Feasibility of using smart meter water consumption data and in-sewer flow observations for sewer system analysis: a case study

N. S. V. Lund **IWA**, J. K. Kirstein **IWA**, H. Madsen, O. Mark,
P. S. Mikkelsen **IWA** and M. Borup **IWA**

ABSTRACT

Globally, smart meters measuring the water consumption with a high temporal resolution at consumers' households are deployed at an increasing rate. In addition to their use for billing or leak detection purposes, smart meters may provide detailed knowledge of the wastewater inflow to the sewer systems in space and time and open up new types of system analyses aimed at closing the urban water balance. In this study, we first validate the smart meter data against other, independent water distribution data. Subsequently, we use a detailed hydrodynamic sewer system model to link the smart meter data from almost 2,000 consumers with in-sewer flow observations in order to simulate the wastewater component of the dry weather flow (DWF) and to identify potential anomalies. Results show that it is feasible to use smart meter data as input to a distributed urban drainage model, as the temporal dynamics of the model results and in-sewer flow observations match well. Furthermore, the study suggests that in-sewer flow observations may be subject to unrecognised uncertainties, which make them unsuitable for advanced investigations of the DWF composition, and this underlines the necessity of collecting data from independent sources. The study also exemplifies that digital system integration in the water sector may be complicated. However, overcoming these obstacles may improve both offline and real-time urban drainage management.

Key words | anomaly detection, distributed model, dry weather flow, smart meters, urban drainage, wastewater flow

N. S. V. Lund **IWA** (corresponding author)
J. K. Kirstein **IWA**
P. S. Mikkelsen **IWA**
M. Borup **IWA**
Department of Environmental Engineering
 (DTU Environment),
Technical University of Denmark,
Bygningstorvet, Building 115, 2800 Kgs. Lyngby,
Denmark
E-mail: nalu@env.dtu.dk

H. Madsen
O. Mark
DHI,
Agern Allé 5, 2970 Hørsholm,
Denmark

N. S. V. Lund
Present address:
EnviDan,
Fuglebækvej 1A, 2770 Kastrup,
Denmark

J. K. Kirstein
Present address:
NIRAS,
Sortemosevej 19, 3450 Allerød,
Denmark

O. Mark
Present address:
Krüger,
Gladsaxevej 363, 2860 Søborg,
Denmark

HIGHLIGHTS

- Smart meter-simulated wastewater flow is a valuable step towards closing the water balance in urban drainage systems.
- Coupling between independent data sources opens up enhanced anomaly detection; here the discovery of potentially erroneous in-sewer observations.
- It can be tedious to gain access to data and models as these may be stored in different silos, but the effort is worthwhile.

INTRODUCTION

The dry weather flow (DWF) describes the flow in the sewer system during periods without rain. This flow consists of wastewater from, for example, households, industry and institutions, as well as groundwater infiltration and rain-induced infiltration into the sewer that can take place for days and weeks after the end of a rain event due to the long transportation time of the stormwater through soil.

Accurate knowledge of the sources of DWF as well as an estimation of its spatial distribution can be used for a wide range of management objectives:

- Estimation of total suspended solids, chemical oxygen demand, organic matter, drugs and nitrogenous pollution loads if the DWF estimation is supported by compound measurements or estimations for different consumer types (Métadier & Bertrand-Krajewski 2011; Schilperoort et al. 2012; Plósz et al. 2013).
- Optimised dimensioning (Cole & Stewart 2013) and operation (Brito et al. 2017) of sewer systems including control during dry weather with focus on energy minimisation and minimisation of sewer overflow during rain events. This could both be based on water quantities and estimated water quality, cf. the first bullet. Both types of control may be carried out using model predictive control (Lund et al. 2018).
- Optimised dimensioning and operation of wastewater treatment plants (WWTPs) (for example, chemical dosing) by knowing how much of the inflow to the plant originates from households, from industry and from groundwater, since these fractions will determine the expected pollutant concentrations (Nguyen et al. 2018), cf. the first bullet.
- Asset management (Djebbar & Kadota 1998; Brito et al. 2017) based on, for example, knowledge of leaking sewer pipes.
- Improved heat recovery from sewage systems (for example, Abdel-Aal et al. 2018) and better planning of decentralised wastewater reuse (Elías-Maxil et al. 2014) by using the estimation of the wastewater flow distribution to pinpoint optimal heat pump locations and optimal wastewater reuse locations, respectively.

In residential areas, the wastewater flow is dominated by the citizens' behaviour. In upstream parts of the system or in small systems, the flow is intermittent (i.e. occurring at irregular intervals) and the detailed flow dynamics may not be captured if too large a sampling interval is used (Butler & Graham 1995; Elías-Maxil et al. 2014). The DWF of smaller areas is thus, in general, more difficult to estimate (Djebbar & Kadota 1998). Further downstream, the aggregation of different dynamic inputs from many small upstream sources results in a change in the nature of the wastewater flow, which forms a less dynamic pattern (Butler & Graham 1995). The DWF from a catchment has previously been studied by a range of authors. For example, Métadier & Bertrand-Krajewski (2011) analysed DWF data and recognised different flow patterns depending on the weekday and date, but also found a relatively large variation within each pattern; Djebbar & Kadota (1998) estimated DWF peaks and average DWF using a neural network model based on the land use and population; and Brito et al. (2017) fitted a partial least-squares model to DWF data to estimate the DWF in situations with missing data. All of these methods may be used to establish DWF patterns but will not give a real-time picture of the DWF. Such real-time information could be obtained from in-sewer measurements, but flow sensors are often scarcely distributed, since it is both expensive and impractical to cover a large urban drainage system (Djebbar & Kadota 1998). Highly spatially distributed real-time DWF information is therefore currently not realistic to obtain using only in-sewer observations.

Water supply and urban drainage systems are intrinsically linked, since most of the consumed water ends up in the sewer system. The wastewater flow can thus be approximated by estimating the water consumption. Butler & Graham (1995) and Elías-Maxil et al. (2014) used a questionnaire and a probabilistic model to estimate the water consumption. Both studies subsequently modelled the resulting flow in the sewer system to obtain a spatial and temporal distribution of the wastewater flow. However, these methods only provide generalised flow patterns. Contrarily, smart meters measure the real-time water consumption in each

household every hour or more frequently, and are increasingly implemented as part of the digitalisation of the water sector, mainly for billing and leakage detection purposes (Boyle *et al.* 2013; Monks *et al.* 2019). Data collected from the water supply system should be pre-processed before use including validation and re-estimation of missing and invalid data (Kirstein *et al.* 2019). After this, the smart meter data can potentially be used to estimate the wastewater flow with a high spatial and temporal resolution. Furthermore, knowledge of the wastewater flow component can, in comparison with in-sewer flow data, be used to estimate the addition or loss of water through infiltration and exfiltration and thus contribute to closing the water balance of the urban drainage system. The concept of applying smart meter data to estimate wastewater flows has been mentioned in the recent literature (Cole & Stewart 2013; Nguyen *et al.* 2018; Monks *et al.* 2019). Zhang *et al.* (2021) used a combination of pressure and flow data from the water distribution system, smart meter data and a water distribution network model to successfully link water consumption with sewage flow. In this effort, they scaled the water consumption based on the sewage flow observations and thus did not take, for example, infiltration/exfiltration and the uncertainty of flow measurements into account.

This study investigates the hypothesis that smart meter consumption data can be used directly to estimate the magnitude, timing and spatial distribution of wastewater flow, without the need of a water distribution network model. This hypothesis is tested using data from the city of Elsinore, Denmark. We validate the smart meter data by comparing the data with observations from the waterworks outlet, observations from the WWTP inlet, and annual water consumption data from a year prior to the installation of smart meters. The smart meter data are also routed through a 1D hydrodynamic urban drainage model for simulating the flow dynamics. We compare the smart meter data and the simulated flow to in-sewer flow measurements from five locations in Elsinore city centre. This comparison forms the basis for assessing anomalies, including unreliable data and other DWF components than wastewater. The comparison is undertaken by using both quantitative methods and logical reasoning to evaluate the wide range of possible sources of uncertainty that arise when employing large amounts of data in a real full-scale urban catchment study.

CASE STUDY AREA

The city of Elsinore, Denmark, is located 30 km north of Copenhagen and has 47,000 inhabitants. It covers an area of 18 km^2 and has a dense medieval centre, surrounded by suburbs developed mainly in the period 1930–1970. The potable water is supplied by four waterworks and the city contains one water tower. The drainage system is predominantly a combined system that leads most of the water to a centralised WWTP. The average yearly precipitation in the city is 670 mm. The upstream parts of the city have a low degree of imperviousness (20–40%), whereas the downstream part containing the city centre is more paved (imperviousness of 60–80%). The utility company has installed permanent smart meters in all the consumers' homes, and in-sewer sensors have been installed for a temporary monitoring campaign in the sewer system, which enables the type of investigations performed in this study. Table 1 explains the main characteristics of the wide range of different data sources that form the basis for this study, which focuses on three week-long periods in October and November 2018 and February 2019.

