real time application

$$V(t) = (Q_{in}(t) - Q_E(t) - Q_{TP}(t))\Delta t + V(t - \Delta t) \quad (3)$$

A drawback of Equation (3) compared to other discretization schemes is the fact that errors in $V(t)$ are propagated to subsequent time steps.

Deterministic reverse level pool routing

Reverse model

As for the forward case, the numerical scheme in Equation (2) can be rearranged to calculate $Q_{in}(t)$ from measurements available at time t as follows:

$$Q_{in}(t) = Q_{TP}(t) + Q_E(t) + \frac{V(t) - V(t - \Delta t)}{\Delta t} \quad (4)$$

The output of the reverse model Q_{in} is thus calculated from measured data only and does not depend on previously calculated values. However, the derivative of V is approximated by two consecutive values and is thus more sensitive to measurement errors than in a trapezoidal scheme.

Filtering of input data

To remove noise and reduce its effects on the calculated inflow, the measured time series of Q_{TP}, Q_E and V were smoothed using a moving average filter. In preliminary experiments, this filter has proven more suitable than exponential smoothing, which is another common method for pre-processing of measurement data (Olsson & Newell 1999). For online application, the filter has to be formulated as a 'causal' filter. Thus, each value of the filtered time series $\hat{y}$ is calculated from current and past values of the raw data y as follows:

$$\hat{y}(t) = \frac{1}{n}(y(t) + y(t - \Delta t) + \cdots + y(t - (n-1)\Delta t)) \quad (5)$$

The filter parameter n, that is, the size of the moving window, must be chosen with regard to the characteristics of the data to be smoothed. The choice should be made with regard to the dynamics of the catchment, that is, the 'time to peak' and the duration of peaks. Furthermore, the fact that a causal filter introduces a lag of $1 + n/2$ time steps to the time series should be considered (Olsson & Newell 1999).

Bayesian estimation of inflow

The second approach is referred to as the BGA. The unknown inflow hydrograph Q_{in} is considered as a vector of random variables that are to be estimated through comparison between modelled (using the forward model) and measured data, that is, the stored volume V in the presented case. Making use of regularization constraints avoids non-uniqueness and overfitting of the solution. It is assumed that the covariance of the parameters can be modelled by means of geostatistical functions and tools. The type of the covariance model is chosen as prior information in terms of Bayes' formula. The parameters of the covariance model, referred to as 'structural parameters', are then inferred from the data. The model parameters, that is, the values of Q_{in}, are optimized based on local linearization of the model. Thus, the procedure is computationally less demanding than Monte Carlo-based methods. The estimation of structural parameters and model parameters is carried out in alternate iterations (D'Oria et al. 2012). The uncertainty of Q_{in}, in terms of posterior covariance, also results from the method. Physical constraints, such as non-negativity, can be considered by estimating the parameters in a transformed solution space. The parameter estimation is carried out using bgaPEST, a freely available software (Fienen et al. 2013).

Apart from the parameters of the covariance model, another structural parameter, referred to as 'epistemic uncertainty', σ_R, influences the solution. This parameter describes the residuals, that is, the difference between the output of the forward model and the available observations. It thus comprises uncertainties from different sources, most important measured data and model simplifications. The epistemic uncertainty defines the covariance of the likelihood function, which is characterized by a Gaussian distribution. Its parameters can be estimated together with the other structural parameters based on a prior assumption or set to a fixed value.

Each optimization of Q_{in} requires the evaluation of the Jacobian matrix, expressing the sensitivity of model output V with respect to Q_{in}. Usually, this matrix is calculated by a finite difference method, which requires one forward model run per parameter plus one base run. Consequently the total computational effort increases with the length of the time series Q_{in}. In an offline application the computation time is usually not a critical issue, and Q_{in} can be estimated for an entire event or even longer periods. In contrast, an online application is only suitable if the estimation of inflow $Q_{in}(t)$ is provided with a short delay after the availability of measurements of $V(t)$, $Q_E(t)$ and $Q_{TP}(t)$. If the estimations are used to control the system, the acceptable time delay depends strongly on the characteristics of the latter, for example, the filling time of the storage tank or the time of flow travel to other control devices.

