Water Research 126 (2017) 88–100

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

An automated toolchain for the data-driven and dynamical modeling of combined sewer systems

Sara C. Troutman [a], Nathaniel Schambach [b], Nancy G. Love [a], Branko Kerkez [a, *]

[a] Civil and Environmental Engineering, University of Michigan, 2350 Hayward St, G.G. Brown Building, Ann Arbor, MI 48109, USA
[b] Mechanical Engineering, University of Michigan, 2350 Hayward St, G.G. Brown Laboratory, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 15 May 2017
Received in revised form
9 August 2017
Accepted 29 August 2017
Available online 3 September 2017

Keywords:
Sensor networks
Data-driven modeling
Smart water systems
Combined sewer

ABSTRACT

The recent availability and affordability of sensors and wireless communications is poised to transform our understanding and management of water systems. This will enable a new generation of adaptive water models that can ingest large quantities of sensor feeds and provide the best possible estimates of current and future conditions. To that end, this paper presents a novel data-driven identification/learning toolchain for combined sewer and stormwater systems. The toolchain uses Gaussian Processes to model dry-weather flows (domestic wastewater) and dynamical System Identification to represent wet-weather flows (rainfall runoff). By using a large and high-resolution sensor dataset across a real-world combined sewer system, we illustrate that relatively simple models can achieve good forecasting performance, subject to a finely-tuned and continuous re-calibration procedure. The data requirements of the proposed toolchain are evaluated, showing sensitivity to spatial heterogeneity and unique timescales across which models of individual sites remain representative. We identify a near-optimal time record, or data "age," for which historical measurements must be available to ensure good forecasting performance. We also show that more data do not always lead to a better model due to system uncertainty, such as shifts in climate or seasonal wastewater patterns. Furthermore, the individual components of the model (wet- and dry-weather) often require different volumes of historical observations for optimal forecasting performance, thus highlighting the need for a flexible re-calibration toolchain rather than a one-size-fits-all approach.

1. Introduction

Combined sewers convey large quantities of wastewater and stormwater to downstream treatment facilities. The delivery of these waters is highly dynamic, being dependent not only on diurnal wastewater patterns, but also on highly uncertain precipitation inputs. The latter is also true in separated sewer systems, which often become susceptible to infiltration due to aging (Ellis, 2001; Karpf and Krebs, 2011; Neshaei et al., 2017; Pawlowski et al., 2013). The sheer size and complexity of these systems makes it nearly impossible for operators to anticipate transient changes and optimally control every field-deployed asset, especially during spatially variable storms. These assets include, but are not limited to, pumps, gates, inflatable pillows, and large storage basins, which store, divert, and discharge excess flows during large storms. Improving how all of these assets are controlled and coordinated in real-time will not only reduce harmful combined sewer overflows (CSOs) (Borsanyi et al., 2008; Löwe et al., 2016), but will also minimize variability of the wastewater inflows that impact treatment operations and performance (Meirlaen et al., 2002; Risholt et al., 2002; Schütze et al., 2004; Seggelke and Rosenwinkel, 2002). To that end, autonomous and coordinated real-time control stands to change how sewer networks are operated across the scale of entire cities (Bach et al., 2014; Beeneken et al., 2013; Seggelke et al., 2013; Vanrolleghem et al., 2005).

The efficacy of any system-scale control must be underpinned by accurate estimates of field conditions, such as water flow, levels, and quality. The recent availability and affordability of wireless sensing technologies will lead to highly instrumented water systems in the near future (Häck and Wiese, 2006; Hill et al., 2014). Once sensors become dispersed throughout sewer networks, the data obtained will form the backbone for real-time management and decision-making. However, it will not be sufficient to just use

* Corresponding author.
E-mail addresses: stroutm@umich.edu (S.C. Troutman), bkerkez@umich.edu (B. Kerkez).

http://dx.doi.org/10.1016/j.watres.2017.08.065
0043-1354/

the latest measurements for decision making. Depending on the size and complexity of infrastructure, once a problem is detected in the field, it may already be too late to respond. In many instances, the combined sewer system or treatment plant will need to be prepared hours or days in advance of storms to ensure that existing assets are maximally leveraged in anticipation of any given input scenario; e.g., releasing flows from basins or in-line storage to make room for an incoming storm, or rerouting flows during a storm to maximize system-wide storage. Akin to steering a large ship around an obstacle, control actions in large sewer networks will need to be proactive rather than reactive. In a control theoretic context, this brings up the important need for model predictive control (making decisions based on predicted future outcomes) (Schütze et al., 2011), rather than strict feedback control (making decisions based just on real-time conditions). Effective control strategies will require the most up-to-date knowledge of system dynamics, which may change over time and require model re-calibration. A reliable forecast of future flow thus becomes imperative.

Once calibrated, a water model does not remain calibrated indefinitely. Many water systems exhibit *uncertainty*, which is driven by short-term shifts in wastewater patterns, seasonal runoff dependencies, or long-term climate and land use changes (Rosén, 2001; van Daal et al., 2017; Vaze et al., 2010). Thus, a vision for smart water infrastructure, which adapts itself in real-time to human and natural inputs, demands the development of a new generation of adaptive models, which will ingest unprecedented quantities of streaming sensor feeds to provide the best possible estimates of current and future conditions. The development of such flexible modeling toolchain will, however, require a holistic approach that combines our domain knowledge of water systems with modern advances in real-time data processing.

To this end, the goal of this paper is to enable a fully automated and data-driven approach for the dynamical modeling and prediction of dry- and wet-weather flows in a combined sewer system. The core innovation behind our toolchain relates to its automated identification, whereby the toolchain continually re-calibrates the underlying model using real-time sensor feeds to ensure the best possible forecasts of future system flows. The reliance on real-time measurements ensures that system operators and future control algorithms will always be informed by the most up-to-date understanding of system dynamics, especially as these dynamics evolve due to changing weather or land uses. While this paper does not explicitly address control strategies, the toolchain is inherently structured to support predictive control in the future. The specific contributions of this paper are:

- A new data-driven identification toolchain for combined sewer and stormwater systems, based on Gaussian Processes and dynamical System Identification,
- A characterization of system uncertainty, which guides how often components of a model need to be re-calibrated to reflect the uniquely changing nature of urban water systems.

To justify the need for this approach, we begin by providing an overview of existing models for combined sewer systems. The proposed toolchain will then be introduced and evaluated using a novel cloud-based data architecture. Finally, this entire end-to-end solution will be evaluated on sensor data collected in a large, real-world combined sewer system.

1.1. Existing approaches

1.1.1. Physical modeling

The most longstanding approach for the modeling of sewer networks has been the physical model. These models seek to characterize the entire sewer collection system, including the drainage subcatchments, the wastewater generation patterns, the pipe network, and many other physical components. Due to this large degree of characterization, physical models have greatly added to our understanding and management of urban water systems. Such a high level of detail requires a correspondingly high level of parameterization including land use, soil types, and pipe characteristics (e.g., slope, diameter, roughness), as well as less well-defined information (e.g., roof downspout connections). Maintaining these models at the city-scale is laborious and expensive, especially when considering the need to update model parameters in response to urban development, urban contraction, and the implementation of new distributed stormwater solutions (Doglioni et al., 2009; Fletcher et al., 2013; Liu et al., 2015). As such, uncertainty in the dynamics of the system limits the useful life of physical models. Additionally, the most common physical models are constructed using systems of partial differential equations, such as the Saint-Venant equations, requiring advanced analytic or numerical techniques to generate solutions (Vanrolleghem et al., 2005), and demand significant computational effort for model simulations. Hence, the challenge of using large and high-resolution physical models for real-time control concerns the computational expense and complexity related to re-calibration (Vanrolleghem et al., 2005; Meirlaen et al., 2002).