Data from the water supply system

Around 19,000 MULTICAL21 smart meters with a temporal resolution of 1 h and a measurement uncertainty of up to ±5% (Kamstrup 2019) are permanently installed in households and industries for billing purposes (Figure 1). In the urban area, the meters are placed inside the households. The smart meters cover the entire city, except 14 consumers who still have manually read meters. There are no unmetered consumers. Around 2,000 of these smart meters are situated in the catchment upstream of the in-sewer sensors (Figure 1), which is the main area of interest in this study. Contractor 1 undertook the installation of the smart meters and provided data from three separate weeks in 2018 and 2019 (Table 1). The raw smart meter data are accumulated volume readings. Due to the various data transmission and collection mechanisms of smart meter data, the raw data arrive at non-uniformly distributed time steps (for example, Kirstein *et al.* 2021). Contractor 1 filled the data gaps primarily using linear interpolation to obtain

Table 1 | Information about applied data sets

Data		Temporal resolution	No. of sensors	Declared uncertainty	Provider	Period 1	Period 2	Period 3
Smart meter data		hour	1,970[a]	±5%	Contractor 1	8–14 October 2018	19–25 November 2018[b]	23 February–1 March 2019
Water consumption from 2012		year	2,075[a]	±2%	Contractor 2	2012	–	–
Waterworks outflow data		hour	4	<± 5%	Utility Elsinore	11 October 2018	18 November 2018	23 February 2019
In-sewer flow	Data	2 min	5	±12%	Contractor 4	8–14 October 2018	19–25 November 2018	23 February–1 March 2019
	Latest calibration EU	–	–	–	Contractor 4	2 October 2018	6 November 2018	4 January 2019
	Latest calibration WU	–	–	–	Contractor 4	6 September 2018	6 September 2018	14 January 2019
	Latest calibration ED	–	–	–	Contractor 4	19 September 2018	16 November 2018	29 January 2019
	Latest calibration CD	–	–	–	Contractor 4	6 September 2018	6 September 2018	14 January 2019
	Latest calibration WD	–	–	–	Contractor 4	23 August 2018	7 November 2018	29 January 2019
WWTP inflow data	Data	day	1	±10 m³/h	Utility Elsinore	8–14 October 2018	19–25 November 2018	23 February–1 March 2019
	Latest calibration	–	–	–	–	March 2018		
Last day with rain before each period		–	–	–	SVK data	7 October 2018	14 November 2018	21 February 2019

WWTP, WasteWater Treatment Plant; SVK, SpildeVandsKomiteen. EU, WU, ED, CD and WD are abbreviations for sensor locations and their upstream catchments.
[a]upstream of the in-sewer flow sensors; [b]there is also additional smart meter data from November 18.

a uniform interval between flow values of 1 h. Kirstein et al. (2021) showed that this interpolation procedure sufficiently represents the total consumption within an area as long as a great proportion of the individual consumers' raw smart meter data are not missing and if the raw data resolution is less than 2 h. Contractor 1 furthermore reported the quality of the data. The time periods in Table 1 were selected based on the availability of good quality data in periods without rainfall creating direct runoff in the sewer system.

Hourly outflow data from the four waterworks supplying clean water to the city were obtained from MAGFLOW flowmeters type MAG3100 Water with an uncertainty of less than 5% (Siemens 2020) for specific days in each time period (Table 1). The utility estimated one of the waterwork outlets and the outlet from the water tower based on other observation points, meaning that the actual uncertainty is higher than the declared uncertainty for the instruments.

Furthermore, Contractor 2, who maintains the utility's hydraulic water distribution network model, provided a database with the annual water consumption readings for each individual consumer in Elsinore from 1 January to 31 December 2012. This database contains readings from manually read water meters (read once a year for billing purposes) and is the newest available independent source of water consumption data from before smart meters were installed in Elsinore. In Denmark, it is legally required that each consumer has a water meter at home, and these meters are owned by the utility that performs random control samples to ensure that the meters provide accurate and unbiased readings. The high Danish water price means that most people and the utility are very aware of large changes in the water bills, which further increases the credibility of the consumption data. The 2012 data have a declared uncertainty of ±2% and are thus the best estimation of the annual water consumption we can get that is independent of the smart meters.

Additionally, the utility provided the total amount of distributed drinking water per year in 2012 and 2018 and per month in 2018 and 2019. The utility also provided a list of known pipe bursts in the three time periods.
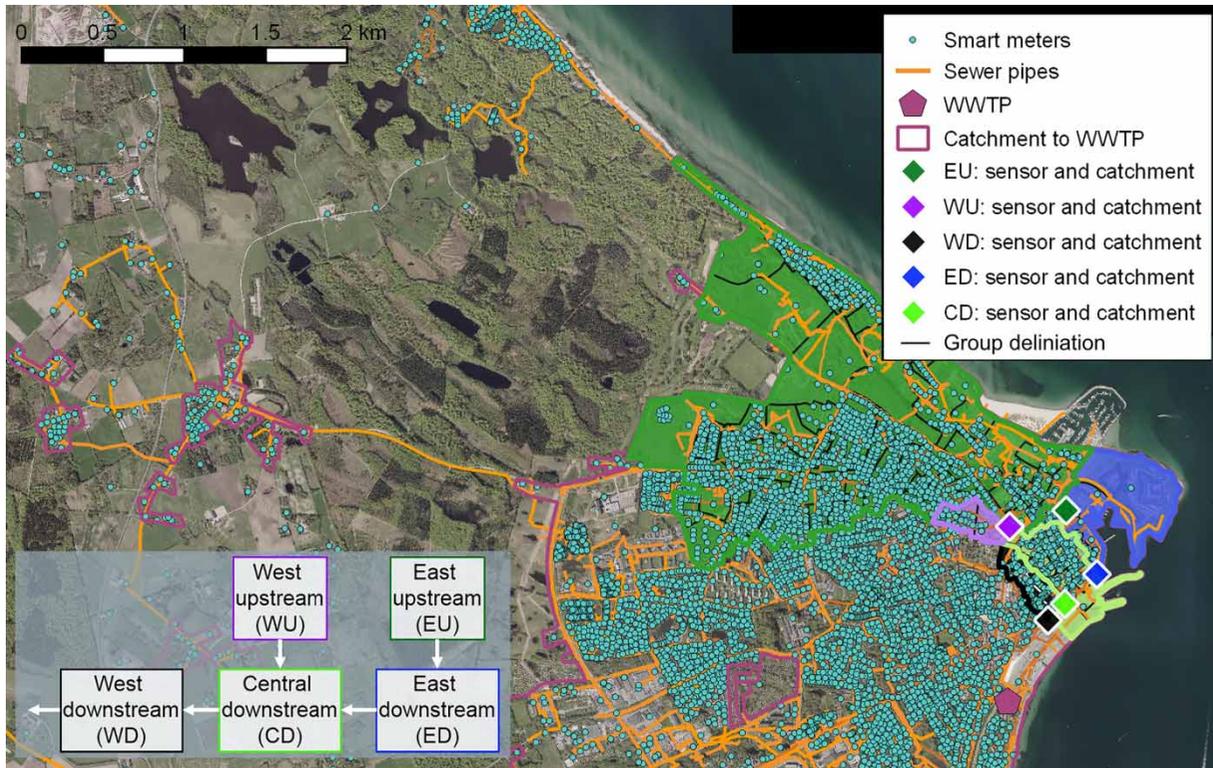
Figure 1 | Map of part of Elsinore, Denmark, showing smart meters, the sewer system, five in-sewer flow observation locations with related sub-catchments (sub-divided into 111 smaller groups for smart meter aggregation), the WWTP and the catchment discharging water to the WWTP. The flow between the sub-catchments in the sewer system is displayed in the lower left corner. EU, ED, CD, WU and WD contain 1,558, 64, 214, 45 and 134 meters, respectively. The extent of the urban drainage system and water supply system goes beyond what is shown in the figure.

Data from the urban drainage system

Utility Elsinore has divided the city into wastewater catchments depending on the layout of the sewer system. These catchments, together with asset data, are described in the municipal wastewater plans, which in detail explain how wastewater and rainwater flows are and will be managed. The asset data have been used as the basis for a detailed 1D hydrodynamic model of the sewer system, which is constructed in MIKE URBAN and maintained by Contractor 3.

Contractor 4 installed level and velocity sensors in Elsinore city centre, as part of a measurement campaign running from May 2018 to June 2019, in order to obtain data for calibration of the model. The sensors were installed at five locations: East Upstream (EU), East Downstream (ED), Central Downstream (CD), West Upstream (WU) and West Downstream (WD) (Figure 1). These abbreviations will be used both to refer to the sensor locations and to their upstream catchments. The level and velocity

observations stem from pressure and Doppler sensors, respectively. The uncertainty of the resulting flow data for the given sensors is estimated to be around ±12% (Franck, T. 2019, Personal Communication, BlueNorth). Two of each sensor types were installed at each of the five locations to ensure that data were present even in case of failure as well as for validation purposes. The contractor also combined the level and velocity data with the geometry of the pipes (multiplying the velocity with the wetted area based on the geometry) to obtain flow observations. The pipes are circular in four of the locations (WU, ED, CD and WD), while one location has an egg-shaped geometry (EU).