The estimation of Q_{in} for a shorter period reduces the required computational effort, but not to an extent which is suitable for urban drainage problems, that is, for measurement time steps of a few minutes. However, the computational effort can be reduced significantly if the Jacobian matrix of the model can be assessed directly, that is, without a large number of model evaluations. In the present case, where the model is linear, the Jacobian matrix is even constant and can thus be calculated in advance.

Estimation of inflow in real time

To provide an estimation of inflow $Q_{in}(t)$ in real time, the procedure has to be repeated as soon as new input data are available, that is, after a few minutes. As the Bayesian approach is based on a certain model for the covariance of the unknown parameters, the estimation of a single value would not be meaningful. Hence, time series of a certain length must be considered. In place of considering the entire time series of measured data, Q_{in} is estimated for a period of $T = n\Delta t$, ranging from time $t-T$ to t. In each estimation, this window of length T is moved forward by Δt. For further reduction of computational effort, the results from the previous estimation are used as initial parameter values for the current one.

As a result, n estimations are performed for a certain point in time t in the course of an online application. The data management and the control of bgaPEST for an online application have been implemented in the R software (R Core Team 2013).

RESULTS AND DISCUSSION

In this section, we first present the results of the deterministic approach to estimate the inflow to the CSO structure. As discussed above, the deterministic approach (Equation (3)) can be applied to both online and offline estimations. These results are compared to inflow estimated with the Bayesian approach for the entire event, that is, in an offline application. Finally, results from the online application of the Bayesian approach are presented. In this context, we discuss the choice of the estimation period T and some practical issues.

All methods have been applied to different CSO events with the same filter parameters and prior values of structural parameters, respectively. As the results show a similar performance of the different methods, only those for the data set in Figure 1 are presented.

Deterministic reverse level pool routing

The results of the application of Equation (3) to both raw and filtered measurements are shown in Figure 1. The inflow estimated from raw data shows many oscillations and even negative values, in particular for periods of decreasing volume. In those periods, the sum of outflows is small and errors in V are thus more visible in Q_{in}. The large negative spike during the plateau in the measurements of V ($t = 13.5$ hours) occurs, however, not only due to measurement errors. It is also related to real processes which are not represented in the simple conceptual model: when the decreasing water level in the storage tank reaches the level of the weir crest separating sewer and offline tank, the volume in the sewer is first emptied by the throttle valve. Only then is the valve of the tank opened and the measured water level decreases.

The relative volume error, calculated as the difference between total inflow (considering only positive values) and outflow volumes with respect to the total outflow volume, is used as a measure to assess the estimated inflow. For the case of raw data, the relative volume error is 7.3%.

A causal moving average filter with parameter $n = 4$ was applied to smooth the input data. The deterministically estimated inflow from the filtered data shows fewer oscillations, but negative inflow values cannot be avoided completely. In addition, the main inflow peak is smoothed and thus reduced by about 0.098 m^3/s or 15%. Furthermore, the lag introduced by the filter results in a shift of the peak of two time steps (4 minutes) with respect to the estimation from raw data. The relative volume error of the inflow calculated from filtered data is 1.1%. Increasing the parameter n would further reduce oscillations but also introduce a larger lag and more damping of the peaks, which occur in reality.

Bayesian estimation of inflow

Offline application

In an offline application, the inflow hydrograph for the entire event is estimated at once. To study the effect of the epistemic uncertainty on the solution, estimations were carried out with both a fixed value for σ_R and an estimation of it. Based on the assumed uncertainty of the measurements of V, the fixed value was set to $\sigma_R = 1.25\ m^3$. For the estimation of σ_R a prior value of $1\ m^3$ was chosen.

Results of both offline estimations are very similar: the largest differences in Q_{in} occur at the main peak and range from -0.03 to 0.006 m^3/s. A value of $\sigma_R = 1.39\ m^3$ was estimated for the epistemic uncertainty in the second case, which corresponds rather well with the assumed uncertainty of V.

