



## Using multi-event hydrologic and hydraulic signatures from water level sensors to diagnose locations of uncertainty in integrated urban drainage models used in living digital twins

Pedersen, A. N.; Pedersen, J. W.; Borup, M.; Brink-Kjær, A.; Christiansen, L. E.; Mikkelsen, P. S.

*Published in:*  
Water Science and Technology

*Link to article, DOI:*  
[10.2166/wst.2022.059](https://doi.org/10.2166/wst.2022.059)

*Publication date:*  
2022

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Pedersen, A. N., Pedersen, J. W., Borup, M., Brink-Kjær, A., Christiansen, L. E., & Mikkelsen, P. S. (2022). Using multi-event hydrologic and hydraulic signatures from water level sensors to diagnose locations of uncertainty in integrated urban drainage models used in living digital twins. *Water Science and Technology*, 85(6), 1981-1998. <https://doi.org/10.2166/wst.2022.059>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Using multi-event hydrologic and hydraulic signatures from water level sensors to diagnose locations of uncertainty in integrated urban drainage models used in living digital twins

A. N. Pedersen <sup>a,b,\*</sup>, J. W. Pedersen <sup>b,c</sup>, M. Borup <sup>b,d</sup>, A. Brink-Kjær <sup>a</sup>, L. E. Christiansen <sup>e</sup> and P. S. Mikkelsen <sup>d</sup>

<sup>a</sup>VCS Denmark, Vandværksvej 7, 5000 Odense C, Denmark

<sup>b</sup>DTU Environment, Technical University of Denmark, Bygningstorvet, Bygning 115, 2800 Kgs. Lyngby, Denmark

<sup>c</sup>Present address: Danish Meteorological Institute, Lyngbyvej 100, 2100 Kbh Ø, Denmark

<sup>d</sup>Present address: Krüger A/S, Veolia Water Technologies, 2860 Søborg, Denmark

<sup>e</sup>DTU Compute, Technical University of Denmark, Richard Petersens Plads, bygning 324, 2800 Kgs. Lyngby, Denmark

\*Corresponding author. E-mail: anp@vandcenter.dk

 ANP, 0000-0002-8254-4196; JWP, 0000-0003-4839-2069; MB, 0000-0002-6531-9464; LEC, 0000-0001-5019-1931; PSM, 0000-0003-3799-0493

### ABSTRACT

Digital twins of urban drainage systems require simulation models that can adequately replicate the physical system. All models have their limitations, and it is important to investigate when and where simulation results are acceptable and to communicate the level of performance transparently to end users. This paper first defines a classification of four possible 'locations of uncertainty' in integrated urban drainage models. It then develops a structured framework for identifying and diagnosing various types of errors. This framework compares model outputs with in-sewer water level observations based on hydrologic and hydraulic signatures. The approach is applied on a real case study in Odense, Denmark, with examples from three different system sites: a typical manhole, a small flushing chamber, and an internal overflow structure. This allows diagnosing different model errors ranging from issues in the underlying asset database and missing hydrologic processes to limitations in the model software implementation. Structured use of signatures is promising for continuous, iterative improvements of integrated urban drainage models. It also provides a transparent way to communicate the level of model adequacy to end users.

**Key words:** digital twins, error diagnosis, hydraulic models, rain uncertainty, time series, validation

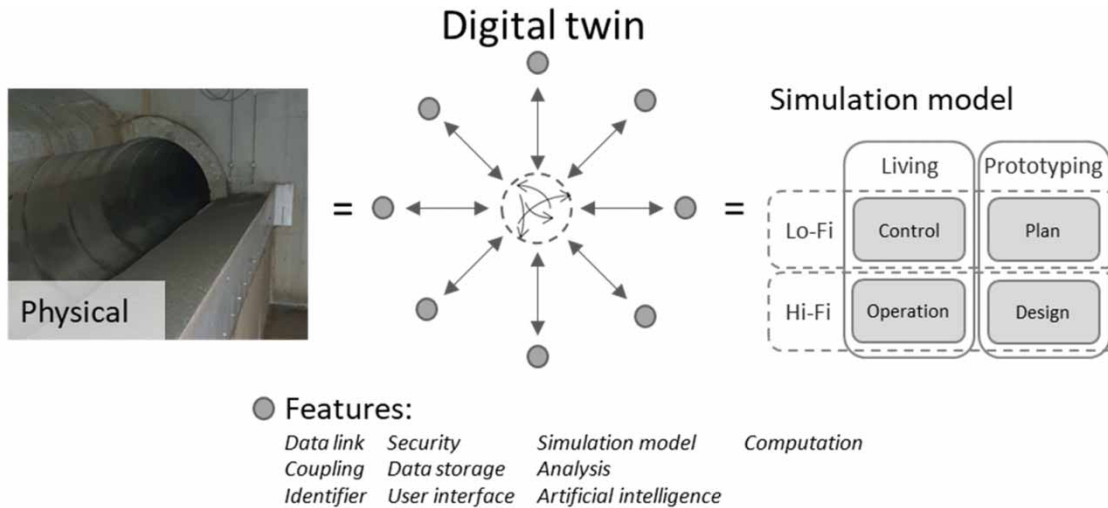
### HIGHLIGHTS

- Transparency of the performance of simulation models in digital twins is needed.
- Observation data can indicate that reality is different from what is perceived as true.
- Signatures are strong tools to diagnose model errors.

### INTRODUCTION

Stimulated by the emergence of increasing amounts of monitoring data the expectations to digital twins (DTs) in the water sector are high, as they are anticipated to provide improved insights and overview of the infrastructure systems for water distribution, drainage, and treatment (Fuertes *et al.* 2020; Therrien *et al.* 2020; Pedersen *et al.* 2021a; Valverde-Pérez *et al.* 2021). Multiple companies currently explore the DT concept which is also being widely debated in academia, and Pedersen *et al.* (2021a) suggested, based on a thorough literature review across engineering disciplines, a definition based on Autiosalo *et al.* (2020), that is used here as a conceptual frame. A DT for an urban drainage system (Figure 1) is a systematic virtual representation of the elements and dynamics of the physical system, organized in a star structure (a representation of the Internet of Things (IoT)) with a set of features connected by data links. Coupled to the physical system, simulation models are among the most important features, and (at least) four different model categories are distinguished. Two of these are prototyping models used for planning or design purposes, and two are living models used for control or operation purposes, living here referring to the coupling of close to real-time observations from an ever-changing physical twin (which may

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).



**Figure 1** | Illustration of the concept of a digital twin (DT). Elaborated from Pedersen *et al.* (2021a) and Autiosalo *et al.* (2020). ○ refers to the feature data link, which is the center of a star structure surrounded by other features. Living and prototyping DTs can include both high fidelity (Hi-Fi) models and simplified low-fidelity (Lo-Fi) models.

change over time) with a simulation model, through a data link connecting the two. Whereas control models are often simple, conceptual, and even purely data driven (e.g., Lund *et al.* 2018; Stentoft *et al.* 2021; van der Werf *et al.* 2021), operation models as defined here are detailed, high-fidelity representations of reality (thus called Hi-Fi models). Bach *et al.* (2014) refer to these as ‘integrated urban drainage models’, which is a ‘classical’ industry-standard type of model first developed in the 1980s (e.g., Huber & Dickinson 1992) for use in simulating entire urban drainage systems, from a raindrop hits the ground until it leaves the system via outlets and overflow structures or due to sewer surcharge or flooding. A utility’s motivation for running an operation model and comparing with monitoring data, e.g., once per day, is to continuously observe, understand and document the performance of the urban drainage system as well as the monitoring system and through this gradually improve the knowledge base for massive future investments in network maintenance and upgrades (Pedersen *et al.* 2021a).

Several software products used in the urban drainage profession worldwide include this type of model, e.g., SWMM, InfoWorks, and Mike Urban. In its representation of the pipe network, such a models’ physical realism often means that its users regard it with a high degree of confidence. The detailed and highly distributed nature of these models however also means that there are many places where model errors can creep in. It is thus important that users can assess where and why model errors exist so they can be remedied. Experts have earlier often done this manually in practice, but with the emergence of the IoT technology many new sensors in urban drainage systems are expected. As a result, manual evaluation of the models will be unrealistic. A more automated, systematic approach to model assessment and, if possible, improvements of simulation models is needed. This requires the choice and/or engineering of multiple diagnostic metrics that target various processes and their spatial variation, which within the field of urban drainage includes runoff from pervious and impervious areas, storm-water conveyance, infiltration-inflow, storage, overflow, and others.

Gupta *et al.* (2008) proposed using hydrologic signatures to increase the power of diagnostic model evaluation. Hydrologic signatures are metrics that quantify specific properties of a hydrologic time series, observed or modeled (McMillan *et al.* 2017; Mizukami *et al.* 2019). The method has been applied in general hydrology for many years (Westerberg & McMillan 2015; McMillan 2020a), with flow measurements as one of the typically available observations. In urban hydrology, the method has only been used in a study of improving models of stormwater infiltration measures when simulating catchment flow response (Hamel & Fletcher 2014). The hydrologic processes of urban areas are different from natural catchments and strongly affected by the hydraulic infrastructure. Additionally, water level sensors are more robust than flow sensors in harsh sewer environments and are thus more prevalent in drainage systems. Thus, urban drainage professionals need to provide the domain knowledge required to define useful signatures in this space.