Daily inflow data were also available for the WWTP (Table 1). These were aggregated from an ultrasound inlet flow sensor with an uncertainty of $\pm 10 \, \text{m}^3/\text{h}$ (Laursen, B. 2019, Personal Communication, Utility Elsinore). Furthermore, the utility and Contractor 4 provided information on last calibration dates for the WWTP inlet sensor and the in-sewer sensors, respectively (see Table 1).

METHODS

Extraction of smart meter consumption data

Utility Elsinore supplied data from 18,804 meters, of which 18,449 were georeferenced. About 338 of the remaining 355 meters were successfully referenced using the QGIS tool 'MMQGIS'. We manually verified that the 17 non-referenced meters are located outside the upstream area of the in-sewer sensors and thus outside the main area of interest. Due to the general data protection regulation, we could not obtain water consumption data on household level for the 18,787 georeferenced meters, and they should thus be aggregated into groups. It is important that all smart meters within a group discharge water to the same part of the sewer system. Thus, we initially based the groups on the catchments from the Elsinore wastewater plans. 20% of the 18,787 meters were discarded because they did not discharge wastewater to the WWTP used in this study, leaving 15,011 meters. It was important to get an accurate description of the flow dynamics in the area upstream from the in-sewer sensors, and we therefore further manually sub-divided the groups in this area from 19 groups in the wastewater plans to 111 groups with a total of 2,015 meters and between 2 and 38 meters in each group (see the group divisions in Figure 1). This division was based on the general flow paths in the sewer system (described by the asset database). The overall result was 15,011 meters distributed in 245 groups containing each up to 869 meters.

We provided a list with the meters in each group to Contractor 1 who returned the aggregated smart meter water consumption for each group in hourly time steps. 2.4% of the 15,011 meters were missing in the files provided by the contractor, of which 45 were located in the area upstream of the in-sewer sensors. These meters were missing because they had either not yet been replaced with smart meters (which are at most 14 meters) or because they were older meters no longer in use. The total water consumption for these 45 meters was estimated to be around 9,500 m^3 in 2018 (corresponding to a yearly average of 0.3 L/s), which is less than 2% of the water consumption in the area. The final result included 14,694 smart meters in 245 groups, of which 1,970 smart meters belonged to the 111 groups in the area upstream of the in-sewer sensors.

Furthermore, Contractor 1 returned the aggregated consumption data from all smart meters receiving water from the Elsinore Waterworks.

Summed smart meter data

We calculated the summed smart meter data for each of the five in-sewer sensor sub-catchments, $Q_{SM,loc}$, by summing the smart meter consumption, $Q_{SM,loc,i}$, of the N,loc groups belonging to these sub-catchments at all time steps, t:

$$Q_{SM,loc}(t) = \sum_{i}^{N,loc} Q_{SM,loc,i}(t) \tag{1}$$

Simulated wastewater flow

The applied MIKE URBAN model (supplied by Contractor 3) simulates the flow in the sewer system pipes by solving the full 1D Saint-Venant equations. It can furthermore simulate the behaviour and flows through the most important structures and elements in the system such as basins, overflow weirs, pumps and gates. The typical application for this sort of model is to simulate stormwater flows to assess the hydraulic capacity of the system, but it is also capable of simulating flows under dry weather conditions, as in this study. The applied model has a drainage area of 386 ha and contains 1,959 nodes of which seven are basins, as well as 1,986 links, 30 pumps, 33 weirs and one orifice. All asset attributes have been derived from physical properties, such as pipe lengths, slopes and pipe materials. There are no other water sources included in the model than the wastewater flow, which in this study is taken directly from smart meter data. During dry weather, none of the overflows are activated and the only sink is thus the WWTP. The model is thus well suited for routing the wastewater from the consumers to the WWTP and thereby taking into account the time dynamics of the system. These time dynamics have not been calibrated for the dry weather situation.

To simulate the wastewater flow, Q_{sim}, with the MIKE URBAN model, we identified the model entry (boundary) node for each of the 245 smart meter groups. The aggregation of a set of smart meters into one group will cause some water consumption to be delayed and some to arrive too fast to the sewer system, depending on the consumers' distance to the

entry node in the sewer system model. Thus, we decided to add the aggregated smart meter data to the node in the model that was closest to the geometrical centre of each group. The error introduced here will be in the form of a time displacement of the wastewater inflow, and the magnitude of this error is expected to be minor due to the limited size of the smart meter groups. We found the node nearest to the centre of the shape of each group with the Distance Matrix tool in QGIS. The M1D API (DHI 2019) in C# was hereafter employed to automatically use the 245 water consumption time series files as boundary conditions for the 245 localised entry nodes in the MIKE URBAN model. Hereafter, we used C# to run the model, which allowed us to calculate the simulated wastewater flow in the entire sewer system model, including in the five sensor locations ($Q_{sim,loc}$). The model was run with adaptive time steps between 5 and 60 s (but saving the results in 1-min time steps), and it took less than 5 min to run one day of simulations on a Windows server 2012 with an Intel® Xeon® CPU E5-2670 v2 @ 2.50 GHz processor with 20 cores and 20 logical processors when the runoff file was generated previously. The setup used here could thus be applied in a real-time setting.

Phenomena affecting the observed DWF

Q_{SM} and Q_{sim} may differ from Q_{obs} if the system is affected by other water inflows and outflows than the smart meter-measured water consumption or due to erroneous data.

To assess these phenomena, we calculated the average flows for the five sub-catchments (loc) over the three 1-week time periods (168 h or 10,080 min), where $\Delta t_{obs} = 2\,\text{min}$, $\Delta t_{SM} = 1\,\text{h}$ and $\Delta t_{sim} = 1\,\text{min}$ (Table 1):

$$\bar{Q}_{obs,loc} = \frac{\sum\limits_{t=2\,\text{min}}^{10,080} Q_{obs,loc}(t)}{10,080\,\text{min}/\Delta t_{obs}}$$

$$\bar{Q}_{SM,loc} = \frac{\sum\limits_{t=1\,\text{h}}^{168} Q_{SM,loc}(t)}{168\,\text{h}/\Delta t_{SM}} \qquad (2)$$

$$\bar{Q}_{sim,loc} = \frac{\sum\limits_{t=1\,\text{min}}^{10,080} Q_{sim,loc}(t)}{10,080\,\text{min}/\Delta t_{sim}}$$

Furthermore, we calculated the mass balances (change in volume, ΔV) for each of the five sub-catchments based on the water consumption in the given sub-catchments as well as inflows from upstream sub-catchments ($Q_{obs,in,loc}$) and the outflow ($Q_{obs,out,loc}$) in the three investigated 1-week time periods:

$$\Delta V_{loc} = \sum_{t=1\,\text{h}}^{168} (Q_{SM,loc}(t)) \cdot \Delta t_{SM}$$
$$+ \sum_{t=2\,\text{min}}^{10,080} (Q_{obs,in,loc}(t) - Q_{obs,out,loc}(t)) \cdot \Delta t_{obs} \qquad (3)$$

A positive mass balance means that more water enters than leaves a catchment (thus, there may be a loss of water in the sewer system); a negative mass balance contrarily means that there may be an additional source of incoming water. The relative importance of the difference in volume as a function of the outflow from each sub-catchment was calculated as:

$$\text{Rel\%} = \frac{\Delta V_{loc}}{\sum\limits_{t=2\,\text{min}}^{10,080} Q_{obs,out,loc}(t)} \cdot 100 \qquad (4)$$

The difference between observed and simulated flows in the five sub-catchments, $Q_{res,loc}(t)$, was calculated based on the observed and simulated flows:

$$Q_{res,loc}(t) = Q_{obs,loc}(t) - Q_{sim,loc}(t) \qquad (5)$$

These residuals may both be positive and negative, and exhibit constant, diurnal or seasonal variations. They may also vary according to the outside temperature, previous rainfall and pipe geometry. Table 2 lists three overall possible reasons for deviations as well as expected residual patterns. Some of the uncertainties can be quantified, such as the declared uncertainty from the measurement devices (Table 1). The uncertainty associated with the equipment installed *in situ* may, however, be higher than declared by the manufacturer, which can be difficult to quantify without comparing the data with other, independent data sources. The remaining potential uncertainties listed in Table 2 are also difficult to quantify without additional data, since these are often highly contextual and case-specific. One would thus need to make additional independent measurements in the case area for the same time period as the in-sewer

Table 2 | Potential reasons for the deviation between observed and simulated sewage flow

Deviation type	Potential reason	Expected residual pattern
Observed flow smaller than simulated flow (negative residuals, Equation (5))	Consumed water not discharged to the sewer	Diurnal
	Exfiltration	Seasonal
Observed flow larger than simulated flow (positive residuals, Equation (5))	Unaccounted for consumers	Diurnal
	Pumping of groundwater to the sewer system	Constant or seasonal
	Rainwater harvesting	Diurnal
	Snow melting (only in winter)	Temperature-dependent
	Infiltration	Constant-, seasonal- or rainfall-dependent
	Sedimentation	Pipe geometry-dependent
General reasons	Erroneous smart meters, data transmission or data handling	Constant or diurnal
	Wrong conceptualisation of the sewer system	Diurnal
	Erroneous in-sewer sensors	Constant or diurnal

observations were performed. Even here, however, the actual corresponding uncertainty would be difficult to quantify. Complex systems, such as the one presented in this study, are often characterised by higher-order levels of uncertainties rather than pure statistical uncertainty, and in this case pragmatism is called for (Warmink *et al.* 2010). To prevent the pitfall of underestimating the uncertainty by including only the uncertainties that can be quantified explicitly (Warmink *et al.* 2010), the potential causes for deviations between the smart meter and in-sewer data are evaluated based on their effect on the residual pattern calculated by Equation (5). The three sub-sections below elaborate on the three deviation types individually.