Figure 2(a) shows the results for the estimation with a fixed value of $\sigma_R = 1.25\ m^3$. The estimated inflow hydrograph is similar to the general shape of the deterministic results but free of spurious oscillations and negative values. It is thus a physically meaningful result. Even during the plateau in measured volume, low positive inflow values are estimated. In contrast to the use of filtered data, the inflow peaks are not smoothed. Thus, the imposed covariance together with the transformation of the solution space allows a physically meaningful solution without unrealistic smoothing. The relative volume error is -0.026% and thus considerably smaller than both

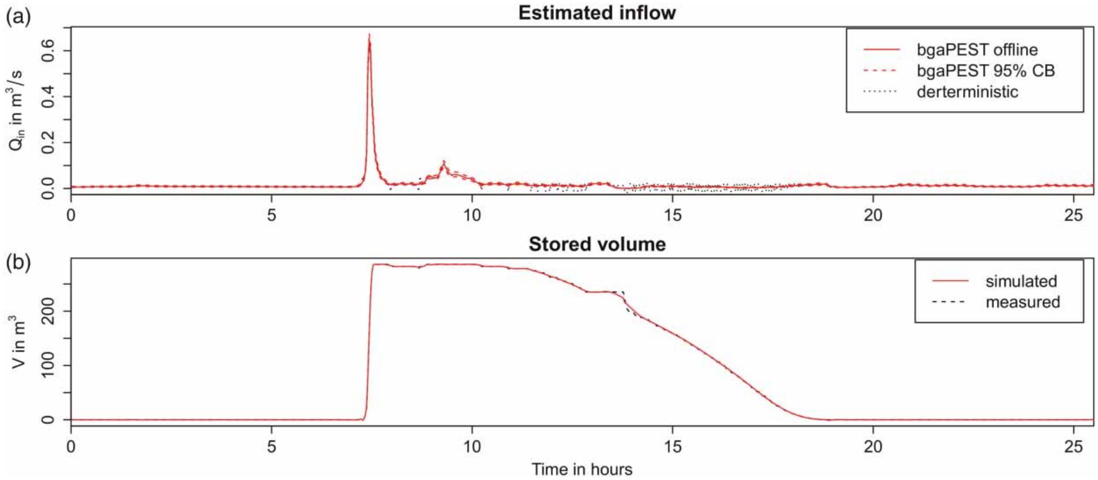

Figure 2 | (a) Inflow estimated with the Bayesian approach for the entire event, 95% credibility bounds (CB) and deterministic estimate (from raw data); (b) stored volume simulated from estimated inflow using the forward model and measured data.

deterministic results. The 95% credibility range of the inflow derived from the posterior covariance has a maximum width of 0.056 m³/s at the main peak, and an average width of 0.007 m³/s (see Figure 4). This is in the same order of magnitude but still considerably smaller than the damping of the peak due to filtering or the range of oscillations in the deterministic results. The volume simulated with the forward model is shown in Figure 2(b). Except for the plateau at $t = 13.5$ hours, it shows good agreement with the measurements. An improvement of the estimation during the period of the plateau would require an adaption of the model structure of the forward model to represent the real process more accurately.

Online application

The Bayesian inflow estimation was applied in real time using four different estimation periods T. According to the concept of the approach and confirmed by preliminary experiments, the estimation fails if T is too short. Based on those tests, experiments were performed with $T = 50$, 100, 200 and 400 minutes. If T is increased further, the required computation time would not be useful for real time applications in the case study (see discussion below). For $T =$ 50 minutes, several estimations covering the plateau in measured volume failed. As this should not occur in a real world application, these results are not included in the following discussion. For the longer estimation periods, each single estimation was successful. The epistemic uncertainty was fixed to $\sigma_R = 1.25$ m³ in all estimations. In the following discussion, the result of the offline estimation with $\sigma_R =$ 1.25 m³ is used as a reference.