An automated model assessment framework based on a range of hydrologic and hydraulic signatures will produce an enormous amount of metrics once implemented at a water utility with several hundred sensors installed in the local urban

drainage system. There is thus a need to identify common model errors and classify them in a structured, ordered way. Such a classification of model performance at the whole-system level would assist iterative model improvements, allow assessing the uncertainties in the model, and help understand for which processes and management objectives a given model can provide trustworthy estimates. There has been a wide focus in the urban draining community on ameliorating model errors through parameter calibration techniques targeting the hydrological (rainfall-runoff) part of integrated urban drainage models (e.g., [Deletic \*et al.\* 2012](#); [Breinholt \*et al.\* 2013](#); [Tscheikner-Gratl \*et al.\* 2016](#); [Vonach \*et al.\* 2019](#)). However, there has been relatively little focus on technical and structural errors that cause large problems in practice, such as faulty asset data, missing physical processes, and others. This paper aims to present a classification scheme that embraces all errors in integrated urban drainage models. It does so by combining the ‘uncertainty frameworks’ of [Walker \*et al.\* \(2003\)](#) and [Gupta \*et al.\* \(2012\)](#), in a manner somewhat similar to what was previously done by [Del Giudice \*et al.\* \(2015\)](#). However, here we additionally include the uncertainty ‘locations’ defined by [Walker \*et al.\* \(2003\)](#), i.e., ‘context,’ ‘input,’ ‘model structure,’ and ‘parameter’. [Deletic \*et al.\* \(2012\)](#) also looked into these uncertainties with a special interest in calibration methods and acknowledged the difficulty in addressing the model structural uncertainties we aim to detect.

The objective of this paper is to develop a framework tailored for practitioners in water utilities for use in their iterative improvements to the hydraulic network representation in existing integrated urban drainage models often used in DTs. To do so, the following three things are considered: Firstly, a scheme for classifying various types of errors and uncertainties in integrated urban drainage models is presented. Then the paper introduces the concept of hydrologic and hydraulic signatures to the urban drainage community and highlights their use as the backbone of a systematic approach to error diagnostics. Finally, the framework is illustrated through application to a real case study with up to 10 years of monitoring data. Its utility is illustrated by identifying three different classes of errors in the available hydrodynamic model through multi-event signature comparisons.

## CONCEPTUAL FRAMEWORK

### Urban drainage modeling and surrounding states

Integrated urban drainage models consist of two distinct modules: a hydrologic surface runoff module that converts precipitation data into inflow to the pipe system for each sub-catchment in the system and a hydraulic model that distributes water throughout the entire pipe system by solving the full St. Venant equations. Additional model forcing components can be defined, such as wastewater inflow, pumped flows, and infiltration-inflow. The surface runoff module is lumped-conceptual on the sub-catchment scale but distributed at the entire urban drainage system scale ([Hansen \*et al.\* 2014](#)). Several types of surface modules are available. The hydraulic model is physically distributed, and as for the surface module, several versions are also available. This urban drainage modeling approach is sometimes referred to as ‘detailed’ or ‘high-fidelity’ (Hi-Fi) modeling, referring to the detailed manner based on physical asset data in which pipe-flow is described. However, the pipe-flow module also includes hydraulic structures in which the hydrodynamics are described in a lumped, conceptual manner, for example, manholes, overflow weirs, and pumps.

In Danish engineering practice, the ‘classical’ urban drainage modeling approach is generally expected to simulate small- and medium-sized rain events leading to increased flow and occasionally overflow and surcharge well. This is because routine monitoring campaigns typically capture these events, where mostly paved surfaces contribute to rainfall-runoff, and therefore model comparisons are often made with these events. Furthermore, with the increasing exploitation of the models to simulate other operational domains, it is expected that the performance may be useable for dry-weather situations. However, there may be big challenges with simulating the wastewater flow in dry weather (infiltration-inflow may increase the flow seasonally or after long wet-periods, which is usually not included in the model) and with simulating flows during storm events where unpaved areas start contributing to the runoff, a process not well described in state-of-the-art models used in Danish engineering practice. In addition, weather patterns may influence the simulation quality; spatially wide-spread rainfall during frontal weather systems may, for example, be well represented by a single rain gauge located in or close to the catchment in question. In contrast, small-scale convective rainstorms causing local flooding may be missed by a single rain gauge. These processes and phenomena are either not included or not well represented in ‘classical’ urban drainage models, and there is a need to qualify under which conditions this is of concern, so that we can avoid attempting to find solutions to uncertainty originating from outside the model boundaries through calibration of parameters inside the model.

We use the term ‘surrounding states’ for any relevant information outside the model boundary that can help identify relationships and processes relevant to the system but are currently not included in the model. These may be indicators that potentially explain the uncertainty of the modeling approach or why the model fit is expected to differ for different operational conditions. For example, precipitation at low temperature may not give high water levels in the sewers, as precipitation will remain on the surface as snow and only later reach the sewers. High soil moisture may also lead to higher flows due to runoff from pervious surfaces and infiltration-inflow. Some rainfall types with limited spatial variability may furthermore be difficult to capture with only a few rain gauges. The impact of the surrounding states can be site- and model-specific, and therefore the ones found to have an impact in this paper may not have an impact in other models.

**Classification of uncertainties**

Three dimensions can characterize uncertainty: location, level, and nature (Walker et al. 2003; Warmink et al. 2010). The *location* is where the model’s uncertainty manifests itself (e.g., context, input, model structure, parameters). The *level* of uncertainty ranges from statistical uncertainty to scenario uncertainty, qualitative uncertainty, and ignorance, suggesting that quantification is not easy for all kinds of uncertainty. Finally, the *nature* of the uncertainty distinguishes epistemic uncertainty (due to imperfect knowledge), natural variability (e.g., rainfall variability), and ambiguity (when there are multiple parallel knowledge frames). Gupta et al. (2012) identifies three different locations of model uncertainty: (a) conceptual model, which incorporates the physical and process structure of the system; (b) the mathematical model, which incorporates the spatial variability and equation structure; and (c) the computational model relating to the model technical uncertainty. To identify and classify the sources of error in a model, we combine the two uncertainty frameworks (Table 1). We add temporal variability to the mathematical model location, which is a large uncertainty in urban drainage system modeling. The different uncertainties are in Table 1 exemplified by a non-exhaustive list of examples. For example, input uncertainty may be related to both rainfall data (external forcing) and the elevation of a crest level (a system attribute).

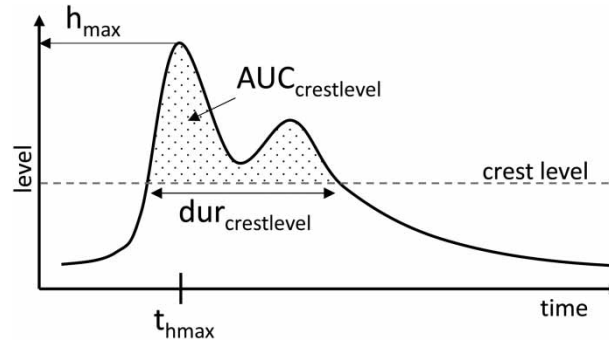
**Signatures for urban drainage systems**

Signatures are quantitative measures derived from parts of time series to analyze patterns – such as peak level, duration of the level above a crest level, the area under the level curve, time of the peak, and others (Figure 2). They may provide insights into the underlying processes (in the case of observed time series) or in the model used to simulate the processes (in the case of simulated time series). Several studies provide examples of signatures (McMillan 2020a; Gnann et al. 2021). Analyzing each signature can help diagnose the likely source of a discrepancy (error) between model output and observations. In addition, peak level can give insight into local physical attributes. In hydrology, they started identifying various process-based signatures (McMillan 2020b), which needs to be done as well in the urban drainage field.

**Table 1** | Classification of errors related to ‘location uncertainty’ for urban drainage models, with examples. Based on Walker et al. (2003) and Gupta et al. (2012)

Locations		Example descriptions of uncertainty in each location
<b>Context</b>	Context	Are the boundaries within the model objective?
<b>Input</b>	External forcings	Incorrect model of rain or observations forcing the model
	System attributes	Asset data are wrong, e.g., pipe diameter, invert levels, weir levels, regulation of valves, pump curves, and others
<b>Model structure</b>	<b>Conceptual model</b>	
	Physical attributes	Description of the manhole structure
	Process	Description of the soil moisture or infiltration models
	<b>Mathematical model</b>	
	Spatial and temporal variability	Resolution of external input too coarse to represent the phenomenon A mismatch between lumped runoff models and detailed pipe-flow models
	Equation	Do we use the right equation? For example, kinematic or dynamic flow equations in the pipe-flow module Do we have the right equations?
	<b>Computational model</b>	
	Computational	Instability in the models or errors in soft- and hardware
<b>Parameter</b>	Parameter	Manning number, imperviousness percentage





**Figure 2** | Simple illustration of how signatures may be used to characterize different parts of a time series.

This study uses signatures for a specific operational domain in the urban drainage system, namely rain-induced events. We look both at signatures characterizing hydrologic processes in the rainfall-runoff module and signatures characterizing hydraulic processes in the hydraulic module.

### Graphic interpretation of multi-event signatures

Many hydrology studies use signatures to extract statistical properties of the system to be replicated in a lumped model, and the model is then assumed to be sufficient. Properties from a gauged area can even be transferred to similar ungauged catchment areas to model these (Hrachowitz *et al.* 2013). This approach can also be performed in urban hydrology. However, urban hydrology, to a larger degree than general hydrology studies, is influenced by physical assets and direct runoff from paved areas, direct comparison of time series event by event, model vs. observation, is often required by regulators and utility companies.