Observed flow smaller than simulated flow (negative residuals, Equation (5))

Consumed water not discharged to the sewer. Most of the consumed water ends up in the sewer system either directly, such as water used for showering, or delayed, such as the water used for washing clothes. However, some consumed water will end up elsewhere, such as water that evaporates during cooking or from wet laundry or water that infiltrates in the soil during car washing or gardening; the latter is assumed to be season dependent and may even infiltrate into the sewer system as discussed later. In Denmark, a conservative estimate is that at least 86% of the used water is discharged to the sewer system (HOFOR 2016). This is in alignment with results from, for example, Zhang *et al.* (2021). Since more water is used during the day than at

night, the residuals will be larger during the day and thus exhibit a diurnal pattern. In coastal cities like Elsinore, discrepancies may also arise if ferries and trains take clean water on board in the city for kitchens, toilets and cleaning but discharge the foul water somewhere else or at another time of the day. This may lead to irregular residual patterns.

Leakage may occur in the supply pipe after the smart meter location, which means the metered consumption would not enter the urban drainage system and the observed sewer flow would be smaller than the simulated flow. If the smart meter in this case is placed at the property boundary, the water from a leakage in the supply pipe between the property boundary and the house is likely to infiltrate into the soil, a situation hard to diagnose. If the smart meter is instead placed inside a household, leakage after the smart meter is likely to lead to building damage which is easier to diagnose.

Exfiltration. Exfiltration from the sewer system may occur when there are cracks and fractures in the pipes and the groundwater level is below the sewer system. The groundwater level is generally lower in the summer than during winter, meaning that the exfiltration rate is expected to follow a seasonal pattern with peaks in the summer.

Observed flow larger than simulated flow (positive residuals, Equation (5))

Unaccounted for consumers. Q_{sim} only contains data from consumers that have smart meters installed. Some consumers may, however, have older meters requiring manual

readings. In fact, large consumers will often have such meters installed. Such a deviation would exhibit a diurnal pattern unless missing industrial users consume water, for example, during the night. Some of the water discharged to the sewer system may furthermore originate from somewhere else. This may be the case for, for example, ferries and trains that take on board clean water when they set out and discharge it through the sewer system at their destination. This may result in irregular residual patterns.

Pumping of groundwater to the sewer system. Buildings and construction sites that are located partly underground may need to pump away groundwater and possibly intruding seawater if they are located on the coast. This water may be discharged into the ocean, infiltrated somewhere else or discharged to the sewer system. Such pumping would most likely occur both day and night leading to a constant deviation. If the groundwater level is only an issue in the winter, the residual pattern will be seasonal. Sea-level variations may, however, also affect the groundwater level and thus impact the amount of pumping.

Rainwater harvesting. Rainwater may be used for toilet flushing and laundry, and this will increase the flow in the urban drainage system compared with the recording by the smart meters. Since more water is consumed during the day, the residual pattern would follow a diurnal pattern. An increase in water consumption from the water distribution system would be seen again when the rainwater tanks are empty.

Snow melting. In winter, precipitation falling as snow can cause a delayed runoff into the sewer system. This may increase the measured flow in the urban drainage system. The temperature determines when the snow melts, and the residuals will thus neither follow a constant, diurnal or seasonal pattern.

Infiltration. Infiltration into leaky sewers can arise from either groundwater, rain, groundwater pumped from perimeter drains around buildings and construction sites, or leaking water supply pipes:

- Groundwater levels change slowly over the course of the year, and groundwater infiltration thus exhibits a seasonal change in the infiltration rate, which may also be affected by the sea water level.
- Rain-induced infiltration increases the sewage flow only after rain events and slowly decreases with time.
- Infiltration stemming from the pumping of groundwater exhibits a constant or seasonal pattern (see 'Pumping of groundwater to the sewer system').
- Leakage from the water distribution system into the urban drainage system is expected to be a function of the pressure in the water distribution system. If the system pumping is pressure controlled or if the pressure fluctuates due to variations in demand, the leakages could, to some extent, display a diurnal pattern. In Denmark, the water supply pipes in general have too large diameters due to former expectations of growth in the water consumption and requirements of being able to supply fire hydrants. Therefore, the pressure drops in the distribution systems are small, and temporal variations in demand will only have a minor effect on the pressure in the pipe system. The residuals would thus predominantly be constant over the day.

Sedimentation. Some sewer system flow observations, like the ones in Elsinore, are obtained by multiplying the velocity with the wetted area of the cross-section of the pipe. If sedimentation is present at the location of the sensor, the water level will rise and lead to a larger calculated than actual flow. The residuals will vary in size depending on the geometry of the pipe. Usually sensor companies will try to avoid installing flow sensors in pipes with a lot of sediments, but sedimentation is a dynamic process that may change after the installation of the sensor.

General reasons for deviations between observed and simulated flows (negative and positive residuals, Equation (5))

Erroneous smart meters, data transmission or data handling. Deviations may occur due to erroneous smart meter observations. In general, individual meter errors are not likely to have a noticeable impact on the simulated results since each consumer only uses a minor fraction of the water in each sub-catchment, and suspicious individual meter errors would due to the high water price in Denmark be noticed during the billing process. A general bias of the

smart meters could, however, impact the results. Depending on whether this bias is proportional or additive, it may both result in a diurnal pattern or a constant deviation. Furthermore, poor transmission conditions can result in even larger uncertainties, and wrong mappings of the geographical location of the smart meters will also cause errors.

Wrong conceptualisation of the sewer system. The conceptualisation of the urban drainage system layout is essential to make the correct coupling between the smart meter data and in-sewer sensors. If one or more catchments are falsely connected to the sewer system upstream of the in-sewer sensors, the simulated flow would be larger than the observed. Contrarily, missing connections in the asset database would lead to a smaller simulated than observed flow. This deviation would follow a diurnal pattern.

Erroneous in-sewer observations. In-sewer observations of wastewater velocity, level and flows may be erroneous due to turbulence, presence of solids, aggressive environment (Hager 1994), low water depth (Larrarte *et al.* 2008), or poor maintenance and calibration. The residuals may both follow a diurnal pattern or be a constant offset. Normally, flow observations are assumed to have uncertainties up to 20% (Bertrand-Krajewski *et al.* 2003), leading to a diurnal residual pattern.

RESULTS AND DISCUSSION

Here, we (1) assess if smart meter data can be used to estimate the wastewater component of DWF, (2) identify anomalies, i.e. other sources to the DWF and erroneous data, by comparing the simulated wastewater component with the observed in-sewer flow, and (3) discuss the added value from using a hydrodynamic 1D model.

Assessment of smart meter data for the estimation of wastewater flow

The smart meter data can be used to estimate the wastewater component of the DWF if the smart meter data set is trustworthy.

Figure 2 compares the total outflow from the waterworks and water tower, $Q_{obs,WW}$, to the aggregated smart meter data from the whole of Elsinore, $Q_{SM,Elsinore}$. First, it is noticed that the consumed water exceeds the total outflow from the waterworks for a short period in October. Data about the outflow were missing for this time period, and the period can thus not be used to assess the validity of the smart meter data. Looking at the other two time periods, the smart meters registered on average 14–17 L/s (15–17.5%) less water than $Q_{obs,WW}$. This deviation could be due to leakage in the water distribution network, errors in the waterworks' flow sensor (higher than the <5% reported uncertainty, potentially due to the partial estimation of the total waterwork outflow) or errors in the smart meter data set. The smart meter data set may be erroneous due to meter errors (higher than the 5% reported uncertainty), data transmission errors, faults in the data handling or missing consumers. The latter is, however, not assumed to be an issue, since all consumers are metered and less than 0.1% of the consumers still have manually read meters. The data in Figure 2 thus do not give a clear indication of the validity of the smart meter data.
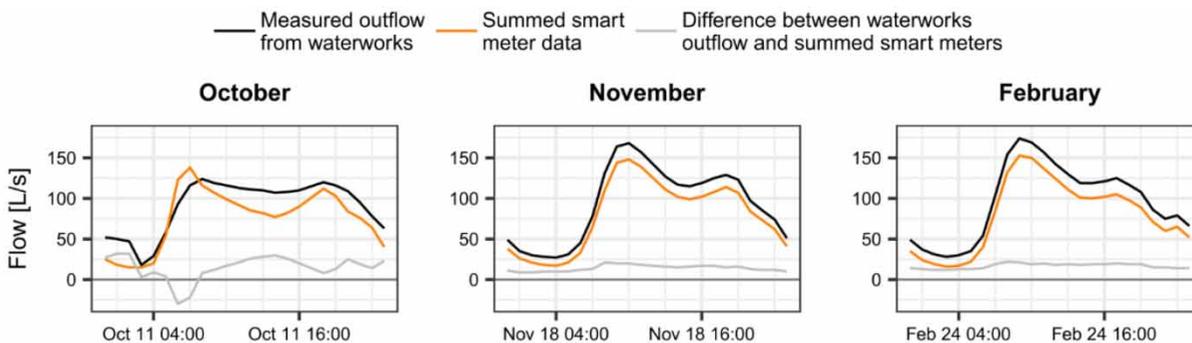

Figure 2 │ Hourly outflow data from the four waterworks and water tower ($Q_{obs,WW}$) compared with summed smart meter data ($Q_{SM,Elsinore}$) for the whole of Elsinore for single days in the three time periods. Note that measured outflow data for October are incomplete.