1.1.2. Data-driven modeling

The development of data-driven approaches has been increasing in the modeling of urban water systems. This is most evident in the use of Neural Networks (NNs), a form of *black-box* model in which hidden parameters, or weight layers, are adjusted to *"learn"* the relationship between measured input and output data (Haykin, 1999). Most often, the input data comprise a rainfall time series and the NN is trained to predict the corresponding flows (Dawson and Wilby, 2001; Kisi et al., 2013; Kurth et al., 2008; Li et al., 2010; Mounce et al., 2014; Smith and Eli, 1995). This approach relies only on data, which has made it a popular and powerful tool across many disciplines beyond water resources. Unlike in physical models, characterization of the actual water system is not required. The application of neural networks for the modeling of hydrologic and hydraulic systems has generally reported good model performance (El-Din and Smith, 2002; Kurth et al., 2008; Li et al., 2010; Mounce et al., 2014). In part, this can be explained by the ability of NNs to model highly non-linear and nuanced relationships between input-output data sets (El-Din and Smith, 2002; Haykin, 1999; Maier and Dandy, 2000). Furthermore, once trained, NNs are highly computationally efficient in making fast predictions of future system states (Kurth et al., 2008; Li et al., 2010; Mounce et al., 2014).

Unlike physical models, however, the parameters of NNs often lack physical interpretation (Li et al., 2010; Maier and Dandy, 2000; Solomatine and Dulal, 2003; Todini, 2007). Since the majority of optimization and control approaches depend on an explicit description of system dynamics (Ruano, 2005), this limits the use of NNs in robust management and safety-critical control approaches. Most importantly, perhaps, requirements pertaining to data quality and measurement or model uncertainty have yet to be clarified, which limits the extent to which these models can be transferred between study areas or accommodate changing conditions.

2. Toward a holistic real-time modeling toolchain

More so than just a model, the real-time forecasting in sewer networks demands an end-to-end toolchain. While a model represents the underlying dynamics of the system, it is only one part of

a more complex processing chain, which must ingest noisy sensor data and convert it to actionable forecasts. The complexities associated with such a task are best illustrated visually, as shown in Fig. 1, in which rain and flows in a combined sewer system are plotted. Real-world sensor data is inherently noisy, often to a degree beyond which it is difficult to visually interpret. Once filtered, however, the underlying dynamics become more apparent, such as stormwater inputs or the diurnal wastewater inputs generated by households. Since these diurnal dynamics and wet-weather flows are described by fundamentally separate dynamics, calibration of the model becomes challenging because the two input data streams must first be decoupled.

Furthermore, depending on the location of the sensors in the collection network, the magnitude of the underlying flows differs, making it impractical to translate model parameters of one site to another. System-based uncertainty in the measured signals also introduces significant challenges, as a set of model parameters that were calibrated in the past may not adequately describe present dynamics due to changing climate, baseflows, or human inputs. While many more challenges exist, for instance deploying sensors for water quality as in Banik et al. (2017), a set of core requirements arises in the development of a robust real-time modeling toolchain:

- Streaming sensor data must be ingested, quality-controlled, and decoupled (dry-weather flows, wet-weather flows, etc.) rapidly before they can be used to calibrate or re-calibrate the underlying model.
- The approach should be automated and readily transferable between sites, requiring only data as an input.
- The underlying model must be computationally efficient to ensure that forecasts can be made within time windows that are suitable for decision-making and control.
- Ideally, the model dynamics should be structured in a way to interpret the physical meaning of parameters and enable the use of formal control theoretic or optimization approaches.

In order to meet these requirements, our approach is data-driven, yet aims to preserve intuition by decoupling the signal into physically-meaningful flow components (i.e., dry- and wet-weather) — that is, rather than being fully abstracted as a black-box, the underlying models are of a structure that should be familiar and interpretable to those working in the water domain.

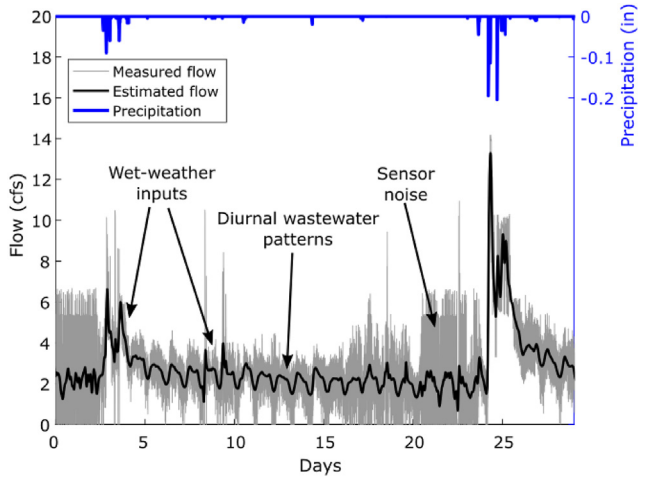

Fig. 1. Example of sensor signal in a combined sewer system, sampled at 5-min resolution, highlighting the complexities that are present in real-world measurements (Site Q05). The estimated flow was determined using a low-pass Butterworth filter to remove high frequency sensor noise. More information about sensor measurements is included in Section 3.3.1.

3. Methods

Given a rainfall measurement or weather forecast p, our problem is framed by the need to predict the flow q at time t in a combined sewer or stormwater conduit. In this section, we first begin by introducing the model structure. We then describe the core contribution of the paper: an automated identification toolchain used to continually re-calibrate the underlying model in response to streaming sensor data.

3.1. Model structure

The instantaneous flows in a combined sewer are described by the sum of wet-weather flows (rainfall runoff) and dry-weather inputs (domestic wastewater). Specifically, we assume that

$$q(t) = h(t) + d(t), \tag{1}$$

where q is the combined flow, h is the wet-weather flow, and d is the dry-weather or wastewater flow. The flexibility of this representation permits it to be used for a number of models. For example, a stormwater network can be represented when the dry-weather flows are not modeled, while a non-combined sewer network can be represented without accounting for wet-weather flows. Our approach does not explicitly quantify groundwater infiltration into the network since this component is implicitly included in the "dry-weather" estimates. Infiltration can be a significant and important component in many systems, and could be estimated, if needed, by further segmenting the flows.

The dry-weather flows $d(t)$ are governed by a repeating probabilistic process, which must be *learned* from the data. The wet-weather flows $h(t)$ are approximated by an n-th order linear differential equation

$$\begin{aligned} &\frac{d^n h}{dt^n} + a_1 \frac{d^{n-1} h}{dt^{n-1}} + \cdots + a_{n-1} \frac{dh}{dt} + a_n h(t) \\ &= b_0 \frac{d^n p}{dt^n} + b_1 \frac{d^{n-1} p}{dt^{n-1}} + \cdots + b_{n-1} \frac{dp}{dt} + b_n p(t), \end{aligned} \tag{2}$$

where $a_1, \ldots, a_n$ and $b_0, \ldots, b_n$ are parameters. When $n = 1$ the model presents the familiar and physically-intuitive *unit hydrograph*, which is used in the hydrologic sciences to conceptualize the hydrograph resulting from one unit of rainfall (e.g., $m^3/s/mm$). Increasing the order of the model permits for more nuanced dynamics to be represented. The biggest benefit of this representation relates to the ability to rewrite the model as the impulse-driven system or transfer function (Golnaraghi and Kuo, 2009), enabling the application of powerful parameter identification tools, which have been developed in the dynamical systems community (Ljung, 1999). Furthermore, this formalism allows for feedback control and model predictive control methods to be applied to the system (Ruano, 2005), thus opening the door to future real-time control applications.