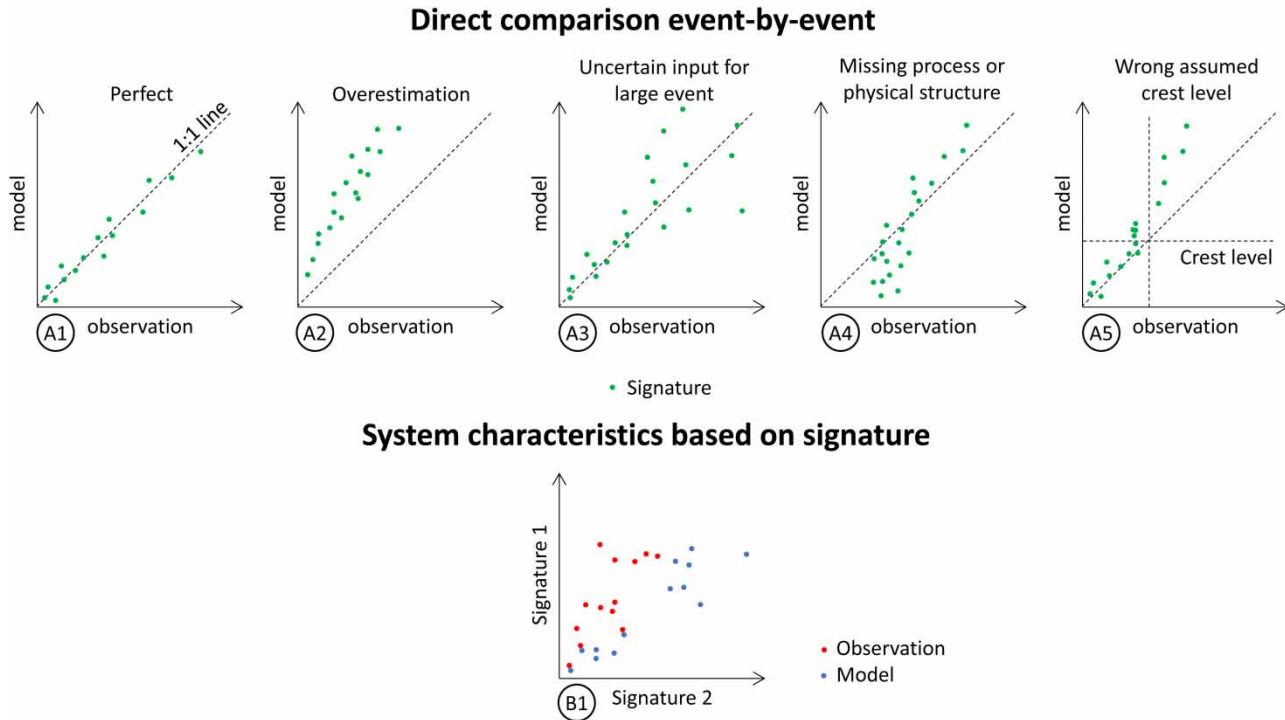
Signatures can be calculated for several physical sites across an urban drainage system, focusing both on characteristics of individual events and on multiple events to investigate persistence and trends in the error structure at the particular physical site. Errors here refer to the discrepancy between the signature from the model event to the observed event, and diagnostics should potentially support identifying and classifying the source of the errors based on these comparisons. Figure 3(a) illustrates different examples of multi-event graphs. Each dot represents the resulting value of the signature for one event, with observed values on the horizontal axis and modeled values on the vertical axis. The identity line (1:1 line) indicates a perfect model fit if the dots are present within an acceptable uncertainty (Figure 3(a1)). Dots systematically above the identity line indicate an overestimation in the model of the effect of the process behind the signature (Figure 3(a2)). Scattered dots for high levels may indicate a high input uncertainty for large events (Figure 3(a3)), and a bend in the dots may indicate a missing process in the model (Figure 3(a4)). With water level data as a signature unit, one could, for example, find the crest level, the maximum level of the pass-forward pipe, and descriptions of other physical attributes in the model that must match reality (Figure 3(a5)). Different characteristics can also be extracted by comparing different signatures (Figure 3(b1)) and looking for clusters and variance in the displayed pattern.

Results presented in Figure 3 make it possible to classify the structural error/uncertainty according to Table 1. An assessment can then be made whether the cause of the error can be rectified or if the model is accepted by an informed evaluation of the model performance for the specific objective at hand. If required, further development of the models could be needed.

## DATA AND METHODS

### Case study – system, data, and model

The case area is located in the outskirts of Odense, Denmark (Figure 4), which is a relatively flat area. We use observation data from level sensors in hydraulic structures of three sites in upstream branches of the urban drainage system (Figures 4 and 5). Site 1 is a manhole in a normal flow-through pipe with a 1 m diameter. Site 2 is a flushing chamber upstream from a storage pipe (trunk sewer for combined sewer overflow (CSO) storage); the chamber fills with water from an internal overflow weir located higher in the chamber. A valve can be opened quickly to flush the volume pipe when it is empty. Finally, site 3 is an internal overflow structure located in a large manhole (2 m diameter); the inlet pipe (0.4 m diameter) is constrained to a pass-forward pipe of just 0.16 m diameter with surplus water overflowing to another pipe. Two of the sensor sites (sites 1 and 2)

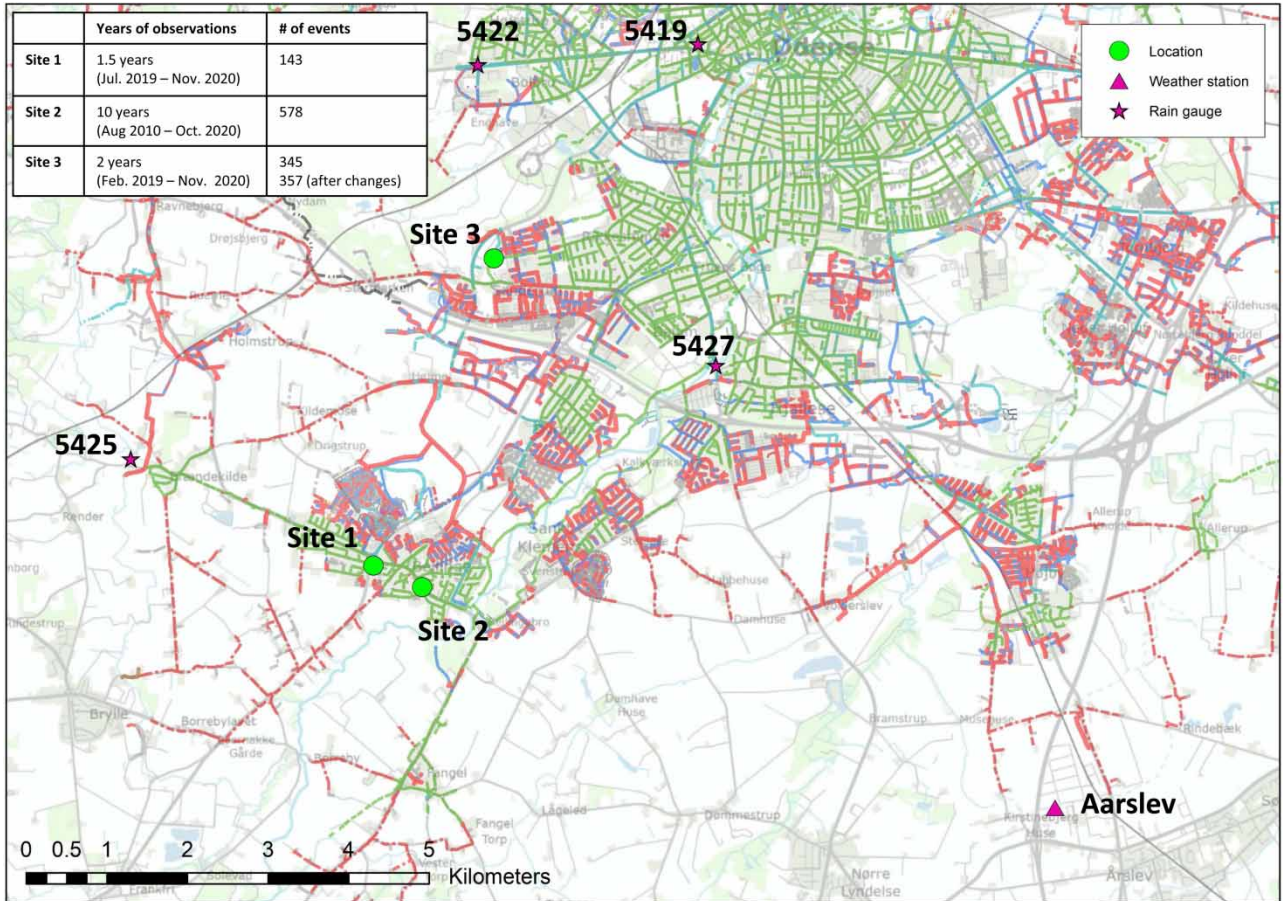


**Figure 3** | Examples of sources of uncertainty identified by visual error diagnosis. The top figures (A) consist of direct comparisons between modeled and measured (observed) signatures, whereas the bottom figure (B) illustrates a comparison of two different signatures for observed and modeled values.

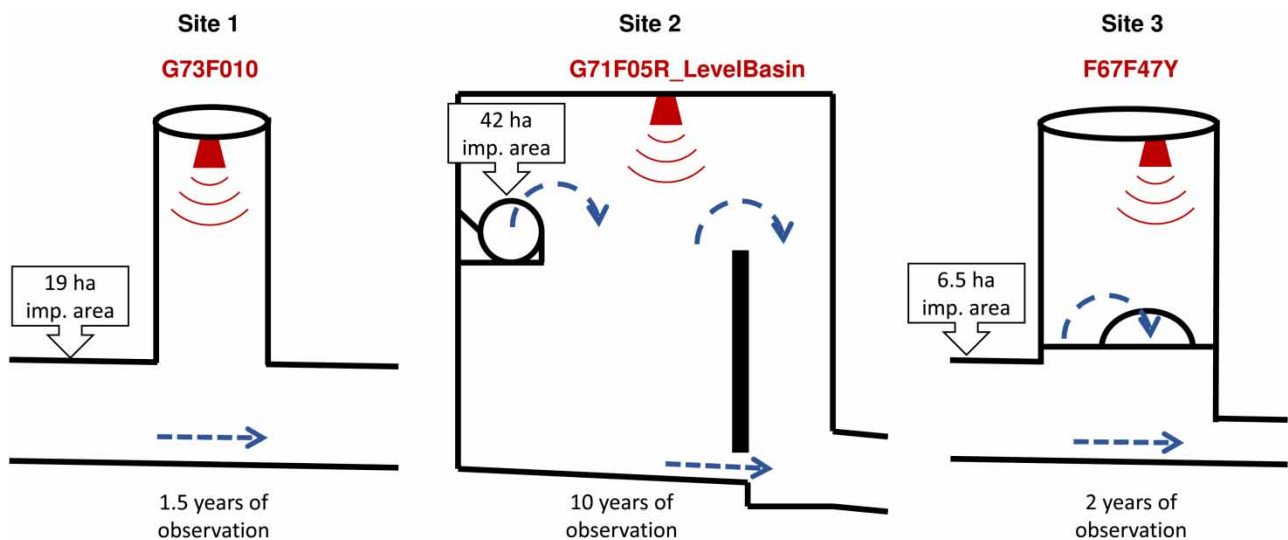
are described in the Bellinge open dataset (Pedersen *et al.* 2021b), whereas the third sensor site (site 3) is described in the Supplementary Material (SM).