We further evaluated the completeness and quality of the smart meter data set by comparing the daily inflow to the WWTP, $Q_{\text{obs,WWTP}}$, with the simulated inflow to the WWTP, $Q_{\text{sim,WWTP}}$, using all smart meters in the WWTP catchment (Figure 3). The model was here used as an easy way to include the flow only from smart meters within the WWTP catchment. Since Figure 3 shows daily values and the model does not affect the time dynamics on this time scale, the results for the summed smart meters would be almost the same as the simulated (apart from insignificant differences due to numerical issues). $Q_{\text{sim,WWTP}}$ is around 1.5 times that of $Q_{\text{obs,WWTP}}$ in October and November. This corresponds to, on average, more than 1,500 m³/day, which seems an unlikely amount of exfiltration or an unrealistic volume of consumed water not ending up in the urban drainage system. In February, however, $Q_{\text{sim,WWTP}}$ and $Q_{\text{obs,WWTP}}$ are very similar. $Q_{\text{sim,WWTP}}$ is in the same range as in October and November, but $Q_{\text{obs,WWTP}}$ has increased notably. This cannot be explained by the calibration of the WWTP inlet sensor since this was last calibrated in March 2018. Neither can it be explained by an increase in the outflow from the area with in-sewer observations shown in Figure 3 (sensor location WD). Here, the average daily observed flow is actually larger in October and November than in February, and this is not explained by the smart meter data. The sensed area, however, only partially covers the WWTP catchment (see Figure 1). The deviation may thus instead be due to increased groundwater infiltration in February (winter) in the unsensed part of the WWTP catchment, but it may also be due to erroneous WWTP inlet observations above the reported uncertainty (Table 1).

Since the 2012 consumption data are independent from the smart meter data, we assume that the completeness and quality of the smart meter data set can be assessed by comparing these with this data (see Figure 4). We have added the declared uncertainty as error bars to the figure. For the water consumption from 2012, we have furthermore adjusted this uncertainty to reflect the annual and seasonal variations. From 2012 to 2018, there was a total decrease in the outflow from the waterworks and water tower of 8%, which fits well with a general decrease in water consumption both in Elsinore and in Denmark in general (DANVA 2017; Elsinore Municipality 2019), and the error bars are thus shifted downwards with 8% of the registered consumption value. This means that the error bars reflect the 2012 data's expected ability to predict the 2018 conditions. Furthermore, the monthly average outflow from the waterworks and water tower ranged between −10 to +11% compared with the mean outflow in 2018 and 2019. In October, November and February, the water consumption was specifically 2, 4 and 10% higher than the mean. We have therefore conservatively added a 10% uncertainty to all three time periods on top of the declared uncertainty of 2%. We have not added error bars to the simulated wastewater flow, $\bar{Q}_{\text{sim}}$, since this is simply a post-processing of the smart meter data, $\bar{Q}_{\text{SM}}$. These two differ from each other due to water generated in empty pipes of the model for the sake of numerical stability and because $\bar{Q}_{\text{sim}}$ is affected by the routing time in the sewer system. Overall, the smart meter data and the consumption data from 2012 match well, taking the uncertainties into account. Figure 4 also shows a large difference between the 2012 water consumption data and $\bar{Q}_{\text{obs}}$, which is larger than the declared
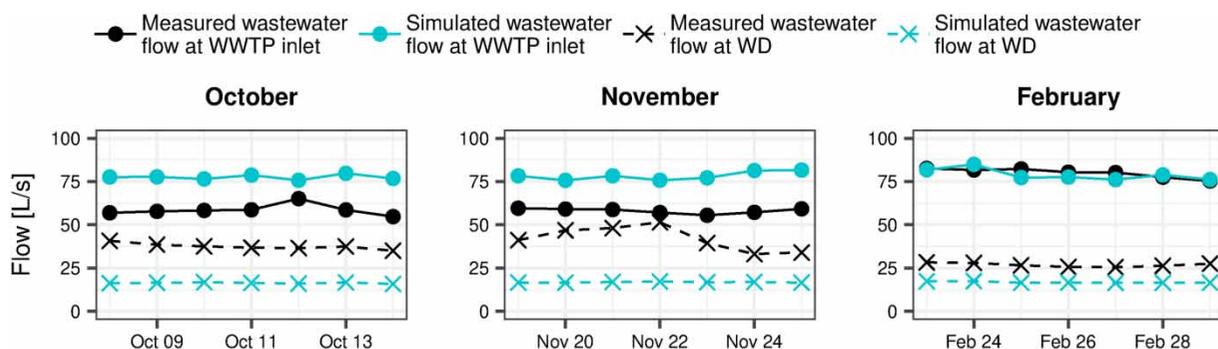

Figure 3 | Observed and simulated flows at the WWTP inlet ($Q_{\text{obs,WWTP}}$ vs. $Q_{\text{sim,WWTP}}$) and at the sensor location West Downstream (WD) ($Q_{\text{obs,WD}}$ vs. $Q_{\text{sim,WD}}$) in the three time periods.
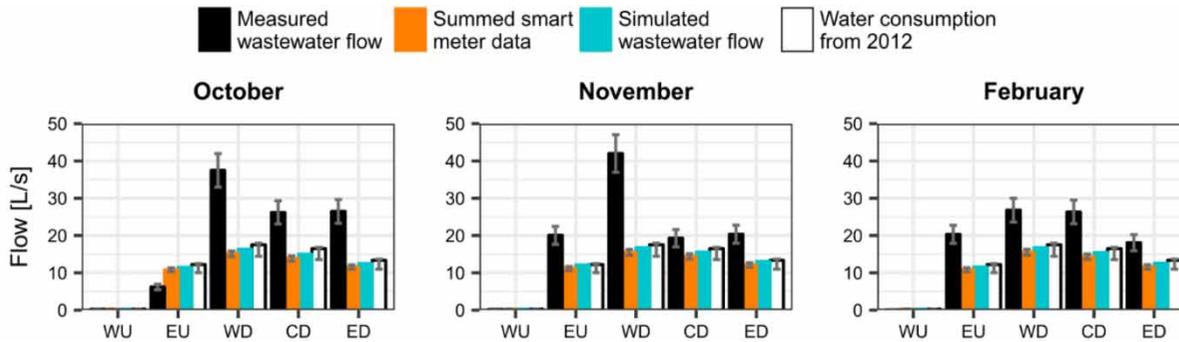
Figure 4 │ Average observed flow to the WWTP ($\bar{Q}_{obs}$), summed smart meter data ($\bar{Q}_{SM}$) and simulated flow ($\bar{Q}_{sim}$) (Equation (2)) over the three time periods for each sub-catchment compared with the average consumption measured independently in 2012 ($\bar{Q}_{consumption,2012}$). The error bars of $\bar{Q}_{obs}$ and $\bar{Q}_{SM}$ indicate the declared uncertainties (Table 1) and are calculated as $\bar{Q} \pm \bar{Q}*$uncertainty. The error bars of $\bar{Q}_{consumption,2012}$ likewise indicate the declared uncertainty (Table 1) as well as the seasonal variation ($\pm 10\%$) and has furthermore been shifted downwards to reflect a decrease in water consumption between 2012 and 2018 of 8%.

uncertainties. Consumers that have private wells will neither be in the 2012 consumption database nor have smart meters installed. However, there are no private or commercial wells from where the water is subsequently discharged to the sewer system in the area upstream of the in-sewer flow sensors (Elsinore Municipality 2019; Pratt, A. 2019, Personal Communication, Elsinore Municipality). Overall, the comparison with the 2012 water consumption data indicates that the amount of unaccounted for consumers and uncertainty related to the smart meters, data transmission and data handling (Table 2) is limited, meaning that the smart meter data set is complete and sufficiently correct to represent the wastewater component of the DWF.

Assessment of anomalies, including other DWF components and erroneous data

Figures 3 and 4 show that there is a discrepancy between the smart meter-based wastewater flows, Q_{sim} and Q_{SM}, and the observed in-sewer flow, Q_{obs}, which is so big that it exceeds

the combined, declared uncertainty of the smart meter and in-sewer sensors (Table 1).