3.2. Toolchain to identify and predict dry- and wet-weather flows

Given sensor measurements of flow q and rainfall p, our problem is framed by the need to identify the model order and the parameters describing the wet-weather flows, as well as the parameters of the dry-weather flows. Once the model parameters are identified or *learned*, the resulting model can then be used in a predictive fashion by taking the measured state q and forward-

modeling it using the forecasted or measured rainfall p.

For a given sensor pair (flow and rain), our toolchain (Fig. 2) identifies the model parameters by *training* the full model on a set of historical observations of flows and corresponding rainfall measurements. The first step in our processing chain involves the *learning* of the dry-weather flows using a Gaussian Process (Section 3.2.2). The full time series of historical dry-weather flows is then estimated and subtracted from the original flows to derive an estimate of the wet-weather flows, which are then used along with measured rainfall to identify the parameters of the wet-weather model (Section 3.2.3). This step is very important, as it is particularly difficult using frequency-based filtering, for example, to separate a storm from the dry-weather flows when the two have similar magnitudes or occur across the same timescale. The approach is inherently flexible, as it can be re-calibrated continuously, or as needed, when more measurements become available. As will be shown, given our dual identification and estimation formulation, the approach is also highly resilient to missing or incomplete measurements. The only input requirements are pairs of flow and rainfall measurements.

3.2.1. Flow separation and pre-processing

The first step in the processing chain involves the separation of the dry-weather flow from the measured signal (Fig. 2a–b). The observed dry-weather diurnal patterns display daily repetition ($1\ \frac{1}{day}$) and approximate semi-daily peaks ($2\ \frac{1}{day}$). Thus, to achieve this separation, a *Butterworth* bandpass filter (Andreas, 1993) parameterized with a range of $0.5–3\ \frac{1}{day}$ is applied. This has the immediate impact of removing high-frequency sensor noise and low-frequency flows (Fig. 2b). The outputs of this operation clearly exhibit the familiar wastewater diurnal pattern, with distinct dynamics for weekdays and weekends. However, since the filter allows daily components through its pass band, daily fluctuations in wet-weather flows are still very apparent in the output and must be removed before the dry-weather diurnals can be characterized. To remove the diurnal wastewater patterns distorted by rain storms, the precipitation input is used to delineate times of "dry" weather. As an additive step, the filtered signal is separated into daily bins. Simple threshold criteria for magnitude and length of the diurnal pattern for each site are used to remove daily signals influenced by wet-weather flows. This may occur when a rain gauge does not

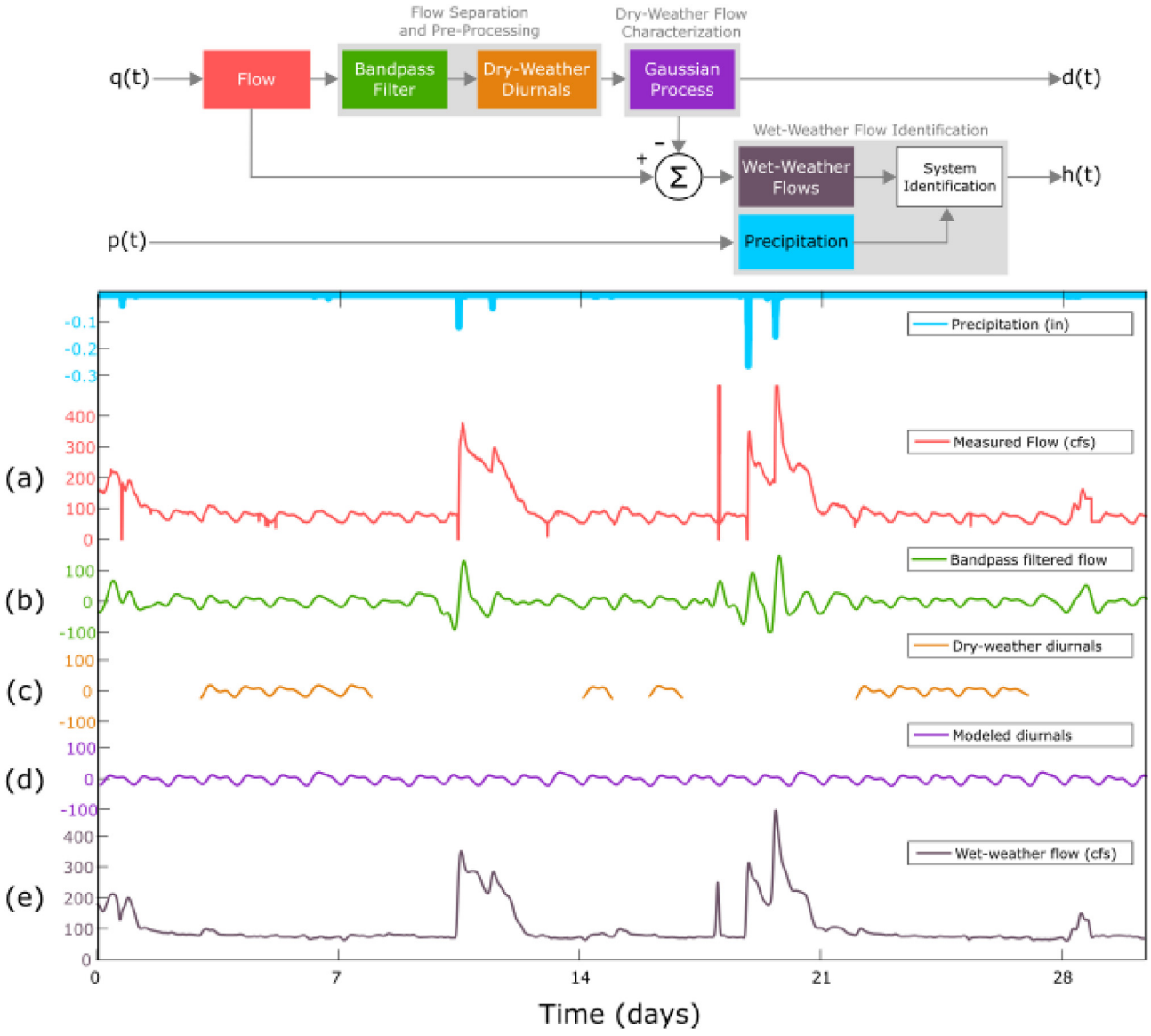

Fig. 2. Proposed identification toolchain applied to measured combined sewer flows.

measure rainfall, but wet-weather flows are still evident. The result of these operations is a sparse dry-weather time series, which only contains those portions of the flow that are not impacted by wet-weather inputs (Fig. 2c). This series is then used to characterize the dry-weather model (Section 3.2.2). Most importantly, perhaps, the dry-weather model can be used to estimate historical dry-weather flows (Fig. 2d), which can then be subtracted from the measured flows to derive a historical estimate of wet-weather flows (Fig. 2e).