We used one-minute model input and level observations for up to 10 years with the number of rain-induced events (events observed in the water level observations that are induced by rain events) calculated as indicated in Figure 4 for each site (143 events at site 1, 578 events at site 2, 345–357 events at site 3). Observation data were cleaned for outliers as proposed in Pedersen *et al.* (2021b), but anomalies naturally occurring in urban drainage systems were not considered. Weather data (solar radiation, wind velocity, humidity and temperature) were extracted from the Danish Meteorological Institute's (DMI) open dataset for a weather station (Aarslev) 7 km away (DMI 2020). In addition, rainfall data were obtained from nearby rain gauge stations 5419, 5422, 5425 and 5427 (Jørgensen *et al.* 1998; DMI (Danish Meteorological Institute) & IDA (The Danish Society of Engineers) 2020).

The applied simulation model is an integrated urban drainage model (see the section *Urban drainage modeling and surrounding states*) developed in the Mike Urban software by VCS Denmark. The model setup is described in the Bellinge open dataset (Pedersen *et al.* 2021b). The surface runoff module is based on the time-area principle, and the pipe-flow module is the MOUSE hydraulic engine. The utility company and its consultants estimated the imperviousness of sub-catchments from satellite data using spectral analysis and standardized imperviousness percentages for individual area classes (LNHwater 2017). The applied runoff model is 'Model A', which considers only impervious areas and is applied due to historical reasons in Danish engineering practice. The pipe-flow module has a fixed calculation time step of 5 seconds. Rainfall input is from two nearby rain gauges, '5425 Brændekilde' and '5427 Dalum'. Wastewater input is included with a daily mean calculated based on the annual water consumption and a mean daily pattern with hourly time steps. Infiltration-inflow is not included in the model even though this is technically possible in the applied software using a built-in RDII-module; this is because of the utility's focus on improving the knowledge base on the physical assets rather than calibration of parameters in the lumped-conceptual model components.



**Figure 4** | Map of the case area with an indication of sites of water level sensors (green dots) as well as the surrounding rain gauges (pink stars) and DMI’s weather station (pink triangle). The urban drainage system is composed of combined sewers (green outline) as well as separate sewers for stormwater (blue outline) and wastewater (red outline). The length of the periods with observation data at the three sites and the number of extracted rain-induced events are indicated in the table.



**Figure 5** | Illustrations of the hydraulic structures at the three sites with sensors installed.



### Identification of rain-induced events

A time-varying, daily water level threshold (from now on referred to as a 'rain threshold') was calculated as the sum of the infiltration-inflow level and the extent of the daily fluctuations in dry weather (Figure 6). Rain-induced events were identified when the observed water level exceeded the rain threshold. The start of events was defined as when the rising limbs started sloping upwards, events had to be at least three time steps long (3 minutes), and unique events were separated by at least 60 minutes. The equations to estimate the different components are defined below.

### Infiltration-inflow level

The infiltration-inflow level in the system was calculated by applying a recursive, exponentially weighted moving average function:

$$y_0 = x_0; \quad y_t = (1 - \alpha)y_{t-1} + \alpha x_t \quad (1)$$

where  $x_t$  is the observed daily minimum level at time (day)  $t$ ,  $y_t$  and  $y_{t-1}$  are the calculated (exponentially weighted) daily minimum value at day  $t$  and  $t-1$ , and  $\alpha = 2/(\text{span}+1)$ , where  $\text{span} \geq 1$  is a user-defined parameter defining how important the current observation is. With a span of 3 days,  $\alpha=0.5$ , which means that the exponentially weighted value calculated today depends 50% on today's observed value, 25% on yesterday's observed value, and 12.5% on the observed value two days ago and so on.

### DWF-height

The height of the dry-weather flow (DWF) indicates the range that the level will change due to variation in water consumption in individual households during a normal dry-weather day:

$$\Delta d = d_{\max} - d_{\min} \quad (2)$$

This gives a variety of values as some days reach near maximum level when rain occurs. In Denmark, it rains approximately every third day. Therefore, it was assumed that the 0.5 quantile of the data would correspond to a standard dry-weather day. The daily DWF-height was calculated based on the previous 365 days.

### Rain threshold

The rain threshold for a runoff event is calculated.

$$\text{Rain threshold} = \text{infiltration-inflow level} + \text{DWF-height} \quad (3)$$

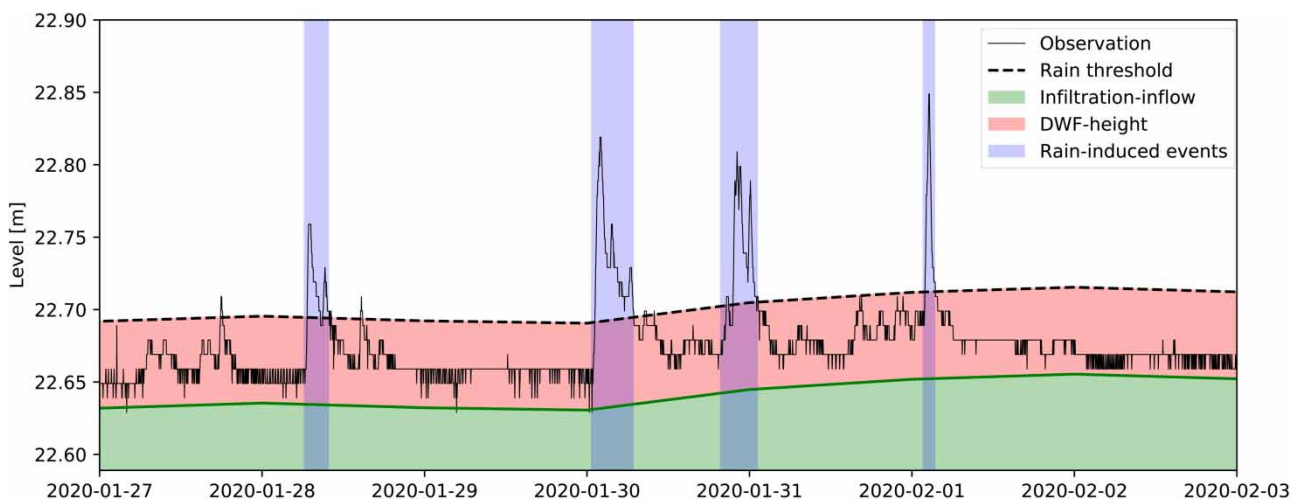


Figure 6 | Illustration of the time-varying rain threshold and identification of rain-induced events.

The rain threshold was calculated with one value each day, and therefore very small events that occur during low DWF may not be detected with this approach.

Finally, we defined the events used for analysis in this study as the union set of events detected in both simulated and observed time series (see the number of events for each site in Figure 4).

### Quantification of signatures

A range of signatures was calculated from the time series of the water level sensors in the three sites (Table 2), typically expressing max/min values, or integrating or differentiating values over time. For each event, the peak level,  $h_{\max}$ , was calculated together with the peak time occurring,  $t_{h_{\max}}$ . The duration of periods with water above the crest level,  $dur_{\text{crestlevel}}$ , of a weir can be used as a signature to estimate the overflow volume. The 'Area Under Curve' (AUC) calculates a 'surrogate volume' for an event from a reference level (similar to the area under a flow hydrograph, but with a different unit). This signature attempts to estimate how much water is passing the manhole. However, attention must be drawn to the fact that different levels in the manhole will give different AUC. It can therefore not be transferred directly to an estimate of volume. The low depths in a manhole may furthermore not contain the same volume as when the levels are higher in a manhole. The level rate of change unit is a very simple way of analyzing the entire time series and is not applied to single events. It calculates the change in water level between two time steps.

### Surrounding states

This paper focuses on quantifying two surrounding states: soil moisture, and spatial variability of the rainfall input.

#### Soil moisture

We used a daily model for drought index prepared by DMI (Scharling & Vilic 2009) that takes five input parameters: solar radiation, wind velocity, humidity and temperature from a local national weather station (DMI 2020), as well as rain input from a nearby local rain gauge. Soil moisture was calculated as a single value per day, and to avoid influence from rain on the day of interest, the value from the day before is applied.

#### Conceptual model for rain input uncertainty

Non-uniform rainfall occurs for convective patterns, where the spatial rainfall extent may be smaller than the catchment area a rain gauge may not capture the event well. More uniform precipitation types are, for example, frontal rain events, where the intensity is roughly the same for a larger area and the uncertainty of the rain input to the model thus is lower.