Figure 3 shows the result of a single estimation for $T =$ 100 minutes to demonstrate the method. Results of the online application are compared to that of the offline estimation. The selected online estimation is performed to estimate inflow at $t = 9.367$ hours, considering only measurements in the previous 100 minutes. The estimated inflow shows good agreement with the offline result, except for the first and last point, where differences increase. In general, the 95% credibility range of the online estimation is smaller than that of the offline estimation in the corresponding period. Only for the last point does the range of the online simulation exceed that of the offline result. For the first point, the range is also considerably higher. This shows clearly that the regularisation imposed by the assumed covariance model is less effective at the edges of the estimated time series. The first point is influenced by the initial condition. In contrast to the other points, the estimation of the last point is underpinned by fewer measurements (D'Oria *et al.* 2014).

If all consecutive online estimations are considered, envelope curves can be calculated from the inflow hydrographs and the corresponding credibility bounds (CB). Figure 4 shows these envelope curves with respect to the offline estimation. The upper and lower envelope curves of the inflow hydrograph were calculated as maximum and minimum values of all online estimations, respectively. They represent the result of the entire online application, considering all n estimations for each time t. Figure 4(a) shows that envelope curves of the online application are in good

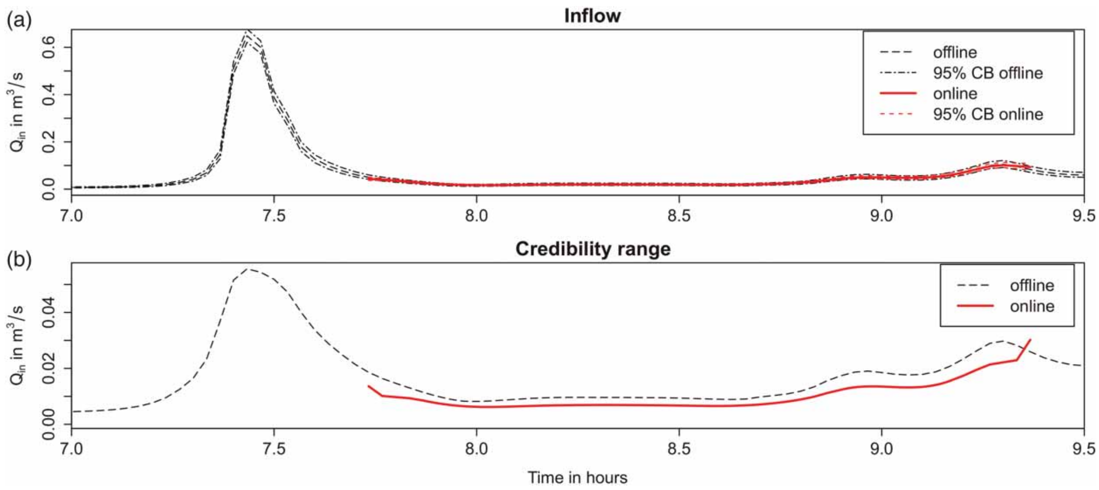

Figure 3 | (a) Estimated inflow for a single online estimation with $T = 100$ minutes and the offline estimation (both with fixed $\sigma_R = 1.25$ m³) and corresponding 95% CB; (b) width of the 95% credibility ranges for both estimations.