3.2.2. Dry-weather flow characterization

As evident in Fig. 2, the task of characterizing dry-weather diurnal patterns in an arbitrary sensor signal is non-trivial due to daily uncertainty in wastewater generation and other factors. If accuracy is desired, there is a need to characterize individual dynamics for each day of the week, and especially the weekends. Furthermore, measurements may not be available on some week days if they have been obscured by wet-weather flows. As such, the approach has to be robust to missing measurements. The high level of non-linearity in the dry-weather diurnal dynamics, as well as their repeating pattern makes them particularly difficult to model as a function of simple inputs. This likely explains why dry-weather signals are often modeled in literature as simple static functions (Saagi et al., 2016), rather than the dynamically changing systems that they really are. Physical models have been proposed as an alternative, but are highly parameterized (Butler and Schütze, 2005; Gernaey et al., 2011).

To address these challenges, our approach represents the dry-weather signal using a non-parametric probability model, which is conditioned on the incomplete dry-weather data received from the first stages of our processing chain. Specifically, we assume that for a measured dry-weather signal y at time t,

$$y(t) = f(t) + \varepsilon, \tag{3}$$

where $f(t)$ is the modeled dry-weather and $\varepsilon \sim \mathcal{N}(0, \sigma_n)$ represents the underlying model and measurement uncertainty, which is normally distributed with a variance σ_n. A powerful model can be obtained by representing the above relation as a Gaussian Process (GP) (Williams and Rasmussen, 2006). Specifically, rather than learning an explicit mapping for $f(t)$, it is possible to characterize the distribution over all (possibly infinite) dry-weather diurnals that describe the observed data. In this setting, the distribution over all dry-weather models is given by the Gaussian Process

$$f(t) \sim \mathcal{GP}(d(t), k(t, t')), \tag{4}$$

which is fully characterized by its mean function $d(t)$ and covariance function

$$k(t, t') = \mathbb{E}[(f(t) - d(t))(f(t') - d(t'))]. \tag{5}$$

Due to their cyclical patterns and irregular shapes, dry-weather diurnal signals are difficult to model using standard linear regression techniques, since the underlying basis functions that describe the model are unknown. Rather than defining an explicit relation that depends on just time, our implementation relies on using the covariance function $k(\cdot, \cdot)$ to determine the degree of *similarity* between points in the dry-weather signal. When making predictions of dry-weather signals at times when no measurements are available, this logic dictates that the missing measurements is most likely similar to its nearest neighbors or to those points that have been observed on the same day during a different week.

This notion can be encoded through the use of specific covariance functions, or *kernels*. Specifically, our models make use of two *kernels*, which are added together to model the covariance of the dry-weather signal. The first is a sinusoidal or *periodic* kernel

$$k_{per}(t, t') = \sigma_p^2 \exp\left[-\frac{2 \sin^2\left(\frac{\pi|t-t'|}{p}\right)}{\ell_p^2} \right] \tag{6}$$

This kernel is characterized by its hyperparameters σ_p,p, and ℓ_p, which must be *learned* before predictions can be made. Given two points t and t' in time, this kernel embeds the notion that similarity between observations is determined according to a repeating pattern. Simply stated as an example: an observation on a Wednesday more closely resembles that made during the same time during another week, rather than a measurement made on a Saturday morning. This also implies that the parameters of the kernel embed meaning about the physical and hydrologic nature of the diurnal wastewater flows, such as periodicity and magnitude, unlike black-box models. The p parameter, for example, encodes how often a signal repeats (daily in our case), while the length-scale parameter ℓ_p determines how closely points are related throughout the day.

Recognizing that the day and time of the week are not the only factors describing the dry-weather wastewater pattern, we also use a rational quadratic kernel $k_{RQ}(\cdot, \cdot)$, which is given by

$$k_{RQ}(t, t') = \sigma_r^2 \left[1 + \frac{|t - t'|^2}{2\alpha\ell_r^2} \right]^{-\alpha}. \tag{7}$$

Once the hyperparameters σ_r,α, and ℓ_r are *learned*, the use of this kernel captures variations in the magnitude of the dry-weather signal that cannot be explained by simple periodicity, but rather by a proximity to neighboring points in time. This includes, but is not limited to, seasonal variations in magnitude or short-term fluctuations (e.g., holidays, sporting events).

While the hyperparameters of these kernels could be manually calibrated, the major task in the use of this probabilistic model relates to using the sensor data to automate this task. This can be accomplished by storing the observed dry-weather measurements and their corresponding weekly time stamps in the vectors $\mathbf{y}$ and $\mathbf{t}$. It can be shown that the posterior distribution over all the dry-weather functions that describe this data is given by applying Bayes' Rule

$$P(f|\mathbf{t}, \mathbf{y}) = \frac{P(f)P(\mathbf{y}|\mathbf{t}, f)}{P(\mathbf{y}|\mathbf{t})}, \tag{8}$$

which, given our assumptions, simplifies to

$$P(f|\mathbf{t}, \mathbf{y}) \sim \mathcal{N}\left(k(t, \mathbf{t})\left[K(\mathbf{t}, \mathbf{t}) + \sigma_n^2 I\right]^{-1}\mathbf{y}, \; k(t, t') - k(t, \mathbf{t})\left[K(\mathbf{t}, \mathbf{t}) + \sigma_n^2 I\right]^{-1} k(\mathbf{t}, t') \right) \tag{9}$$

The hyperparameters can be learned by maximizing the marginal likelihood $P(y|t)$. Unlike in the predictive distribution, no closed form solution exists for this relation. However, a gradient-based optimization algorithm can be used to find the optimal (or near-optimal) hyperparameters by maximizing the log-likelihood, which is given by

$$\log P(\mathbf{y}|\mathbf{t}) = -\frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}| - \frac{n}{2}\log 2\pi. \tag{10}$$

Once model is *learned*, Equation 9 can then be used to make a prediction y^* at a time t^* via

$$P(y^*|t^*,\mathbf{t},\mathbf{y}) \sim \mathcal{N}\ \left(\mathbf{k}(t^*,\mathbf{t})^T\left[K+\sigma_n^2 I\right]^{-1}\mathbf{y},\ k(t^*,t^*)+\sigma_n^2-\mathbf{k}(t^*,\mathbf{t})^T\left[K+\sigma_n^2 I\right]^{-1}\mathbf{k}(t^*,\mathbf{t})\right) \tag{11}$$

3.2.3. Wet-weather flow identification

Once the full historical dry-weather flows are estimated, they are subtracted from the original measurement to derive an estimate of historical wet-weather flows (Fig. 2). This estimate can then be used to identify the input-output relationship between precipitation and wet-weather flow, which our toolchain accomplishes through the use of *System Identification* (Ljung, 1999). In short, System Identification is a form of inverse modeling, where input and output data are provided, a model structure is specified, and the parameters of the model are learned from the data.

In our toolchain, the structure takes the form of a transfer function model (Golnaraghi and Kuo, 2009; Yang and Han, 2006), which represents the frequency (s) response of the flow model (Equation 2) as

$$G(s) = \frac{H(s)}{P(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n}; \tag{12}$$

the input is precipitation data $P(s)$ and output is wet-weather flow data $H(s)$.