To assess possible anomalies, we look at the mass balances for the five sub-catchments for the three time periods (Equation (3)). From Figure 5, it is evident that there are no clear patterns in which catchments are losing or gaining water in the three periods. Trusting the observations, almost 2,000 m^3 more water leaves 'ED' than enters it during October (negative mass balance), which could indicate a large external water inflow such as, for example, infiltration. From the percentage (Equation (4)) shown in Figure 5, it is clear that this possible external inflow would be the main contributor of the outflow stemming from the sub-catchment. This trend corresponds poorly with the results for November and February that show a small positive mass balance for 'ED'. The natural variations in groundwater levels and soil moisture that could explain month-to-month variations in infiltration would give the opposite pattern, since the summer of 2018 was very dry followed by a wet winter. 'CD' displays some
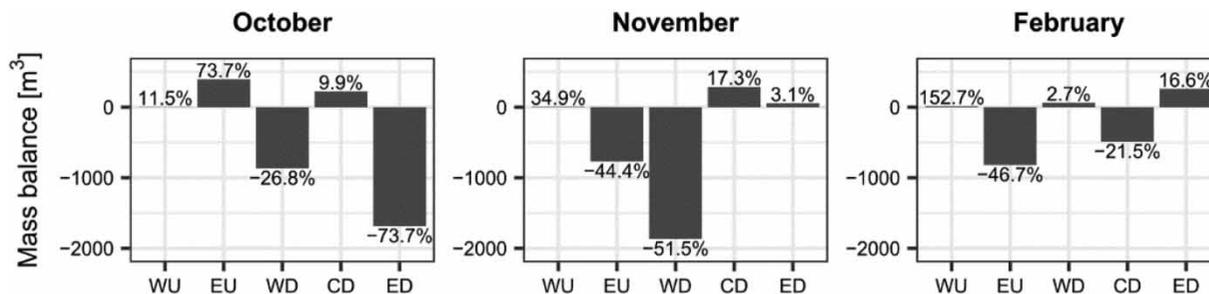
Figure 5 │ Mass balances for each sub-catchment (Equation (3)) and the relative size of the change in volume compared with the outflow (Equation (4)) for the three time periods.

of this expected behaviour, but it would be premature to draw any conclusions from this considering the large unexplainable variation for the remaining sub-catchments.

Figure 6 shows the time dynamics of Q_{sim}, Q_{SM} and Q_{obs}, and is arranged to resemble the flow through the sewer system as conceptualised in Figure 1. The flow is naturally smallest in 'WU', since this is the smallest of the five sub-catchments and does not receive water from any further upstream catchments. The axis for this plot is thus scaled differently than for the remaining plots. The flow increases through the system as more water aggregates. Most of the outflow from the most downstream catchment (WD) is generated in the most upstream catchment, EU, which also has by far the most consumers. The residuals between Q_{obs} and Q_{sim} (Equation (5)) are shown in the upper left corner of Figure 6 for all five sub-catchments.

Figure 6 shows that the flows in WU match well in October and November. There are negative discrepancies between Q_{obs} and the smart meter-based flows (Q_{SM} and Q_{sim}) in EU in October and in WU in February, and many time periods and sub-catchments with positive residuals. In the following, we use Table 2 (besides unaccounted for consumers and uncertainty related to the smart meters, data transmission and data handling, which were previously deemed valid) and Figure 6 to deduct what could be plausible causes for the observed discrepancies between the flow results from various data sources.

The negative diurnal residuals likely do not stem from **exfiltration** from the sewer system since $Q_{res,EU}$ exhibits a clear diurnal pattern in October while exfiltration is expected to vary according to the season. Furthermore, a net exfiltration is not expected to occur in February (where the groundwater level generally is higher) without also occurring in October and November. Neither do the negative residuals stem from **consumed water not discharged to the sewer system** since residential areas like 'EU' and 'WU' should discharge at least 86% of the used water to the sewer system, whereas the observed differences on average are 60 and 40% in EU (October) and WU (February), respectively. Furthermore, the smart meters in Elsinore are located inside the houses, which means that it is fair to assume that leakage on the consumers' side of the smart meters is minimal and can be disregarded. It is not likely that the positive diurnal residuals stem from **pumping of groundwater to the sewer system** because this will not lead to diurnal residuals. Furthermore, the Capital Region of Denmark, who stores records of such pumping sites, confirmed that no such pumping of groundwater to the sewer is taking place in the area of Elsinore where the in-sewer sensors are placed (Vormbak, F. 2019, Personal Communication, the Capital Region of Denmark). Even though it rained 1, 5 and 2 days before the three analysed time periods, respectively, the positive diurnal residuals do not stem from **rainwater harvesting** because this would have led to a systematic increase in the in-sewer flows within each sub-catchment across the three time periods. The positive diurnal residuals also do not stem from **snow melting**, since there was no snow in Elsinore in any of the three considered time periods. Neither **rain-induced infiltration** nor **infiltration from pumped groundwater** exhibit diurnal patterns. The deviation between the waterwork outlet and consumed water is more or less constant over the three time periods (Figure 2), but the gap between, for example, $Q_{obs,ED}$ and $Q_{sim,ED}$ varies heavily and is on average 14, 7 and 1.5 L/s in October, November and February, respectively, which indicates that the positive diurnal residuals are not likely to stem mainly from **infiltration from a leaking water distribution system**. This is supported by the fact that leakages in the water distribution system were actually recorded by the utility in EU in October and March (after the investigated periods), but Figure 6 shows signs of outflowing, not inflowing, water in October.

The constant positive residuals for WD are not likely to stem from **a constant infiltration from groundwater** since infiltration is expected to be larger in February than in October, whereas the observed offset is around 20 L/s in October and November and 10 L/s in February. No repair work was carried out for the sewer pipes in WD between November and February to explain this drop in infiltration. The pipe in which the WD flow sensor was located is heavily influenced by sedimentation, which could result in the constant offset from zero. The sediment was, however, not removed between November and February to explain the drop in the offset level, and it is therefore not expected that the offset is due to **sedimentation**. However, there is a chance that the amount of sediment may have been affected by the storm water flow in the sewer system.

Neither the positive nor negative residuals likely stem from a **wrong conceptualisation of the sewer system**,
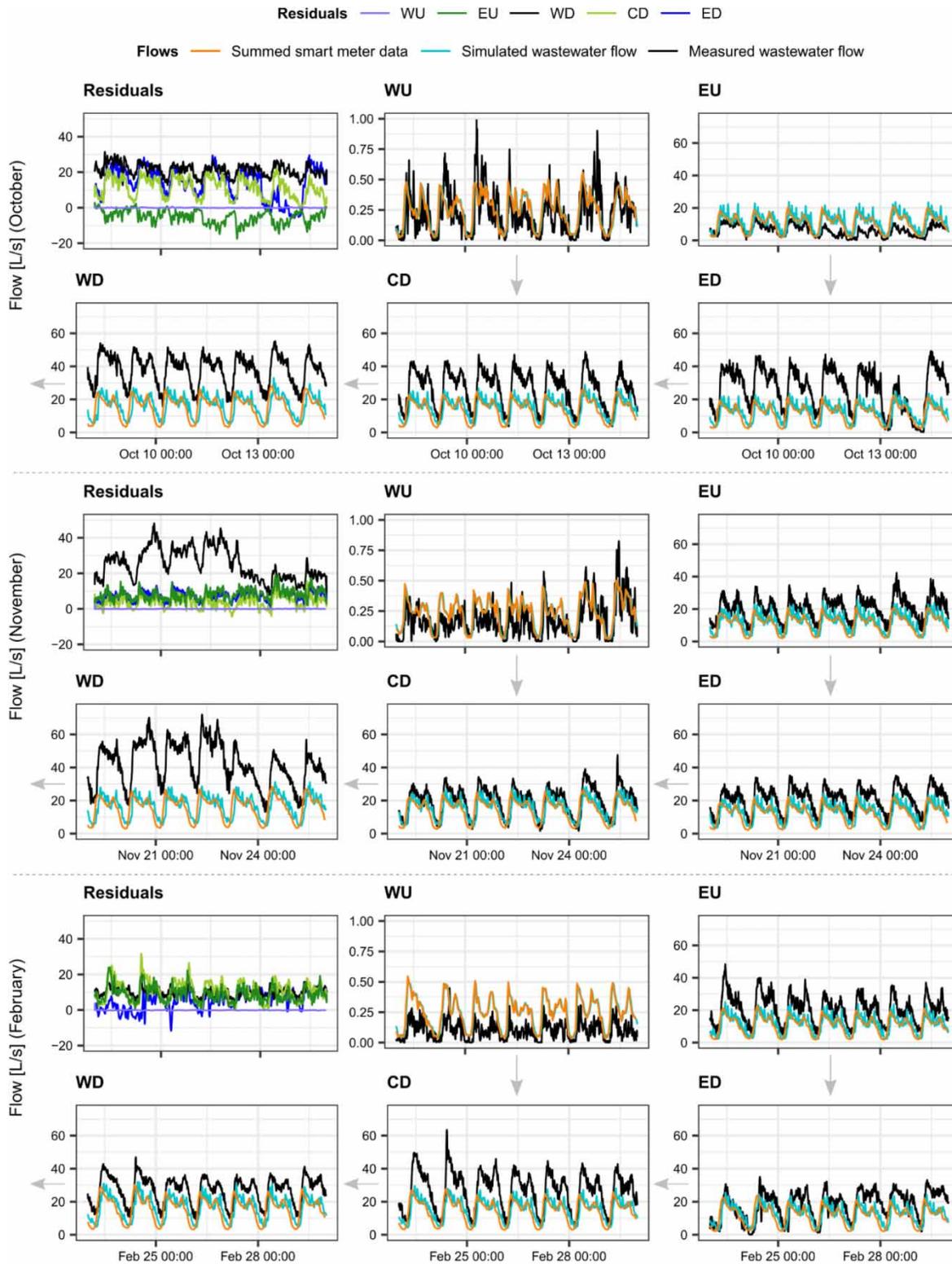
Figure 6 | Summed smart meter data (Q_{SM}) and simulated (Q_{sim}) and observed (Q_{obs}) sewage flows in October (top panel, Mon–Sat), November (middle panel, Mon–Sun) and February (lower panel, Sat–Fri) (Equation (1)). All data are in local time. For Q_{sim} and Q_{obs}, a 30 min moving average filter has been applied for better readability. Notice the different scales used for Residuals, WU and the four remaining plots. Due to the small flow in WU, the residuals appear constant in the Residuals plot.