Rain gauges within a certain catchment distance are assumed to affect the rain uncertainty; 5 km is the best estimate of this distance based on local spatial rainfall statistics in Denmark (Gregersen *et al.* 2013). One way to estimate this uncertainty is to look at the difference in recordings from the surrounding rain gauges. We here used the coefficient of variation (CV, standard deviation divided by mean) calculated from intensities recorded at the surrounding rain gauges that meet the distance criteria to the catchment area. The rain input was considered uncertain if  $CV > 0.5$ , and if the total depth of the event was above 3 mm or the maximum intensity over 10 minutes was above 3.3 mm in at least one rain gauge, which on average occurs five times per year (Arnbjerg-Nielsen *et al.* 2006). This is a rough estimate for a model describing rain input uncertainty. It is assumed

**Table 2** | Basic signatures applied on water level observations

Signature	Name	Description	Unit
$h_{\max}$	Peak level	Peak level of the event	L
$t_{h_{\max}}$	Time of peak level	Time of peak level calculated from the start of the event	T
$dur_{\text{crestlevel}}$	Duration above crest level	Duration above crest level – 2 cm (to encounter that the observation can be uncertain and that the crest level is not always exactly known)	T
AUC	AUC	Area Under Curve, calculated with reference to a defined base level	L*T
$AUC_{\text{topofpipe}}$	AUC top of pipe	AUC calculated using the top of the pipe as base level	L*T
$AUC_{\text{crestlevel}}$	AUC crest level	AUC calculated using the crest level of an overflow weir as base level	L*T
$\Delta h/\text{time}$	Level rate of change	The change in level between two timesteps	L/T

Units: L = Length unit, T = Time unit.

that the rain uncertainty can be significantly improved by introducing weather radar data in the larger and more convective rain events. However, this is out of scope for this paper.

## RESULTS AND DISCUSSION

### Direct time series comparison and signatures

Figure 7 shows the observed and simulated time series for a rain-induced event at site 3. Each signature is tabulated for the event, and differences between observed and modeled values are calculated. The simulated level (model) of the first peak at 00:30 is much higher than the observed, leading to overestimating the surrogate overflow volume (AUC above crest level). There is a flattening of the curve for the second peak at 02:00, indicating the crest level (at level 24.35 m). The water level sensor has a lower bound at 23.81 m (zero point), which is the same as the invert level in the manhole. If the invert slope inside the manhole is large, the difference between the zero point and the invert level can be large, and using values in this range can lead to misinterpretations. The dash-dotted line at 24.06 m indicates the assumed top of the pass-forward pipe (according to the asset database).

### Multi-event signature comparisons

Extracting signatures from several rain-induced events provides insight into general tendencies and increases the robustness of the analysis. Examples from all three sites (Figure 4) were selected to illustrate how the signatures (Table 2) can be used to diagnose errors.

### Signatures illustrating system attribute input

Figure 8 illustrates different multi-event signature comparisons for site 3. Left panels represent the original model output, whereas the right panels show the output of a modified version of the model where detected structural errors have been remedied. This is described in detail in the SM. Figure 8 C+F directly compares the modeled and corresponding observed values for all events (green dots), whereas A+D and B+E show different signatures against each other (blue dots representing modeled values and red dots calculated from observations).

Figure 8(a) shows the signature ‘peak level’ versus the signature ‘AUC’. To some extent the observed peak levels are clustered around the levels app. 24.8 m, 24.35 m, and 24.0 m. 24.0 m indicates the approximate top level of the pass-forward pipe, and the modeled values are slightly higher than the observed values. This could indicate that the pass-forward pipe is not located as assumed. 24.35 m is the crest level of the overflow weir, and here the modeled and observed values correspond

	Model	Obs	Diff
rain threshold	23.81	23.85	0.04
peak level	25.23	24.68	-0.55
rolling median 5 min, peak level	25.23	24.66	-0.57
time of peak level [min]	259	251	-8
duration above crest level [min]	89.0	60.0	-29
AUC	136	73	87 %
AUC above top of pipe	74	34	116 %
AUC above crest level	19	4	345 %
RMSE			0.18
r2			-0.05

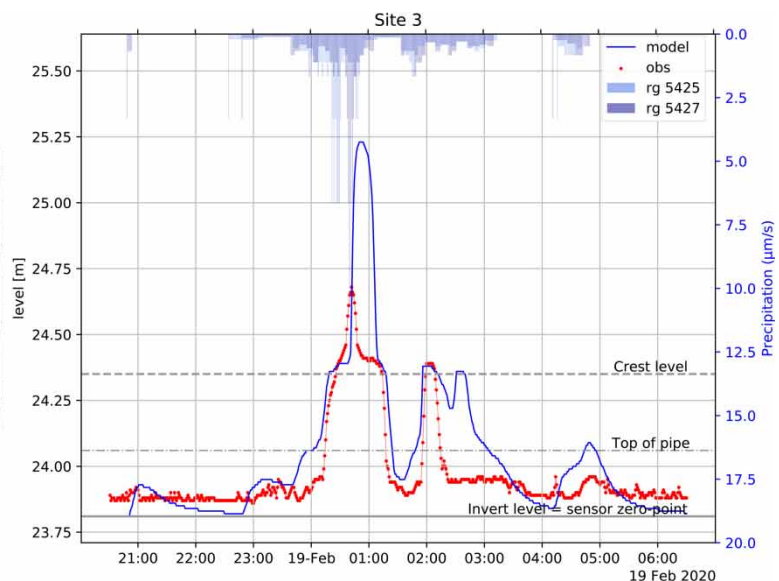
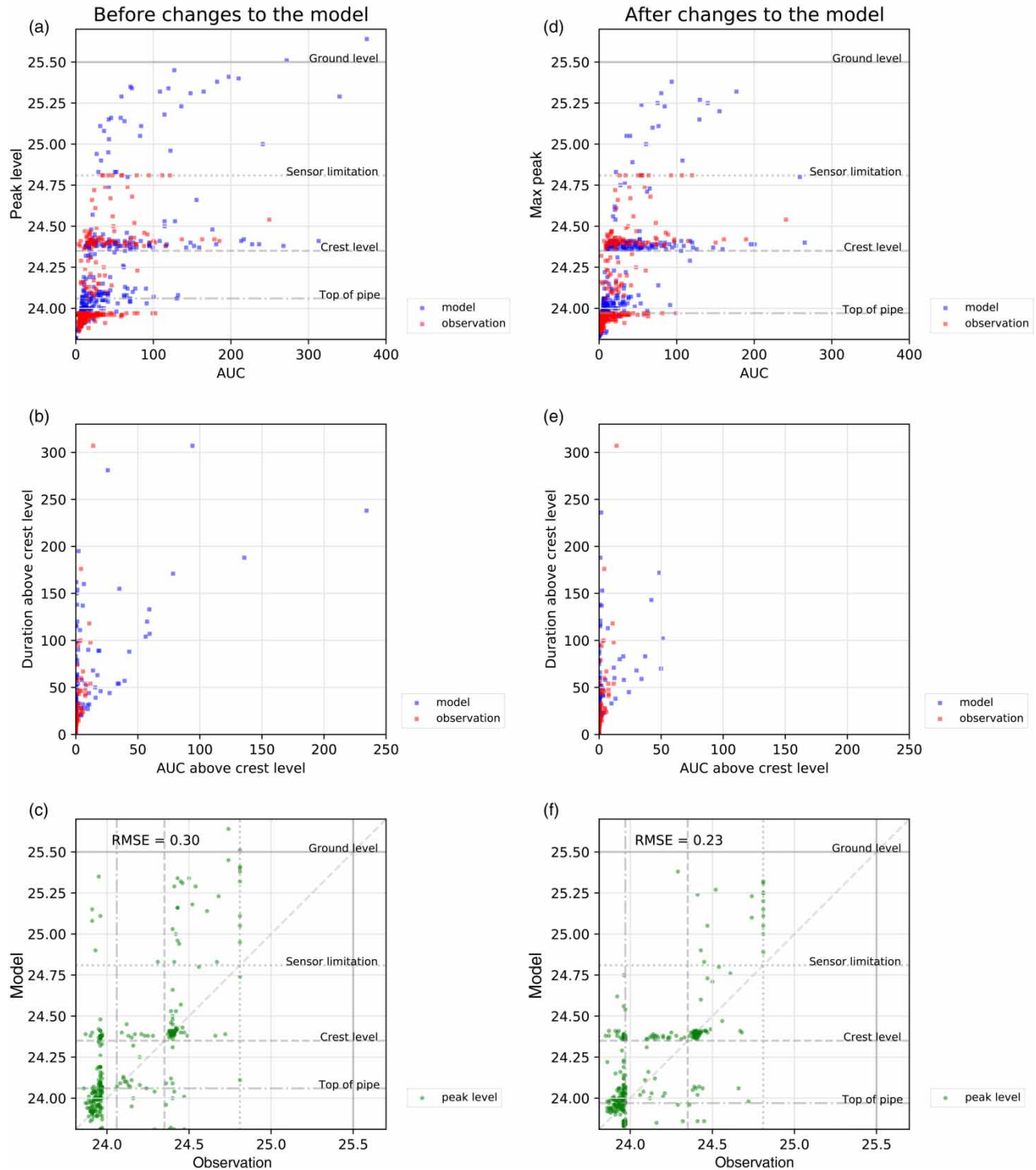


Figure 7 | Rain-induced event at the overflow structure at site 3. The dashed line at 24.35 m indicates the crest level, the dash-dotted line at 24.06 m is the expected top of the pass-forward pipe, assumed to be 250 mm (pipe diameter) above the invert level (23.81 m, solid grey line). Fixed levels indicate recorded system attributes. Obs Observation.

Site 3



**Figure 8** | Multi-event signature comparisons for rain-induced events for site 3. Left panels (a–c) represent the model before changes, while right panels (d–f) represent the model after changes. (c) and (f) show modeled vs. observed values of the signature ‘peak level’ (RMSE of the model-to-observation fit is displayed). In contrast, (a, b, d, e) show different signatures against each other calculated based on model simulations (blue) and observations (red). (a) and (d) show the signature ‘peak level’ against the signature ‘AUC,’ and (b) and (e) show the signature ‘duration above crest level’ against ‘AUC above crest level.’ Note that the events may not be exactly the same before and after changes in the model, as this depends on the algorithm behind it. Fixed levels representing important system attributes are indicated on (c–f) and (a–d) (Ground level 25.5 m; Sensor limitation 24.81 m; Crest level 24.35 m; Top of pass-forward pipe 24.06/23.97 m).



well. This is remarkable as water levels above weir crests are neither constant over time nor the length of the crest, and this highlights the potential power of multi-event signature diagnostics as suggested here. The observed peak values seem bounded with 24.8 m as a maximum, whereas the modeled peak levels are generally much higher, including even pluvial flooding (ground level is at 25.5 m). This could indicate that there is an upper constraint in the measurements. The AUC also tends to be higher for the model.