The parameters of the differential equation $a_1, \ldots, a_n, b_0, \ldots, b_n$ are placed into the vector θ. For parameter identification, let Z^N denote the N-dimensional measured data set, $h(t)$ denote the measured system output, and $\hat{h}(t|\theta)$ denote the predicted system output given system parameters θ. The prediction error is given by

$$\varepsilon(t,\theta) = h(t) - \hat{h}(t|\theta). \tag{13}$$

The desired parameters maximize the fit between observed and modeled dynamics. Mathematically, we seek to minimize a norm of the prediction error, given as

$$V_N\left(\theta, Z^N\right) = \frac{1}{N}\sum_{t=1}^{N} l(\varepsilon(t,\theta)), \tag{14}$$

where $l(\cdot)$ is a scalar-valued norm function. If the standard choice of $l(\cdot)$ as the quadratic norm is used

$$l(\varepsilon) = \frac{1}{2}\varepsilon^2, \tag{15}$$

an estimate of the system parameters $\widehat{\theta}$ for the measured data set is given by

$$\widehat{\theta} = \widehat{\theta}\left(Z^N\right) = \arg\min_\theta V_N\left(\theta, Z^N\right). \tag{16}$$

To minimize V_N, parameters are iteratively changed using numerical strategies such that

$$\widehat{\theta}_N^{(i+1)} = \widehat{\theta}_N^{(i)} - \mu_N^{(i)}\left[R_N^{(i)}\right]^{-1} V_N'\left(\widehat{\theta}_N^{(i)}, Z^N\right), \tag{17}$$

where V_N' is the gradient of V_N, $R_N^{(i)}$ denotes a matrix to dictate the search direction, and $\mu_N^{(i)}$ is the step size to decrease V_N with each iteration. For the quadratic norm,

$$V_N\left(\theta, Z^N\right) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\varepsilon^2(t,\theta), \tag{18}$$

which implies that

$$V_N'\left(\theta, Z^N\right) = -\frac{1}{N}\sum_{t=1}^{N}\psi(t,\theta)\varepsilon(t,\theta), \tag{19}$$

where $\psi(t,\theta)$ denotes the gradient of $\hat{h}(t|\theta)$ with respect to θ.

To estimate the parameters, during each numerical iteration, the search direction is calculated using each using a Gauss-Newton method. Specifically, the search direction is calculated

$$R_N^{(i)} = V_N''\left(\theta, Z^N\right) = \frac{1}{N}\sum_{t=1}^{N}\psi(t,\theta)\psi^T(t,\theta) - \frac{1}{N}\sum_{t=1}^{N}\psi'(t,\theta)\varepsilon(t,\theta), \tag{20}$$

where $\psi'(t,\theta)$ is the Hessian of $\varepsilon(t,\theta)$. However, determining ψ' may be computationally expensive for each iteration. Rather, we assume that there exists θ_0 such that the prediction errors are independent (i.e., $\varepsilon(t,\theta_0) = e_0(t)$). Near θ_0, Equation 20 can be approximated as

$$V_N''\left(\theta, Z^N\right) \approx \frac{1}{N}\sum_{t=1}^{N}\psi(t,\theta)\psi^T(t,\theta) \triangleq H_N(\theta). \tag{21}$$

This ensures the Hessian estimate (H_N) is positive semidefinite and thus converges to a stationary solution (i.e., the global minimum).

3.3. Implementation

The entirety of the toolchain developed in this paper, which includes the full source code, how-to documentation, and implementation details has been made available on an open-source public web repository (http://github.com/kLabUM/DRIPS). While the authors are not at liberty to share all of the raw sensor data used in this study due to privacy considerations, an anonymized example data set (precipitation and corresponding flow) has been included in the web repository to allow others to evaluate our approach and implementation. Users should also be able to apply the model to their own datasets of rainfall and flows.

For this study, all analyses were carried out on a Windows OS laptop and the software is written in *MATLAB* (2016b edition). The implementation of the real-time system is executed on *MATLAB* as well, but is hosted on a cloud server, specifically an Amazon Web Services (AWS) Linux instance. The low computational overhead of the proposed toolchain does not require large server resources, which permits it to run on a low-cost or free cloud instance.

3.3.1. Data architecture and study area

The study was provided with access to observations from 10 flow measurement sites and 18 precipitation measurement sites, spanning three years (2013–2016) at 5 min measurement resolution. Precipitation data were measured using tipping bucket rain gauges and flow sensors varied across the sewer network, including magnetic, Parshall flume, and ultrasonic flow meters. The extent of study area and location of sensors is shown in Fig. 3, with geographic and infrastructure identifiers removed in compliance with the data agreement. The data were retrieved in real-time using software written in the *Python* language and stored in the *InfluxDB* time series database system (InfluxDB, 2017). This database architecture is optimized for timeseries data, permitting large

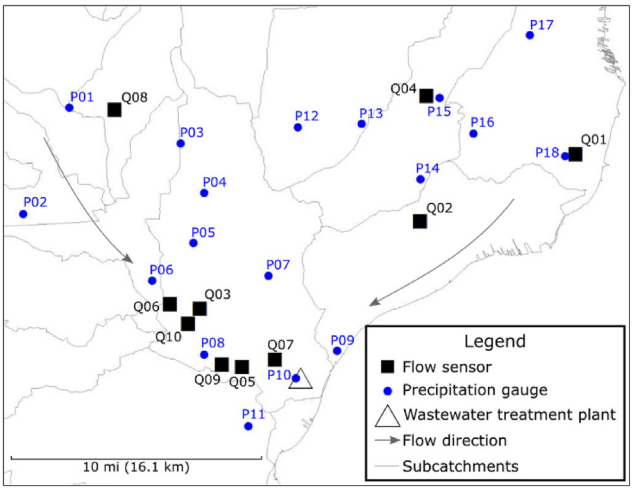

Fig. 3. Study area extent and location of sensors used in the analysis. Subcatchments were delineated based on contributing runoff areas and the pipe system.

series to be seamlessly searched and retrieved using web services and interfaces to popular programming languages, such as *Python* and *MATLAB*. This was particularly important in our approach, which presents an end-to-end, functional, and real-time data process chain. The toolchain was deployed on an AWS instance. A visualization interface was also developed on top of the *Grafana* time series visualization package (Grafana, 2017) and *Google Maps* (Google Maps, 2017). The deployed system not only permits for data to be visualized, but also executes underlying flow models and re-calibrates them based on intervals deemed appropriate in our identification study.

3.3.2. Evaluation

While automated, the processing chain (Fig. 2) for the dry-weather and wet-weather model still depends on input data, which raises two challenges. Firstly, the computational complexity of the GP algorithm scales cubically with the number of input data (Williams and Rasmussen, 2006), while the Gauss-Newton method within System Identification typically has quadratic local convergence rates (Björck, 1996). As such, to improve computational efficiency it is generally in the interest to reduce the number of historical observations that are used to train the models. Secondly, and more importantly, just like all other models, the proposed toolchain is challenged by system uncertainty, which means that the age of the data used to calibrate the model may play a large role in model performance. This is particularly important when considering changing climate patterns, land use, or human inputs, which may operate on unique timescales.

To determine the impact of the size and age of calibration data, various *lookback* periods were evaluated. Wet-weather models were calibrated using one to 24 months of lookback training data. For each site, the three nearest rain gauges were used as inputs into the System Identification procedure. A cross validation was then carried out across a lookback period, whereby System Identification was used on each rain gauge and flow measurement. The model that most accurately predicted the wet-weather response of the remaining storms in the lookback period (*training data*) was then used to make a prediction on a future storm (*testing data*). The decision to train the model on the three closest gauges and then choose the best performing rain-flow pair was guided by a number of factors. Firstly, the choice was motivated by practicality, since in most cases the three closest rain gauges were either located in the same subcatchment as the flow sensor or were close enough to assume some level of rainfall uniformity. Secondly, this specific pairing of gauges with flow meters is actually what the city staff had assigned through their judgment, so it provided a good starting point to begin evaluating the proposed toolchain. It is, of course, possible to train the flow model on every rain gauge, which proved to be out of the scope for our initial evaluation and will be reserved for future studies. The model calibration/identification for each site is carried out in a rolling fashion, whereby the approach automatically re-calibrates or re-trains the wet-weather model based on new measurements. A similar procedure was used to evaluate the dry-weather model performance. To predict dry-weather flows one month in advance, the model was trained across one to 24 month lookback periods.