since this would lead to similar residuals within each sub-catchment in all three time periods.

Erroneous in-sewer observations thus remain the only likely explanation for the deviations between Q_{obs} and Q_{sim}. The observed flow changes greatly from time period to time period; for example, $Q_{obs,CD}$ is much larger than $Q_{sim,CD}$ in October and February, but fits well in the intervening period in November. Looking at October, there is only one day where $Q_{obs,ED}$ and $Q_{sim,ED}$ match; however, this is according to Contractor 4 due to faulty velocity observations. Contrarily, the water consumption, as expected, more or less follows the same diurnal pattern throughout October, November and February. Additionally, the assumed erroneousness of in-sewer observations is supported by the fact that Q_{obs} out of WD, on average, is 21 and 25 L/s larger than Q_{sim} in October and November, respectively. Adding this amount on top of $Q_{sim,WWTP}$ would only exacerbate the difference to $Q_{obs,WWTP}$ (Figure 3). We therefore expect that the in-sewer sensor measurements are the main contribution to the deviation between simulated and observed sewage flows. These deviations are larger than the declared uncertainty of 12% (as supported by Figure 4) and often also larger than what, as stated earlier, is generally considered 'normal' deviations of up to 20%.

Figure 7 shows the relationship between the deviation between Q_{obs} and Q_{sim} and the time since the last calibration of the in-sewer sensors for the three time periods. No clear correlation exists, either within each time period or when comparing each location across the three time periods. This indicates that the differences are not due to slowly drifting sensors.

Since the sensors were installed as duplicate sensors, it might at first seem unlikely that these observations are as uncertain as indicated by this study. It is, however, worth noting that the duplicate sensors are not truly independent since they are installed at the same location and use the same sensing technology. This means that they are likely to produce the same systemic errors when exposed to the same conditions and thereby confirm each other's misconception of the flow. This could, for example, be the case if there is sedimentation or dislocation of pipes close to where a sensor pair is installed. Truly independent data sources are thus key to anomaly detection and data validation. To fully understand the anomalies detected in this study, it would be necessary to perform further independent measurements, for example:

- Independent and reliable flow observations at the five sensor locations to validate that the in-sewer flow observations indeed caused the deviation to the smart meter-simulated wastewater flow.
- CCTV (Closed Circuit Television) footage and groundwater-level observations to estimate the reason for the constant positive residuals for WD and whether these could stem, at least partly, from groundwater infiltration, sedimentation or dislocated pipes.

These issues could not be investigated in the present study as the measurement campaign in the urban drainage system was terminated in June 2019 but should be considered in future studies focusing directly on closing the urban water balance.
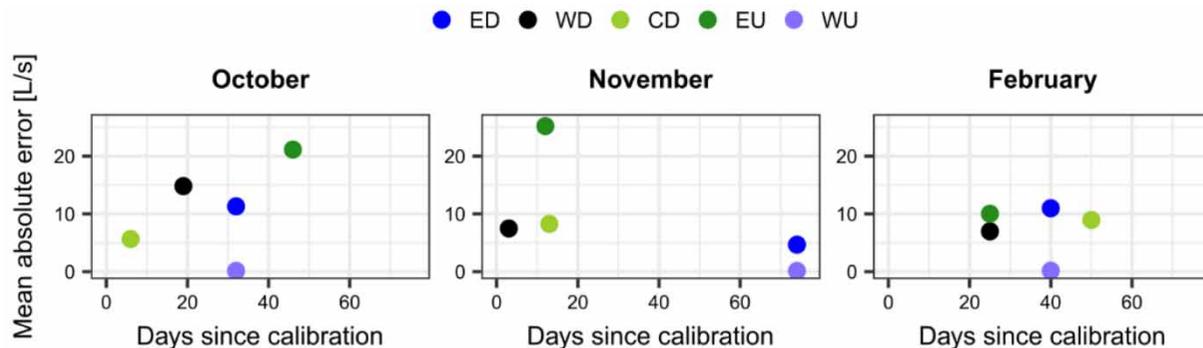

Figure 7 | Days since last calibration of in-sewer sensors compared with the difference between observed flow (Q_{obs}) and simulated flow (Q_{sim}) calculated as the mean absolute error for the three time periods.

Added value from using a hydrodynamic 1D model

A hydrodynamic 1D model poses an easy way of coupling smart meter data and in-sewer observations, since the model automatically ensures that only wastewater from the consumers upstream from an in-sewer sensor contributes to the flow at the sensor location. Furthermore, it is possible to include infiltration/exfiltration dynamics and thus make a better comparison between the simulated and observed data, which will not be possible when simply summing the smart meter data.

Figure 8 shows $Q_{obs,CD}$, $Q_{SM,CD}$ and $Q_{sim,CD}$ for 3 days in each of the considered time periods. The timing of the observed flow is considered reliable, despite the magnitude of the wastewater flow being off as a result of the likely erroneous in-sewer observations. It is clear that the routing of the smart meter data through the MIKE URBAN model catches the timing of the observed peaks and low flows much better than simply aggregating the smart meter data. The model can thus be used to capture the time dynamics of the system in dry weather in the five sensor locations without further calibration (however, the model may still need to be calibrated for rain situations). The dynamics shown in Figure 8 are representative of the entire dataset for all five flow observation locations. The difference in time between the summed smart meter data and the simulated flow will naturally be more pronounced the longer the water has travelled in the system and thus depends on the size of the catchment upstream of the flow observations.

Outlook

Sewer system analysis will likely never be the sole reason for installation of smart meters, which are typically installed for better operation and management of the water distribution system. The urban drainage sector can, however, still take advantage of this data, which is increasingly collected in the future anyway. This study shows that water supply smart meter data can potentially be used to estimate the wastewater flow. The fact that smart meters thus become multi-purposed will likely make their operation more robust. Furthermore, the robustness of the approach also stems from the sheer amount of smart meters, where a failure of a single meter will have much smaller consequences regarding the information level than the failure of a sewage flow sensor.

The use and comparison of data across the water supply, urban drainage and wastewater sectors and the subsequent anomaly analysis performed in this study is a tedious process due to the many aspects affecting both the water distribution and urban drainage networks. Furthermore, it is laborious to systematically gain access to and compare all relevant data sources. This was further complicated by the fact that not one single person could access all the models and data as they were managed and stored in different silos both within the utility and by different contractors. To get the most value out of data, we need to break down these silos in the future and aim for more open standards for data exchange. Many utilities are currently taking the first steps towards using data in new and more integrated ways. This study shows that digitalisation is not easy, and sometimes the data quality remains unknown until the data are used and compared with other data sources. More integrated data analysis and systematic uncertainty assessments are clearly needed to bring this field further ahead. The process of actually using data will provide important learnings regarding good practices within sensing and
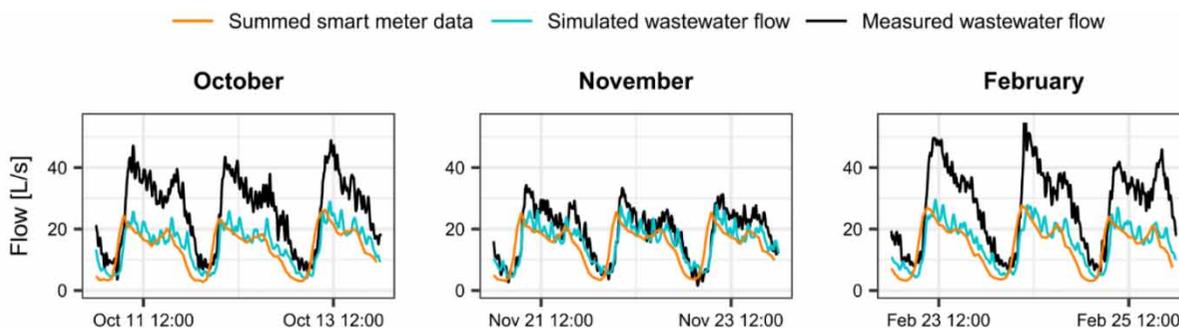

Figure 8 | Observed sewage flow (Q_{obs}) compared with simulated flow (Q_{sim}) and summed smart meter data (Q_{SM}) for 3 days in each of the three time periods for the 'CD' catchment.

data accessibility, and enable the utility and their contractors to further refine their work processes.