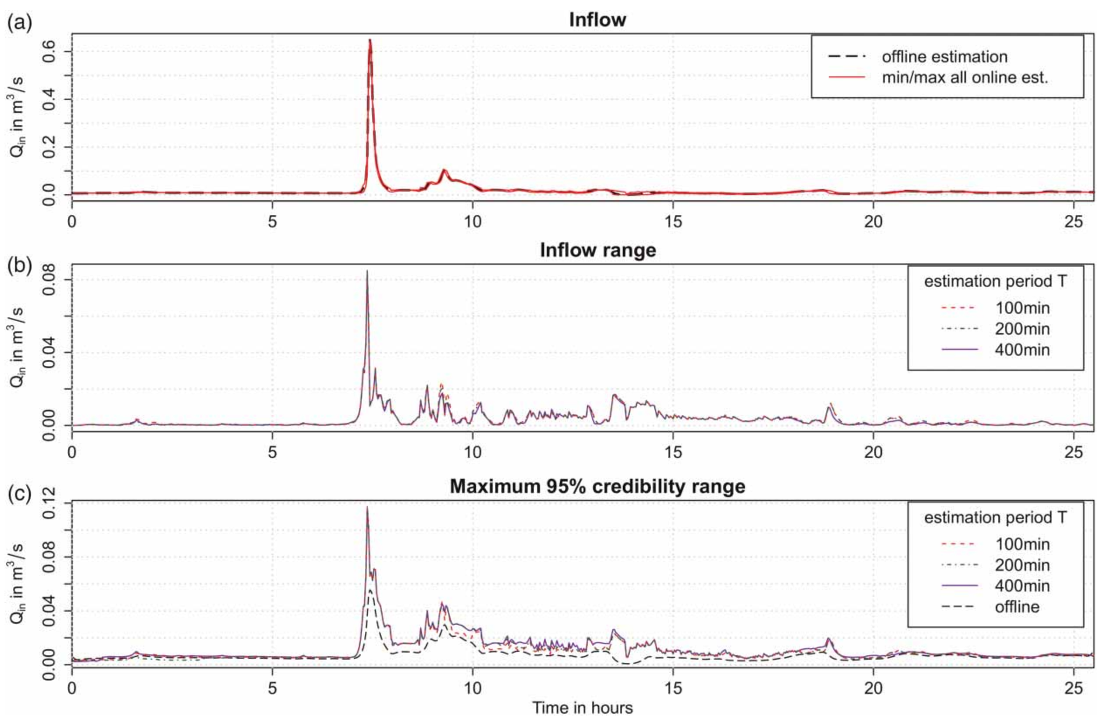

Figure 4 | (a) Upper and lower limit of all online estimations of inflow for $T = 100$ minutes and result of the offline estimation; (b) ranges between upper and lower limits of all inflow estimations for different estimation periods T; (c) ranges between maximum upper and minimum lower 95% CB of all inflow estimations for different estimation periods T, and width of the 95% credibility range for the offline estimation.

agreement with the offline results (the offline result is always between the envelope curves). The ranges between the upper and lower envelope curves for different values of T are shown in Figure 4(b). Differences with regard to the estimation period T are small and show no systematic behaviour. The same applies to the ranges between the maximum upper and minimum lower CB (Figure 4(c)). For periods of $V > 0$ they are generally larger than the credibility range of the offline result, which is caused by the increased uncertainty at the edges of the estimated time series discussed above. These ranges can be considered as the total uncertainty of the online application of the Bayesian approach. As noted for the offline estimation, they are in the same order of magnitude as the damping of the peak and the oscillations in the deterministic results. In addition, they are in the same order of magnitude as uncertainties of typical measurement devices (Bertrand-Krajewski *et al.* 2003). The Bayesian approach can thus provide physically meaningful estimations of inflow in both the offline and online application.

The required computation time for a single estimation including data preparation and management varies in a wide range. Maxima were observed in the period of the plateau in the measured volume. On a desktop computer with an Intel® Core™ i7 2.98 GHz processor, 8GB RAM and 64bit MS Windows 7 operating system, they range from 135 seconds for $T = 100$ minutes to 161 seconds for $T = 400$ minutes. Thus, the online estimation can be performed in time intervals of 3–5 minutes. The choice of the estimation period T should be based on the characteristics of the case study and the available data.

CONCLUSIONS

Two different methods to estimate inflow to a storage structure based on measurements of water level and outflows and knowledge of the geometry have been applied to a data set from a real CSO structure. Both methods are based on the same conceptual storage model, referred to as the level pool routing equation. They can be applied to estimate inflow in real time, which is of interest with regard to control of sewer systems.

The deterministic approach presented is simple to implement for both online and offline applications. However, the results are very sensitive to uncertainties in measurement data and geometric information. Filtering of input data can reduce certain problems, that is, spurious oscillations and negative values in the estimated inflow. However, this is at the cost of loss of important information due to the smoothing of peaks and, in the online case, the introduction of a time lag.