To evaluate the performance of each model, the fit was quantified using the normalized root mean square error (NRMSE)

$$NRMSE = 1 - \frac{\| x_{ref}(t) - x(:t) \|_2}{\| x_{ref}(t) - avg\left(x_{ref}(t)\right) \|_2}, \tag{22}$$

where x_{ref} denotes the measured data, x denotes the modeled data, and $\| \cdot \|_2$ denotes the 2-norm. Since flow magnitudes vary drastically from site to site, this formulation permits for a relative comparison across the study area. Using this metric, a perfect fit receives a value of one and increasingly poor fits approach a value of negative infinity. An NRMSE of zero would indicate that a mean model (simply taking the average of all historical training data) performs as well as the proposed model.

4. Results

4.1. Dry-weather flow model

When tested against future data, the GP-based approach accurately modeled the dry-weather diurnal patterns in the conduits, as evidenced by NRMSE metrics (SI Table 1), as well as a visual inspection (Fig. 4). In particular, the modeled dynamics closely resembled those of the measured values. The approach even captured the nuanced dynamics of individual days, such as weekends (one diurnal peak) and weekdays (two peaks). The GP model also provided estimates during times of wet weather (middle section of Fig. 4), during which measured values were not available. While the accuracy of these wet-weather diurnals can thus not be calculated, the modeled dynamics did reflect what could intuitively be expected on those days.

The impact of lookback periods (how much data was used to train the model) was unique to each site (Fig. 5 and SI Table 1). For most sites (Q02, Q04, Q05, Q06, Q08, Q09), model performance did not drastically improve with longer lookback periods, showing only marginal improvements when more data was used to train the model. This is also apparent through visual inspection of the modeled dynamics (Fig. 4), which showed very similar model predictions regardless of the duration of measurements used to train the model. One site (Q10) yielded notably better performance for short lookback periods, while the remaining sites (Q01, Q03, Q07) required longer lookback periods for satisfactory model performance. On average, a lookback period of 6–9 months yielded the best performing dry-weather model. Beyond this point, the use of longer periods of historical training data actually resulted in a worse model performance on average.

4.2. Wet-weather flow model

For all sites, a third order wet-weather model structure exhibited the best NRMSE performance. Visual inspection revealed

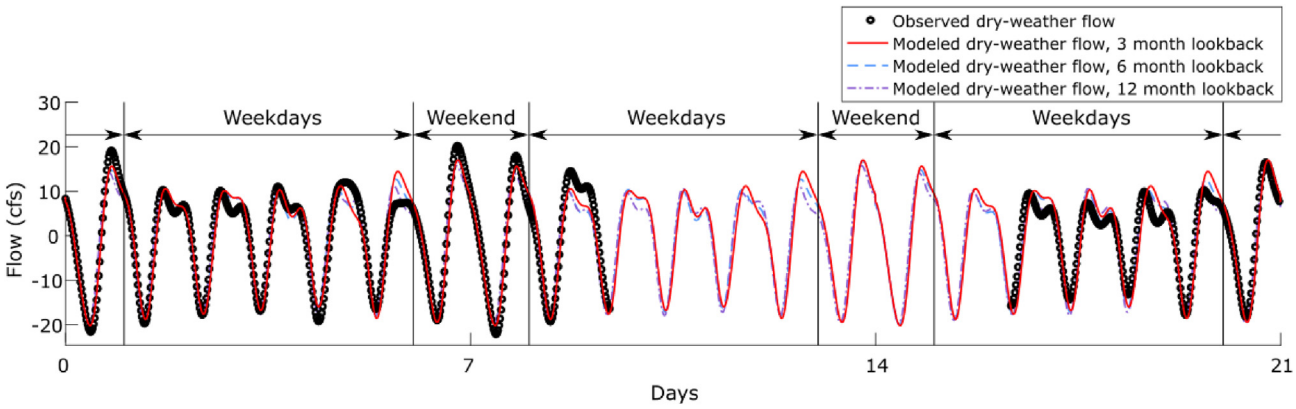

Fig. 4. Predicted dry-weather flow using GP, compared to the observed dry-weather flow (Site Q02) (average NRMSE value of predictions is 0.7427).

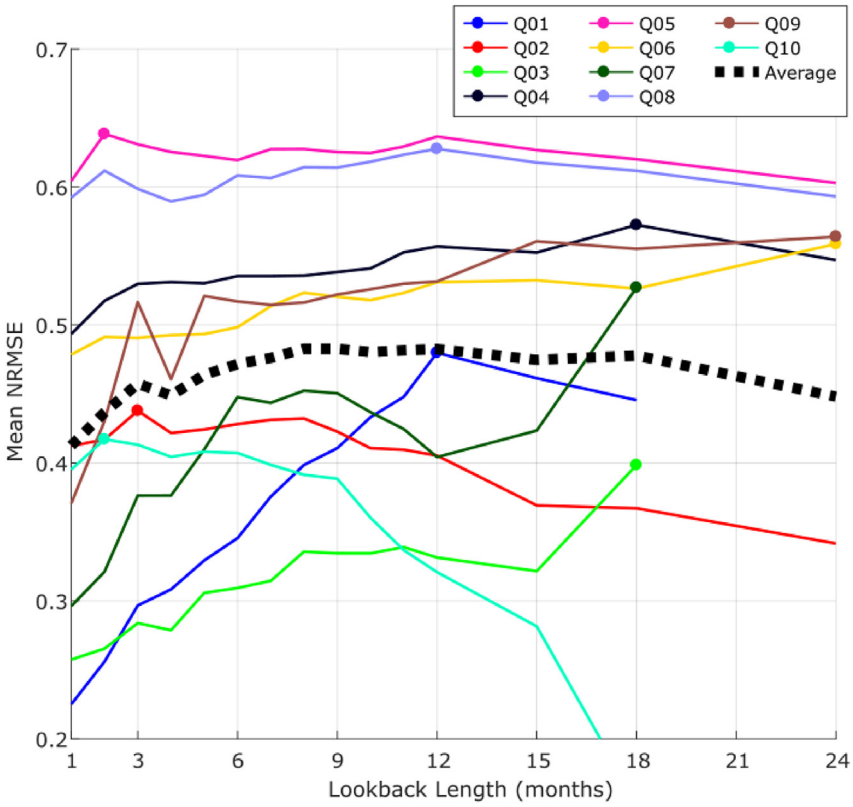

Fig. 5. Dry-weather prediction performance of GP using different training lookback windows. Dot on each curve indicate the lookback period that resulted in the best performing model. Dashed line indicates average model performance.

that the System Identification procedure accurately parameterized the wet-weather model, as seen in the predicted dynamics when given a measured rainfall input (Fig. 6). Generally, the fit of the model improved as the lookback period increased (Fig. 7 and SI Table 2). In other words, when the model was trained using more historical storms, it generally performed better at predicting future events. At some point, however, increasing the lookback period yielded marginal or worse performance. On average, peak model performance was achieved when using 15 months of training data (Fig. 7), after which using more data to train the model actually yielded worse NRMSE performance. For a few sites, the model performed best when using even shorter lookbacks (Q10, Fig. 7 and SI Table 2).