The coupling between smart meter data and urban drainage models can be done offline for post-analysis of data and system performance (as done in this study). If the model is trustworthy, this coupling could also be done, and the sewer model run in real time, to get an up-to-date picture of the system state. This would open up enhanced real-time data validation as well as entirely new ways of operating sewage infrastructure during dry weather, for example, for improved heat recovery from sewage systems or better sewer system control.

CONCLUSIONS

The current study aimed at using smart meter water consumption data to simulate the wastewater flow and to combine this information with in-sewer observations to detect system and data anomalies, such as infiltration, exfiltration and sensor errors.

Smart meter data were validated with data from other independent sources, including data from the waterworks' outflow, WWTP inflows and households' annual water consumption audits for a period prior to the installation of the smart meters. Subsequently, we illustrated the feasibility of using the smart meter data as input to a 1D hydrodynamic sewer model to simulate the wastewater component of the DWF. An estimate of the wastewater flow was thus obtained with a high spatial and temporal resolution. Even though there is still uncertainty related to this estimation, we believe that having this data-based approach to estimate the wastewater flow directly from its source (the water consumption) is an important step towards closing the water balance in urban drainage systems.

The main difference in the results from simply summing the smart meter data and using an urban drainage model is the effect of the routing time in the sewer system. If time dynamics are insignificant, a simple summation may be sufficient. Time dynamics are, however, important when comparing data on a fine time resolution as done in this study. Our comparisons and analysis of observed in-sewer DWF with both smart meter-simulated wastewater flow and WWTP inflow data suggest that the in-sewer flow observations used here may be subject to unrecognised uncertainties at times exceeding the expected 10–20%. These specific in-sewer observations are thus unsuitable as a realistic estimate of the DWF and thus also for decomposition of the DWF into wastewater flow and other input and output sources (such as infiltration and exfiltration). The in-sewer flow sensors were here installed in a very thorough way with two identical flow sensors at each location for validation, but it cannot be excluded that both sensors were influenced by the same error sources under the same conditions. This underlines the necessity of using truly independent data sources from different measuring techniques for future data quality assessments.

The main take-away messages from this study are:

1. Using smart meter data as input to hydrodynamic urban drainage models is feasible and promising as a means of estimating the wastewater flow directly from water consumption data. This is an important step towards closing the water balance in urban drainage systems.
2. Installing multiple in-sewer flow sensors of the same type at the same location is not enough to assess the data quality. Including independent data sources using different sensor types in future investigations is paramount.
3. It is a tedious task to compare and analyse data from multiple sources. However, the only way to start breaking down data silos and realise the real quality of collected data is to start using it in integrated studies as exemplified in the case study.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

We declare no conflict of interest.

DATA AVAILABILITY STATEMENT

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

REFERENCES

Abdel-Aal, M., Schellart, A., Kroll, S., Mohamed, M. & Tait, S. 2018 Modelling the potential for multi-location in-sewer heat recovery at a city scale under different seasonal scenarios. *Water Res.* **145**, 618–630.

Bertrand-Krajewski, J. L., Bardin, J. P., Mourad, M. & Béranger, Y. 2003 Accounting for sensor calibration, data validation, measurement and sampling uncertainties in monitoring urban drainage systems. *Water Sci. Technol.* **47**, 95–102.

Boyle, T., Giurco, D., Mukheibir, P., Liu, A., Moy, C., White, S. & Stewart, R. 2013 Intelligent metering for urban water: a review. *Water* **5**, 1052–1081.

Brito, R. S., Almeida, M. C. & Matos, J. S. 2017 Estimating flow data in urban drainage using partial least squares regression. *Urban Water J.* **14**, 467–474.

Butler, D. & Graham, N. J. D. 1995 Modeling dry weather wastewater flow in sewer networks. *J. Environ. Eng.* **121**, 161–173.

Cole, G. & Stewart, R. A. 2013 Smart meter enabled disaggregation of urban peak water demand: precursor to effective urban water planning. *Urban Water J.* **10** (3), 174–194.

DANVA 2017 *Water in Figures – DANVA Statistics and Benchmarking. Report by DANVA (the Danish Water and Wastewater Association)*. Available from: https://www.danva.dk/media/4662/water-in-figures_2017.pdf (accessed 26 September 2019).

DHI 2019 MIKE 1D API. Available from: docs.mikepoweredbydhi.com/engine_libraries/mike1d/mike1d_api/ (accessed 25 September 2019).

Djebbar, Y. & Kadota, P. T. 1998 Estimating sanitary flow using neural networks. *Water Sci. Technol.* **38**, 215–222.

Elías-Maxil, J. A., van der Hoek, J. P., Hofman, J. & Rietveld, L. 2014 A bottom-up approach to estimate dry weather flow in minor sewer networks. *Water Sci. Technol.* **69**, 1059–1066.

Elsinore Municipality 2019 *Vandforsyningsplan 2019–2030. Helsingør Kommune. Center for by, land og vand, natur og miljø. June 2019*. Available from: https://www.fh.dk/files/media/dokumenter/Om%20os/Lovgivning%20og%20planer/vandforsyningsplan_19-30_helsingoer_kommune_juni_2019_132374-19_v1.pdf (accessed 15 October 2019).

Hager, W. 1994 *Wastewater hydraulics - Theory and practice*. Springer-Verlag, Berlin.

HOFOR 2016 *Spar vand – spar penge. Så skåner du også miljøet*. Greater Copenhagen Utility (HOFOR), København, Denmark.

Kamstrup 2019 Data Sheet. Multical® 21 and flowIQ® 2101. Available from: https://products.kamstrup.com/ajax/downloadFile.php?uid=515d4ab700278&display=1 (accessed 29 August 2019).

Kirstein, J. K., Høgh, K., Rygaard, M. & Borup, M. 2019 A semi-automated approach to validation and error diagnostics of water network data. *Urban Water J.* **16**, 1–10.

Kirstein, J. K., Høgh, K., Rygaard, M. & Borup, M. 2021 A case study on the effect of smart meter sampling intervals and gap-filling approaches on water distribution network simulations. *J. Hydroinf.* **23** (1), 66–75.

Larrarte, F., Bardiaux, J. B., Battagliac, P. & Joannis, C. 2008 Flow measurement and instrumentation acoustic doppler flow-meters. A proposal to characterize their technical parameters. *Flow Measurement and Instrumentation* **19**, 261–267.

Lund, N. S. V., Falk, A. K. V., Borup, M., Madsen, H. & Mikkelsen, P. S. 2018 Model predictive control of urban drainage systems: a review and perspective towards smart real-time water management. *Crit. Rev. Environ. Sci. Technol.* **48** (3), 279–339.

Métadier, M. & Bertrand-Krajewski, J.-L. 2011 Assessing dry weather flow contribution in TSS and COD storm events loads in combined sewer systems. *Water Sci. Technol.* **63**, 2983–2991.

Monks, I., Stewart, R. A., Sahin, O. & Keller, R. 2019 Revealing unreported benefits of digital water metering: literature review and expert opinions. *Water* **11** (838), 1–32.

Nguyen, K. A., Stewart, R. A., Zhang, H., Sahin, O. & Siriwardene, N. 2018 Re-engineering traditional urban water management practices with smart metering and informatics. *Environ. Modell. Softw.* **101**, 256–267.

Plósz, B. G., Reid, M. J., Borup, M., Langford, K. H. & Thomas, K. V. 2013 Biotransformation kinetics and sorption of cocaine and its metabolites and the factors influencing their estimation in wastewater. *Water Res.* **47** (7), 2129–2140.

Schilperoort, R. P. S., Dirksen, J., Langeveld, J. G. & Clemens, F. H. L. R. 2012 Assessing characteristic time and space scales of in-sewer processes by analysis of one year of continuous in-sewer monitoring data. *Water Sci. Technol.* **66**, 1614–1620.

Siemens 2020 SITRANS FM MAG 3100. Available from: https://new.siemens.com/global/en/products/automation/process-instrumentation/flow-measurement/electromagnetic/sitrans-f-m-mag-3100.html (accessed 20 May 2020).

Warmink, J. J., Janssen, J. A. E. B., Booij, M. J. & Krol, M. S. 2010 Identification and classification of uncertainties in the application of environmental models. *Environ. Modell. Softw.* **25**, 1518–1527.

Zhang, Q., Zheng, F., Jia, Y., Savic, D. & Kapelan, Z. 2021 Real-time foul sewer hydraulic modelling driven by water consumption data from water distribution systems. *Water Res.* **188**, 116544.