Signatures calculated for events causing overflow can be seen in Figure 8(b). The modeled 'AUC above crest level' is generally larger than the observed. The overflow duration reaches the same maximum, but modeled values tend to be larger than observed. The characteristics of Figure 8(a) can also be found in Figure 8(c) (pass-forward pipe not well represented in the model).

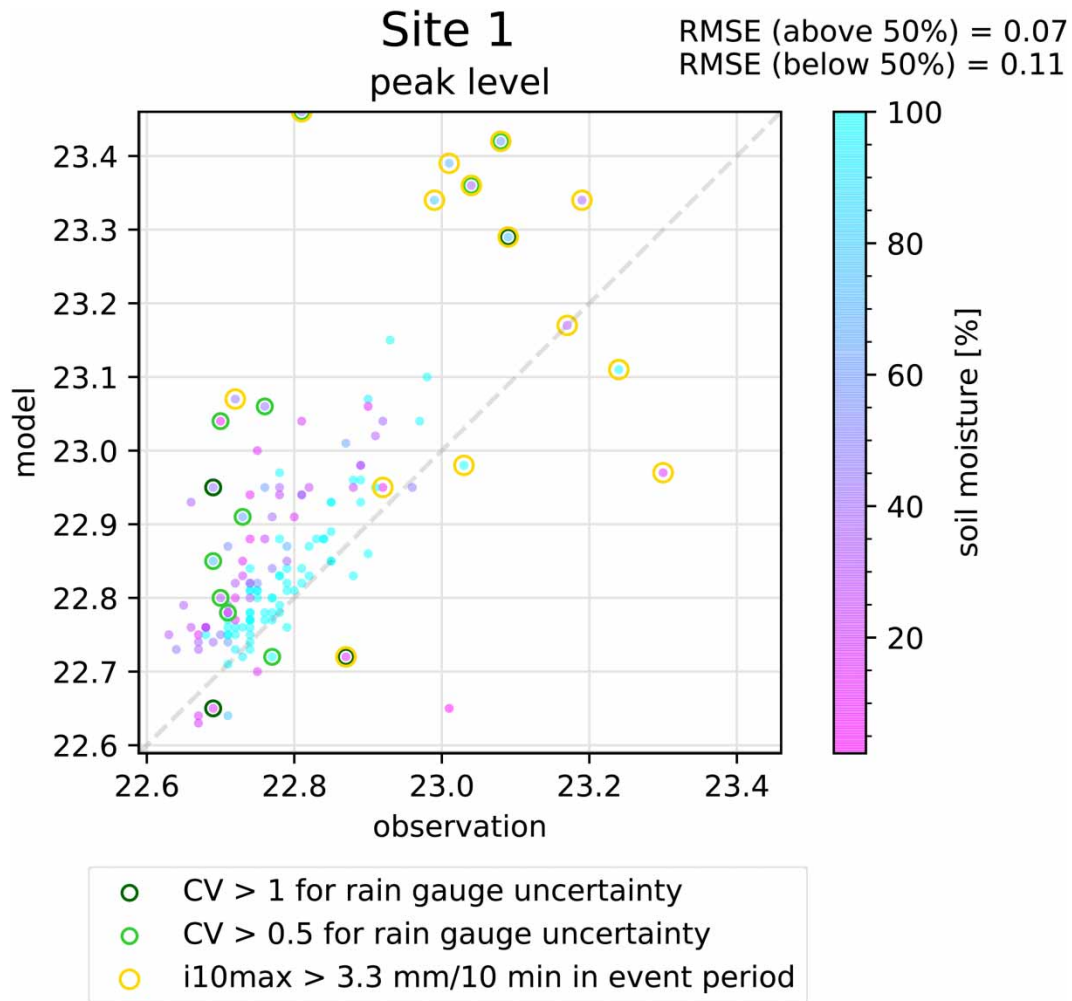
Figure 8(a)–8(c) indicate that there is a system replication deviation in at least two places: (1) the top of the pass-forward pipe seems to be lower than expected and thus simulated in the model, and (2) there is too much runoff in the model indicated by the AUC above crest level and the peak level. A field trip to the overflow structure revealed that the pass-forward pipe had an inlet orifice diameter of 160 mm in the manhole instead of 250 mm, as specified in the asset database for the pass-forward pipe. Furthermore, the construction drawings showed (see SM) that the pass-forward pipe was intended to be 160 mm, but instead a 250 mm pipe was constructed. A 160 mm orifice opening of the manhole thus serves as a throttle in the overflow structure even though the pass-forward pipe is actually larger. Analysis of building construction drawings in the area upstream from site 3 also showed that the connected area was overestimated (see SM). The model was hence updated with a smaller connected area and a correct throttle diameter of 160 mm, and the results are shown in Figure 8(d)–8(f). The model still overestimates the peak levels above 24.8 m and the AUC above the crest level, but there are clear general improvements. Interestingly, the two flood occurrences originally simulated (Figure 8(a)) disappear when simulated with the modified model (Figure 8(d)). The apparent maximum peak level at approximately 24.8 m was finally identified as a limitation in the sensor installation, as it was impossible to measure higher levels with the current sensor.

### Signatures and surrounding states illustrating process knowledge and spatial variability

An example where signatures can help diagnose a gap in process knowledge is a multi-event comparison of modeled vs. observed peak levels combined with states of the surrounding environment in terms of soil moisture and spatial rain uncertainty, as illustrated for site 1 in Figure 9. Events with both high CV and high peak intensity over 10 minutes ('i10max') are highlighted and reveal a large spread in the figure. This confirms the well known limitation of urban drainage modeling based on rain input from only a few rain gauges that spatio-temporal variability can lead to very large simulation uncertainty that prevents systematic comparison with in-sewer measurements. The remaining events are clustered in two groups, those in the wet period (above 50% soil moisture) and those in the dry period (below 50% soil moisture), and RMSE is calculated and depicted for both groups. The dry period events overestimate the peak level compared to the wet period events). This may indicate that either the initial loss in the model needs to be differentiated according to soil moisture, which indicates a *parameter assessment error*, or that the imperviousness is lower for events with a low soil moisture content, which indicates a *process knowledge error*. Without giving a precise answer, we refer to the literature dealing with this specific knowledge gap in urban drainage modeling (Thorndahl *et al.* 2008; Davidsen *et al.* 2018; Nielsen *et al.* 2019).

### Signatures illustrating equation structure and temporal variability

Figure 10 exemplifies for site 2 how signatures can help identify errors due to equation structure and temporal representation of process dynamics in the model. Levels are shown against the signature 'level rate of change' for short time intervals (1 and 5 min duration), which illustrates the response dynamics in the system. For example, a value of +4 m/min at level 14.2 m in the upper right panel (Figure 10) illustrates that the water can rise 4 m during 1 min from level 14.2 to 18.2 m. The observed response is much faster than the modeled, and the shape of the graphs cannot be replicated for shorter time steps than 5 min (bottom panels). The shape of the graphs indicates the physical limits of how high and low the water level in this structure can go. The observed values appear to have a cluster around -2 m/min (a drop) from level 19 to 16 m, which is due to the control of the flow regulator. Opening and closing the valve can occur in 12 seconds because it must flush the storage pipe up to three times when the flushing chamber is completely full. The model cannot simulate this as the control setting is implemented with a temporal resolution of one minute. This is why the patterns modeled as observed values for 1-minute delta levels are different. Figure 10 also illustrates a minor discrepancy between the lowest point in the observed and modeled values because the sensor has a zero point that is higher than the invert level of the model. Therefore, we would never get the same values in



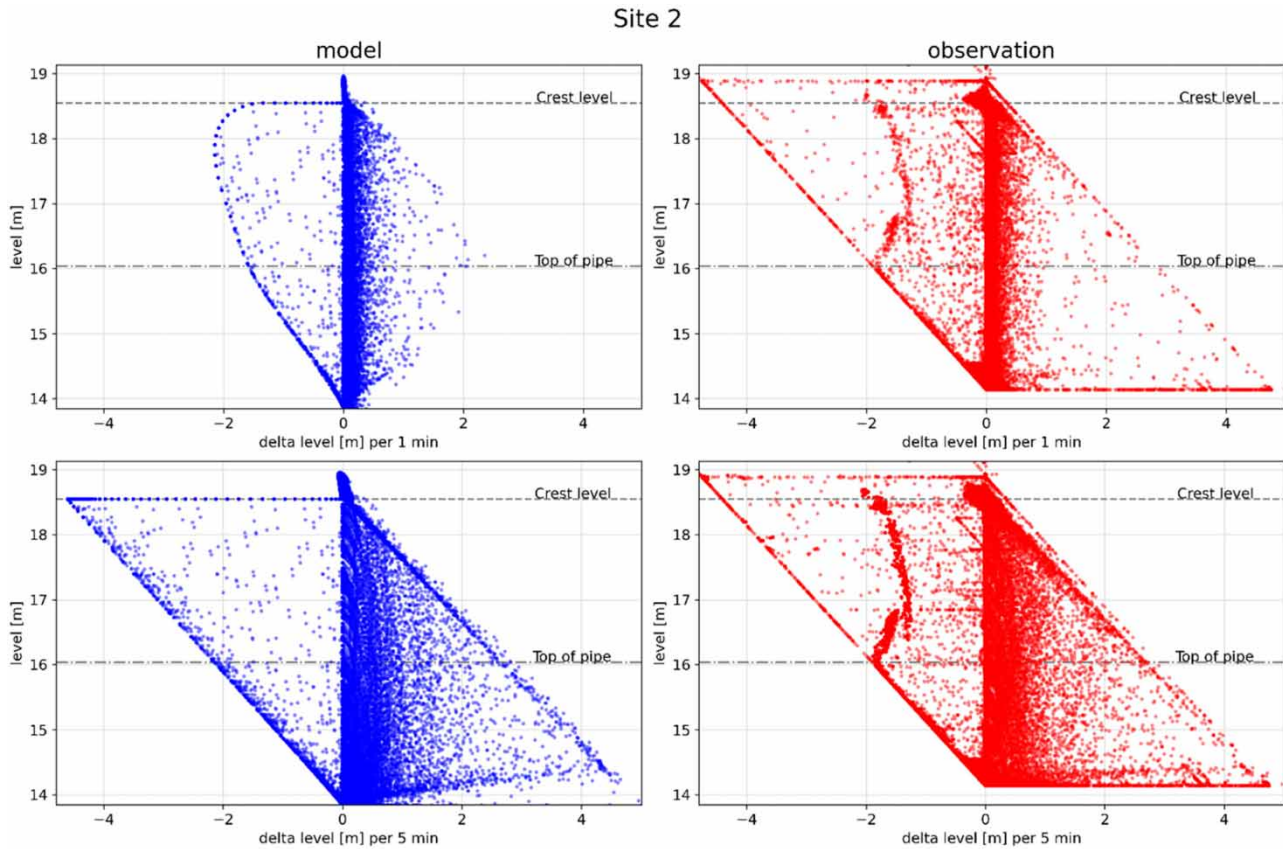
**Figure 9** | Multi-event signature comparisons for rain-induced events for site 1. Modeled vs. observed values of the signature 'max peak' together with two surrounding states: 'soil moisture' (indicated by a color scale) and 'spatial rain input uncertainty' (three categories indicated). RMSE is calculated for events that do not exceed the rain input uncertainty thresholds and is extracted into two categories: above or below 50% soil moisture.