The Bayesian approach considers the inflow time series as a set of parameters to be estimated, using a forward model. A covariance model is used as prior information about the structure of the parameters. The approach provides an estimation of inflow, which is physically meaningful and free of oscillations. Although computationally more demanding than the deterministic approach, it can be applied in real time in combination with certain models, which allow a fast approximation of the Jacobian matrix. Uncertainties are in the same order of magnitude as those of common measurement devices.

The Bayesian approach to estimate inflow clearly outperforms the deterministic method. However, the results remain conditional on the underlying simplified forward model. In order to further improve this procedure, future research should thus focus on methods for the efficient approximation of the sensitivity matrix when non-linear forward models are applied. As the measurement of inflow to storage structures in sewer systems with real sensors remains challenging, the use of the Bayesian approach as an online software sensor is a promising alternative.

REFERENCES

Bertrand-Krajewski, J. L., Bardin, J. P., Mourad, M. & Beranger, Y. 2003 Accounting for sensor calibration, data validation, measurement and sampling uncertainties in monitoring urban drainage systems. *Water Science and Technology* **47** (2), 95–102.

Bogner, R. E. & Constantinides, A. G. 1975 *Introduction to Digital Filtering*. John Wiley & Sons Ltd, New York.

Butler, D. & Davies, J. W. 2004 *Urban Drainage*. Spon Press, London.

Chow, V. T., Maidment, D. R. & Mays, L. W. 1988 *Applied Hydrology. McGraw-Hill Series in Water Resources and Environmental Engineering*. McGraw-Hill, New York.

D'Oria, M. & Tanda, M. G. 2012 Reverse flow routing in open channels: A Bayesian Geostatistical Approach. *Journal of Hydrology* **460–461** (0), 130–135.

D'Oria, M., Mignosa, P. & Tanda, M. G. 2012 Reverse level pool routing: Comparison between a deterministic and a stochastic approach. *Journal of Hydrology* **470–471** (0), 28–35.

D'Oria, M., Mignosa, P. & Tanda, M. G. 2014 Bayesian estimation of inflow hydrographs in ungauged sites of multiple reach systems. *Advances in Water Resources* **63** (0), 143–151.

Fienen, M., Hunt, R., Krabbenhoft, D. & Clemo, T. 2009 Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach. *Water Resources Research* **45** (8), W08405.

Fienen, M. N., D'Oria, M., Doherty, J. E. & Hunt, R. J. 2013 *Approaches in Highly Parameterized Inversion: bgaPEST, a Bayesian Geostatistical Approach Implementation with PEST – Documentation and Instructions*. US Geological Survey Techniques and Methods, Book 7, Section C9, US Geological Survey, Reston, Virginia.

Kitanidis, P. K. 1995 Quasi-Linear Geostatistical Theory for Inversing. *Water Resources Research* **31** (10), 2411–2419.

Li, W., Englert, A., Cirpka, O. A., Vanderborght, J. & Vereecken, H. 2007 Two-dimensional characterization of hydraulic heterogeneity by multiple pumping tests. *Water Resources Research* **43** (4), W04433.

Michalak, A. M., Bruhwiler, L. & Tans, P. P. 2004 A geostatistical approach to surface flux estimation of atmospheric trace gases. *Journal of Geophysical Research: Atmospheres* **109** (D14), D14109.

Olsson, G. & Newell, B. 1999 *Wastewater Treatment Systems: Modelling, Diagnosis and Control*. IWA Publ., London.

R Core Team 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.

Snodgrass, M. F. & Kitanidis, P. K. 1997 A geostatistical approach to contaminant source identification. *Water Resources Research* **33** (4), 537–546.

Snodgrass, M. F. & Kitanidis, P. K. 1998 Transmissivity identification through multi-directional aquifer stimulation. *Stochastic Hydrology and Hydraulics* **12** (5), 299–316.

Zoppou, C. 1999 Reverse routing of flood hydrographs using level pool routing. *Journal of Hydrologic Engineering* **4** (2), 184–188.

First received 18 March 2014; accepted in revised form 14 July 2014. Available online 28 July 2014