Rainfall dynamics played one of the largest roles in explaining differences in modeled and measured wet-weather flows. In particular, relatively worse model performance often connected to storms during which measured rainfall dynamics and measured flows did not correlate. For example, in Fig. 8a, a storm with distinct rainfall peaks is shown. While the modeled flows also revealed these peaks, the actual measured flows did not exhibit one of the major flow episodes (i.e., approximately mid-way through the third day). While not shown here, the converse was also observed at times, where no rainfall was measured but a wet-weather flow was measured. This is likely due to the rain gauge being outside of the contributing watershed and underscores the significance of spatially variable storms. As expected, the model did not predict

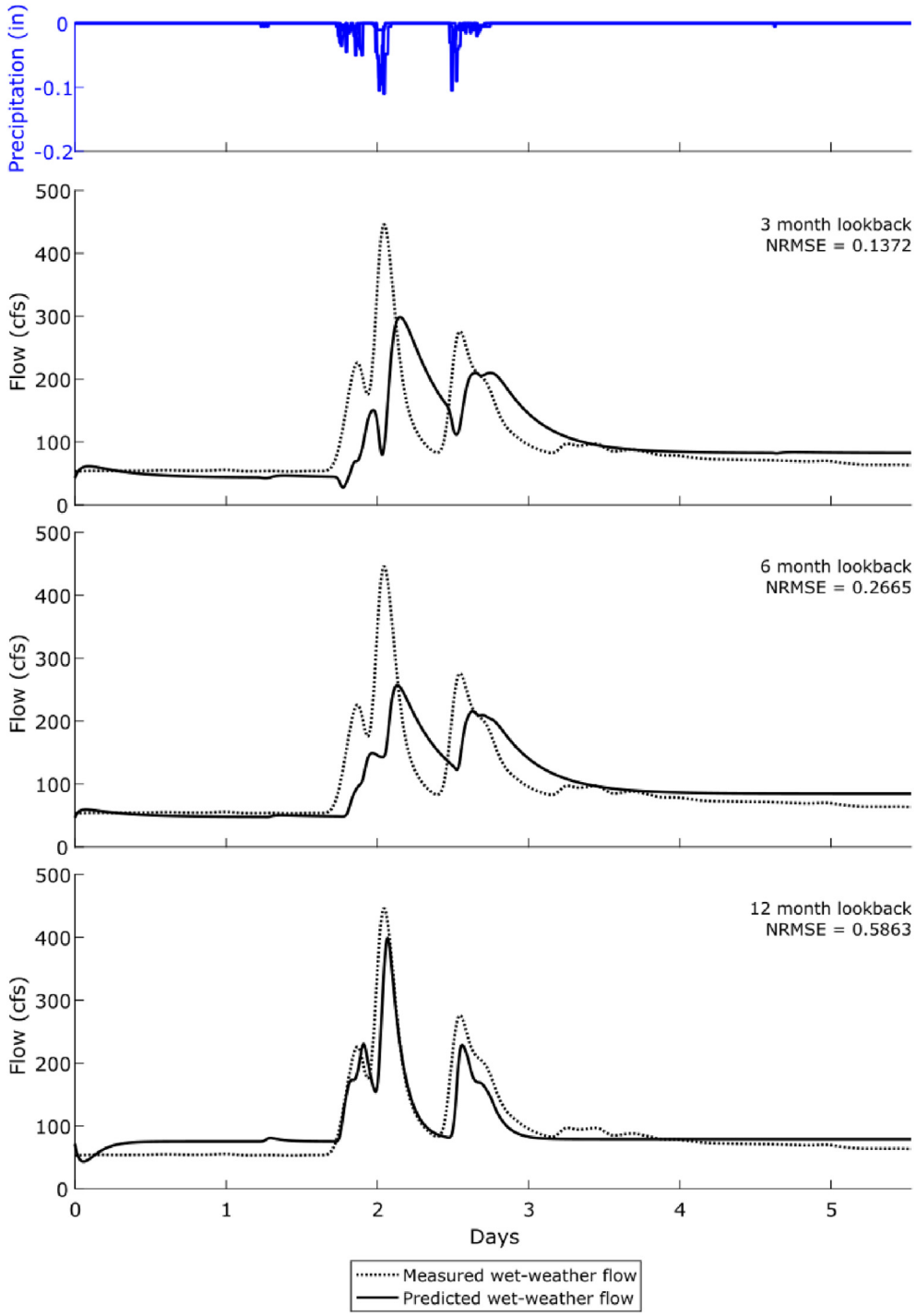

Fig. 6. Predicted wet-weather response during a storm event using System Identification and various lookback windows for model learning, compared to the measured wet-weather flow (Site Q03). The plotted rainfall measurements are those of the gauge that was selected for the model (Section 3.3.2).

these flows during these isolated instances.

Models for five of the ten sites also showed a behavior wherein the modeled dynamics agreed with measurements for most storms, except during large rain events (example Fig. 8b). In these cases, the measured values showed a distinct leveling-off, whereas the model predicted a sharp hydrograph peak. These dynamics could be caused by pipe capacity exceedance or, given the interconnected nature of sewer networks, control actions at upstream infrastructure assets, such as pumping stations, storage dams, and weirs. In this study, it was, however, determined through confirmation with the city that these instances corresponded with flows that exceeded pipe conveyance capacity.

4.3. Combining flows

Once the dry-weather and wet-weather flow models are identified using the proposed toolchain, the combined flows are modeled for a given precipitation input by summing the two

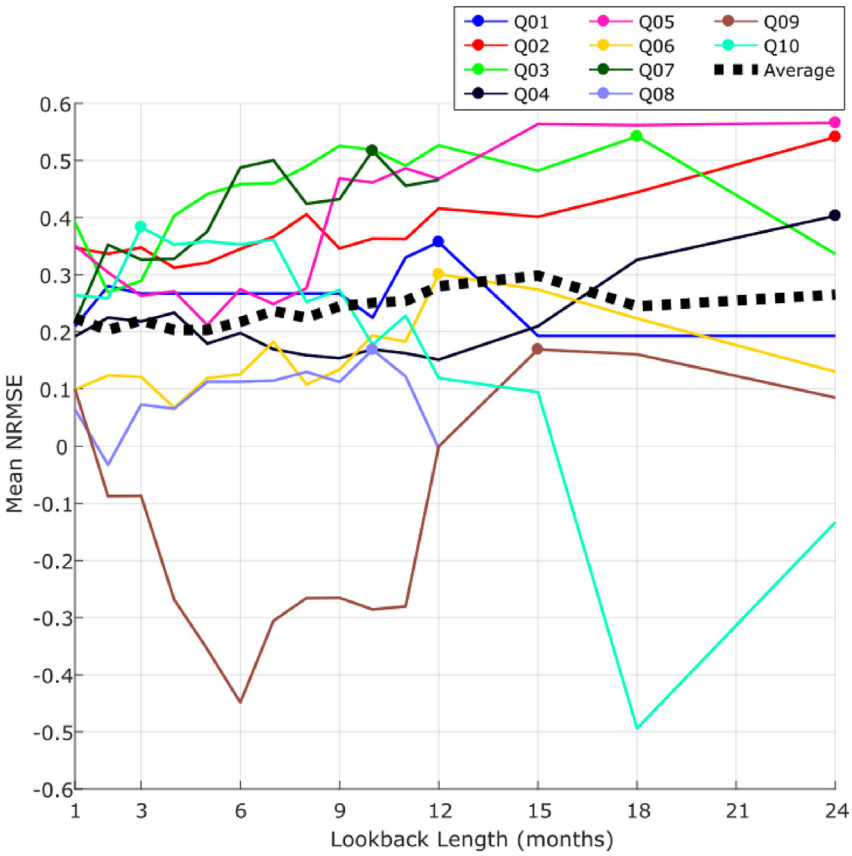

Fig. 7. Impact of model learning lookback window on wet-weather response prediction performance for each site. Dot on each curve indicate the lookback period that resulted in the best performing model. Dashed line indicates average model performance.