these areas. The high levels (vertical axis) show that the observed values tend to cause a drop in the water level when they are above the crest, which the modeled values are not. This may be due to a utopian replication of the control rules, which are not always executed perfectly in reality. For example, the valve must be open to give a drop of 4 m in 1 min, and this is not allowed in the model unless the water level is below the crest level.

## Limitations and perspectives for further research

### Limitations by applying water level sensors

Signature-based error diagnostics are developed for water level sensors in urban drainage systems. However, measured water levels are not always good indicators of flows and volumes because these can be affected by local conditions both upstream and downstream from the measurement site. Therefore, caution must be exercised in its use. Knowledge of the physical system becomes especially important when applying water level sensors to validate the model. However, most sensors in urban drainage systems are water level sensors, and therefore we need to find ways to extract knowledge from these as well. In addition, water level sensors are cheaper to install and more robust than flow meters, and caution must also be taken when using flow meters in partially filled pipes.



**Figure 10** | Average water levels (vertical axis) vs. 'level rate of change' for time steps of 1 and 5 min. Left panels (blue) show modeled values, and right panels (red) show observed values. Positive values indicate that the water level rises and negative values that it drops.

The validity of the observations in this paper is taken for granted. However, there could be time steps where the model cannot replicate the measurement due to, for example, blockages in the system, poor calibration of sensors or manual errors in the sensor setup, which ideally should be identified before conducting error diagnostics in models. We suggest making a model specifically for anomaly detection (e.g., [Bertrand-Krajewski \*et al.\* 2003](#); [Therrien \*et al.\* 2020](#); [Palmitessa \*et al.\* 2021](#)), but this is beyond the scope of the present paper.

### The eyes of the data scientist

When validating urban drainage models, it is necessary to pay attention to the fact that stakeholders with different backgrounds and experiences may have different views of model adequacy, which will lead to different approaches to model construction. [Gupta \*et al.\* \(2012\)](#) illustrate three different viewpoints for defining model structural adequacy: (I) the engineering viewpoint where the evaluation is focused on functional adequacy, (II) the physical science viewpoint where the models must be in accordance with the physical system, and finally (III) the 'system science' viewpoint, which is a hybrid of the two previous ones. In (III), adequate simplified representation of a physical system is in focus, also referred to as the principle of parsimony. The last one favors the information content in observations as the model is compared to observational data ([Gupta \*et al.\* 2012](#)).

Different types of urban drainage models (control, operational, planning, design, cf. [Pedersen \*et al.\* \(2021a\)](#)) require different viewpoints. Control models may require little physical knowledge and instead obtain high accuracy by utilizing observations through data assimilation. Design models may require a good physical foundation and therefore favor the physical science viewpoint. However, the insights presented here on the limitations of an integrated urban drainage model probably require most modelers to accommodate the engineering viewpoint. For example, some hydraulic models (like the one used in this study) cannot resolve flow in pipes shorter than 10 m, which must be handled differently. Short pipes

should either be merged with connecting pipes or kept as is, in both cases leading to an inaccurate representation of reality. With increased automation of the model updating processes in living DTs (Pedersen *et al.* 2021a), the subjectivity of the modeler's choices may be reduced.

### **Are traditional calibration methods the best way to get good models?**

The different viewpoints will greatly influence how a given model error is remedied. For example, the estimation of runoff from an area can be based on physical knowledge of the contributing areas and an acceptance of the apparent error in these estimates. On the other hand, it can also be based on an engineering approach where runoff-relevant parameters are calibrated against observations. Calibration of parameters will likely reduce the apparent error of the model, but it might also reduce the ability to interpret parameter values from a physical point of view. Ideally, calibration is done on a subset of events followed by validation on a separate set of events. However, the author's experiences suggest that many utility companies in practice end up only calibrating parts of the model as observations are obtained in an ad-hoc manner and mainly perform calibration whenever model-observation comparisons show very different behavior than expected. The danger of such an ad-hoc approach is that the tuning of parameters compensates for structural errors located in other parts of the model, like the errors shown in this paper in the asset database, surrounding states, and process dynamics. Therefore, it is suggested that applying a structured error diagnostics framework like the one developed here will be paramount for the success of any later calibration exercise.

The developed framework is highly scalable to many sites in the system and provides the ability to learn across sites. In the service area of VCS Denmark, the suggested methodology will be applied on all observation sites (+400) during the coming years. The general framework of signature-based error diagnostics is in principle applicable to any type of model. This study argues for tailoring its use towards the detailed hydrodynamic part of integrated urban drainage system models that is based on detailed representation of physical system attributes. The results are thus applicable to other model codes than the Mike Urban software applied in this study (SWMM, InfoWorks, etc.). This will improve the understanding of where the models generally fail due to limitations of the applied software, for example, conceptualizations of the hydrologic and hydraulic/hydrodynamic processes or numerical implementation, and where an error is site-specific and justifies a designated, local field investigation.

## **CONCLUSION**

This paper presents a structured framework for diagnosing errors in integrated urban drainage models intended for continuous, iterative improvements of living DTs. The framework provides a classification scheme, which divides different types of errors and uncertainty locations in urban drainage models into four categories: context, input, model structure, and parameter. We show that the concept of hydrologic and hydraulic signatures, i.e., characteristics of system functioning extracted from time series, can successfully be applied to urban drainage systems. While signatures are typically estimated from flow measurements in general hydrology, this paper shows that water level measurements can also be used for feature extraction. This is highly relevant for urban drainage systems where flow measurements can be difficult to obtain due to the harsh sewer environments.

The framework is applied to a real case study in Odense, Denmark, with more than 10 years of observations representing more than 500 rain-induced events available for a diverse set of structures in the local drainage system: a regular manhole, a flushing chamber upstream of a large storage pipe, and an internal overflow structure. Rain-induced stormwater events were identified and extracted from the time series. Two 'surrounding states' related to soil moisture and uncertainty of rain input due to spatial variability were defined to support this process. The diagnostic framework was shown to be an efficient way to identify the source of various errors and identified patterns in model performance that were not clear to the utility company beforehand. The provided examples identified errors in the utility's asset database, errors due to hydrologic processes not represented in the model, and mathematical limitations in the model code in terms of temporal resolution. Especially, the ability to find errors in the asset database is interesting as it is of value to the entire utility organization. With the developed framework it is easier to identify what actions are suitable to mitigate the errors and improve the transparency of the model performance for end users.

With increasing digitalization, the introduction of living DTs in utility companies, and a surge in the number of in-sewer sensors enabled by IoT, there is a clear need for automating model performance evaluation within a structured framework like the one developed here. Therefore, continued research should focus on matching specific signatures to specific types



of model errors for improved guidance on the location of errors related to representation of physical attributes. There is furthermore a need for defining which signatures that can serve as indicators of how suitable a given model is for various different operational domains (dry-weather conditions, 'every day' rain events, extreme rainfall), different objectives (simulation of overflow volumes, simulation of critical level exceedance) and different processes (infiltration-inflow, combined sewer overflows, manhole surcharging). This will greatly improve the transparency of model usefulness to end users that are not trained drainage experts. The future application of DTs is vulnerable to their confidence, which means that if we cannot communicate the uncertainties and application area for the DTs, they will not be trusted.

## ACKNOWLEDGEMENTS

The authors would like to thank their colleagues in VCS Denmark for contributing valuable information about the sensors and field measurements when necessary. The work has been carried out in collaboration with the Technical University of Denmark and VCS Denmark through an Industrial Ph.D. project partly funded by VCS Denmark and Innovation Fund Denmark (grant no. 8118-00018B).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

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

## REFERENCES

- Arnbjerg-Nielsen, K., Madsen, H. & Mikkelsen, P. S. 2006 Regional variation af ekstremregn i Danmark – ny bearbejdning (1979–2005) – spildevandskomiteen. *Skrift nr. 28*, 1–54.
- Autiosalo, J., Vepsäläinen, J., Viitala, R. & Tammi, K. 2020 A feature-based framework for structuring industrial digital twins. *IEEE Access* **8**, 1193–1208. <https://doi.org/10.1109/ACCESS.2019.2950507>.
- Bach, P. M., Rauch, W., Mikkelsen, P. S., McCarthy, D. T. & Deletic, A. 2014 A critical review of integrated urban water modelling – urban drainage and beyond. *Environmental Modelling & Software* **54**, 88–107. <https://doi.org/10.1016/j.envsoft.2013.12.018>.
- 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 Science and Technology* **47**, 95–102. <https://doi.org/10.2166/wst.2003.0094>.
- Breinholt, A., Grum, M., Madsen, H., Örn Thordarson, F. & Mikkelsen, P. S. 2013 Informal uncertainty analysis (GLUE) of continuous flow simulation in a hybrid sewer system with infiltration inflow – consistency of containment ratios in calibration and validation? *Hydrology and Earth System Sciences* **17**, 4159–4176. <https://doi.org/10.5194/hess-17-4159-2013>.
- Davidson, S., Löwe, R., Ravn, N. H., Jensen, L. N. & Arnbjerg-Nielsen, K. 2018 Initial conditions of urban permeable surfaces in rainfall-runoff models using horton's infiltration. *Water Science and Technology* **77**, 662–669. <https://doi.org/10.2166/wst.2017.580>.
- Deletic, A., Dotto, C. B. S., McCarthy, D. T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T. D., Rauch, W., Bertrand-Krajewski, J. L. & Tait, S. 2012 Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth* **42–44**, 3–10. <https://doi.org/10.1016/j.pce.2011.04.007>.
- del Giudice, D., Reichert, P., Bareš, V., Albert, C. & Rieckermann, J. 2015 Model bias and complexity – understanding the effects of structural deficits and input errors on runoff predictions. *Environmental Modelling and Software* **64**, 205–214. <https://doi.org/10.1016/j.envsoft.2014.11.006>.
- DMI 2020 DMI Open Data. <https://confluence.govcloud.dk/display/FDAPI> (accessed 8.17.20).
- DMI (Danish Meteorological Institute), IDA (The Danish Society of Engineers) 2020 The Water Pollution Committee – Rain Gauge System (in Danish: *Spildevandskomiteens Regnmålerstyrgruppe*). <https://ida.dk/om-ida/spildevandskomiteen/regnmaalerstyrgruppen> (accessed 10.23.20).
- Fuertes, P. C., Alzamora, F. M., Carot, M. H. & Campos, J. C. A. 2020 Building and exploiting a digital twin for the management of drinking water distribution networks. *Urban Water Journal* **17**, 704–713. <https://doi.org/10.1080/1573062X.2020.1771382>.
- Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J. K. & McMillan, H. K. 2021 TOSSH: A toolbox for streamflow signatures in hydrology. *Environmental Modelling and Software* **138**. <https://doi.org/10.1016/j.envsoft.2021.104983>.
- Gregersen, I. B., Sørup, H. J. D., Madsen, H., Rosbjerg, D., Mikkelsen, P. S. & Arnbjerg-Nielsen, K. 2013 Assessing future climatic changes of rainfall extremes at small spatio-temporal scales. *Climatic Change* **118**, 783–797. <https://doi.org/10.1007/s10584-012-0669-0>.

- Gupta, H. v., Wagener, T. & Liu, Y. 2008 Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes* **22**, 3802–3813. <https://doi.org/10.1002/hyp.6989>.
- Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G. & Ye, M. 2012 Towards a comprehensive assessment of model structural adequacy. *Water Resources Research* **48**, 1–16. <https://doi.org/10.1029/2011WR011044>.
- Hamel, P. & Fletcher, T. D. 2014 Modelling the impact of stormwater source control infiltration techniques on catchment baseflow. *Hydrological Processes* **28**, 5817–5831. <https://doi.org/10.1002/hyp.10069>.
- Hansen, L. S., Borup, M., Møller, A. & Mikkelsen, P. S. 2014 Flow forecasting using deterministic updating of water levels in distributed hydrodynamic urban drainage models. *Water (Switzerland)* **6**, 2195–2211. <https://doi.org/10.3390/w6082195>.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. & Cudennec, C. 2013 A decade of predictions in ungauged basins (PUB)—a review. *Hydrological Sciences Journal* **58**, 1198–1255. <https://doi.org/10.1080/02626667.2013.803183>.
- Huber, W. & Dickinson, R. E. 1992 *Storm Water Management Model, Version 4: User's Manual*. Rep. No. EPA/600/3-88/001a 1, United States Environmental Protection Agency, Athens, Georgia, USA. pp. 1–720.
- Jørgensen, H. K., Rosenørn, S., Madsen, H. & Mikkelsen, P. S. 1998 Quality control of rain data used for urban runoff systems. *Water Science and Technology* **37**, 113–120. [https://doi.org/10.1016/S0273-1223\(98\)00323-0](https://doi.org/10.1016/S0273-1223(98)00323-0).
- LNHwater 2017 *Kortlægning af Overflader I Odense -Opdatering af Analysen fra 2010*.
- 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. *Critical Reviews in Environmental Science and Technology* **48**, 279–339. <https://doi.org/10.1080/10643389.2018.1455484>.
- McMillan, H. K. 2020a A review of hydrologic signatures and their applications. *Wiley Interdisciplinary Reviews: Water*. <https://doi.org/10.1002/wat2.1499>.
- McMillan, H. K. 2020b Linking hydrologic signatures to hydrologic processes: a review. *Hydrological Processes* **34**, 1393–1409. <https://doi.org/10.1002/hyp.13632>.
- McMillan, H. K., Westerberg, I. K. & Branger, F. 2017 Five guidelines for selecting hydrological signatures. *Hydrological Processes* **31**, 4757–4761. <https://doi.org/10.1002/hyp.11300>.
- Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. v. & Kumar, R. 2019 On the choice of calibration metrics for 'high-flow' estimation using hydrologic models. *Hydrology and Earth System Sciences* **23**, 2601–2614. <https://doi.org/10.5194/hess-23-2601-2019>.
- Nielsen, K. T., Moldrup, P., Thorndahl, S., Nielsen, J. E., Uggerby, M. & Rasmussen, M. R. 2019 Field-scale monitoring of urban green area rainfall-Runoff processes. *Journal of Hydrologic Engineering* **24**. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001795](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001795).
- Palmitessa, R., Mikkelsen, P. S., Borup, M. & Law, A. W. K. 2021 Soft sensing of water depth in combined sewers using LSTM neural networks with missing observations. *Journal of Hydro-Environment Research* **38**, 106–116. <https://doi.org/10.1016/j.jher.2021.01.006>.
- Pedersen, A. N., Borup, M., Brink-Kjær, A., Christiansen, L. E. & Mikkelsen, P. S. 2021a Living and prototyping digital twins for urban water systems: towards multi-Purpose value creation using models and sensors. *Water* **13**, 592. <https://doi.org/10.3390/w13050592>.
- Pedersen, A. N., Pedersen, J. W., Viguera-Rodríguez, A., Brink-Kjaer, A., Borup, M. & Mikkelsen, P. S. 2021b The belling data set: open data and models for community-wide urban drainage systems research. *Earth System Science Data* **13**, 4779–4798. <https://doi.org/10.5194/essd-13-4779-2021>.
- Scharling, M. & Vilic, K. 2009 *Tørkeindeks Version 1.0 – Metodebeskrivelse*. København, Denmark. Danmarks Meteorologiske Institut.
- Stentoft, P. A., Munk-Nielsen, T., Møller, J. K., Madsen, H., Valverde-Pérez, B., Mikkelsen, P. S. & Vezzaro, L. 2021 Prioritize effluent quality, operational costs or global warming? – Using predictive control of wastewater aeration for flexible management of objectives in WRRFs. *Water Research* **196**, 116960. <https://doi.org/10.1016/j.watres.2021.116960>.
- Therrien, J.-D., Nicolai, N. & Vanrolleghem, P. A. 2020 A critical review of the data pipeline: how wastewater system operation flows from data to intelligence. *Water Science and Technology* **82**, 2613–2634. <https://doi.org/10.2166/wst.2020.393>.
- Thorndahl, S., Beven, K. J., Jensen, J. B. & Schaarp-Jensen, K. 2008 Event based uncertainty assessment in urban drainage modelling, applying the GLUE methodology. *Journal of Hydrology* **357**, 421–437. <https://doi.org/10.1016/j.jhydrol.2008.05.027>.
- Tscheikner-Gratl, F., Zeisl, P., Kinzel, C., Rauch, W., Kleidorfer, M., Leimgruber, J. & Ertl, T. 2016 Lost in calibration: why people still do not calibrate their models, and why they still should – A case study from urban drainage modelling. *Water Science and Technology* **74**, 2337–2348. <https://doi.org/10.2166/wst.2016.395>.
- Valverde-Pérez, B., Johnson, B., Wärrf, C., Lumley, D., Torfs, E., Nopens, I. & Townley, L. 2021 *Digital Water – Operational Digital Twins in the Urban Water Sector: Case Studies*.
- van der Werf, J. A., Kapelan, Z. & Langeveld, J. 2021 Quantifying the true potential of real time control in urban drainage systems quantifying the true potential of real time control in urban drainage systems. *Urban Water Journal* 1–12. <https://doi.org/10.1080/1573062X.2021.1943460>.
- Vonach, T., Kleidorfer, M., Rauch, W. & Tscheikner-Gratl, F. 2019 An insight to the cornucopia of possibilities in calibration data collection. *Water Resources Management* **33**, 1629–1645. <https://doi.org/10.1007/s11269-018-2163-6>.

- Walker, W., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M., Janssen, P. & Kreyer von Krauss, M. 2003 **Defining uncertainty: a conceptual basis for uncertainty management**. *Integrated Assessment* **4**, 5–17. <https://doi.org/10.1076/iaij.4.1.5.16466>.
- 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**. *Environmental Modelling and Software* **25**, 1518–1527. <https://doi.org/10.1016/j.envsoft.2010.04.011>.
- Westerberg, I. K. & McMillan, H. K. 2015 **Uncertainty in hydrological signatures**. *Hydrology and Earth System Sciences* **19**, 3951–3968. <https://doi.org/10.5194/hess-19-3951-2015>.

First received 19 July 2021; accepted in revised form 2 February 2022. Available online 15 February 2022