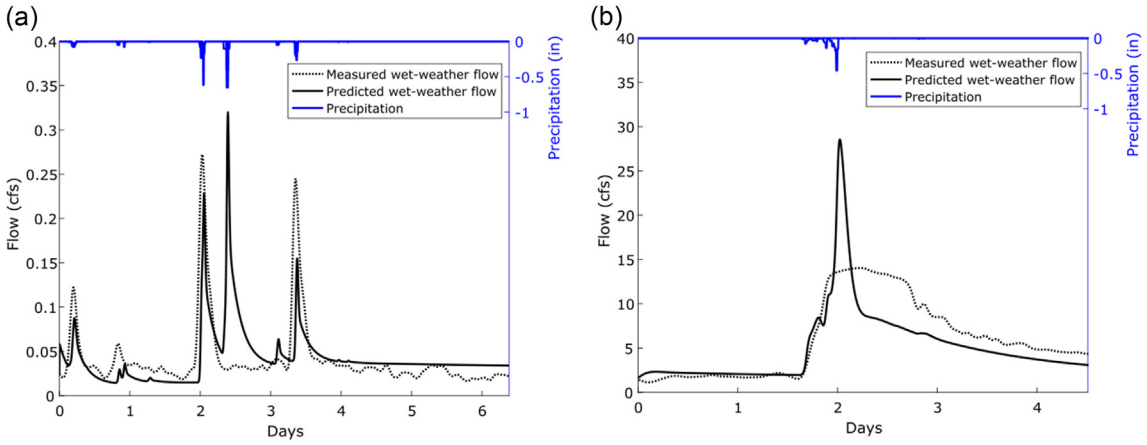

Fig. 8. Wet-weather artifacts of the System Identification procedure. The precipitation data shown here is the rain pair for the given flow site. (a) Example of measured rainfall that does not result in wet-weather response (Site Q06). (b) Instance of pipe capacity exceedance caused by a large storm event (Site Q05).

components. While the performance of the final model is thus clearly underpinned by the performance of the respective sub-models, it is nonetheless worth noting that the final model exhibited good performance in predicting combined flows. For example, as shown in Fig. 9 for site Q05, the combined model closely adhered to the measured flow both during dry weather and wet weather periods, aside from a small deviation measured on day five, where a change in flow was measured but not accompanied by a change in measured rainfall.

5. Discussion

5.1. Dry-weather identification

As indicated by the results, the GP-based model provides a powerful tool by which to predict and describe dry-weather diurnal patterns in urban sewer networks. This is especially evident when considering the complexity of dry-weather dynamics, which not only vary across time, but also between sites. The GP approach

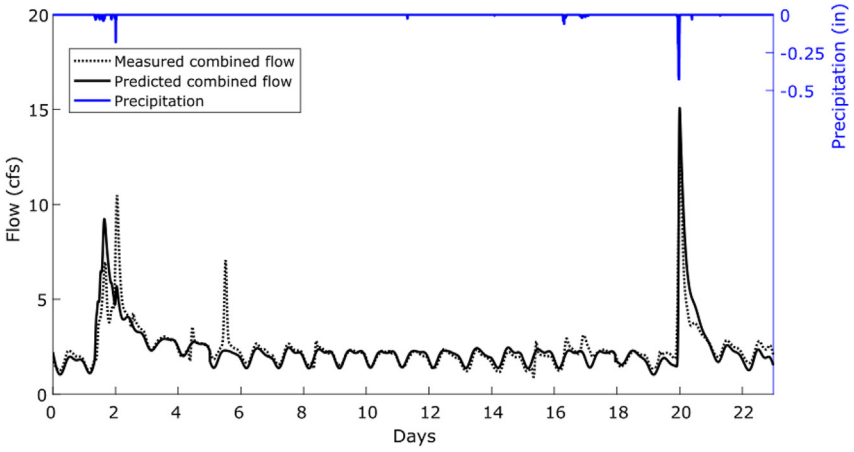

Fig. 9. Measured and predicted combined flow (Site Q05), obtained by combining the forecasts made by the dry-weather and wet-weather models. The precipitation data shown here is the rain pair for the given flow site.

accurately reflected diurnal dynamics of each site using a sparse set of dry-weather observations. This included the nuances of dry-weather diurnal shapes, magnitudes, and weekly trends. This should make it a very powerful tool for modeling and predicting components of sewer flows that have generally been very difficult to model due to variability of the wastewater generation patterns. It also underscores the need for a flexible modeling approach that is able to learn from the site-specific data, rather than being confined to assumed shapes or dynamics. While outside of the scope of this study, future work will include explicitly quantifying groundwater infiltration or baseflows from the dry-weather signal. This can be accomplished by subtracting the diurnal estimates of the Gaussian Process from the overall flow signal during dry-weather periods.

The major take-away from the dry-weather analysis relates to the role of the training data: generally speaking, the best model performance will be achieved when the model is trained on enough, but not too many historical observations. As expected, an insufficient amount of training data (in our case fewer than three months, on average) may lead to relatively poor model performance due to overfitting. Once enough training data are obtained, however, increasing the length of the lookback period did not result in drastically improved modeling performance. The notion of "enough" data will ultimately be guided by practical considerations and modeler judgment. Nonetheless, for the sites studied in this paper a clear guideline would focus on choosing the lookback period with the best NRMSE criterion, or the point at which NRMSE improves only marginally with longer lookback. This underscores the need to acknowledge uncertainty in the dry-weather diurnal dynamics, which may be driven by changing wastewater generation patterns, seasonal impacts, or new connections or repair to the system, indicating that older dry-weather flow signals no longer represent the current dry-weather diurnal patterns. While the best performing dry-weather models required about 3–6 months of training data, there was significant variability around this average, suggesting that changes in long-term diurnal patterns are unique to each site. As such, this underscores the importance of a flexible calibration/identification chain, such as the one presented in this paper, which will continually tune the dry-weather model to ensure that the most relevant measurements are used to inform forecasting.

One benefit of this dry-weather model is the ability to *hindcast*, wherein the dry-weather model is used to project historical estimates of dry-weather flows. These estimates can then be subtracted from the time-series of historical flows to reconstruct the series of wet-weather flows. Without a good dry-weather model, the reconstruction of wet-weather flows becomes very difficult, if not impossible. This is especially true for sites at which dry-weather and wet-weather flows share similar timescales and magnitudes. As was shown (Fig. 2), the resulting wet-weather signal becomes much more clear after this operation, especially during small storm events, which would have otherwise been obscured in the original time series. While improved historical estimates of dry-weather flows are useful in practical applications, such as billing, they are perhaps even more valuable because they can then be coupled with the System Identification procedure to reliably learn the underlying wet-weather model.

5.2. Wet-weather identification

The performance of the wet-weather model was highly underpinned by the availability of training data. The wet-weather lookback period is a proxy for the number of storms used for training. Given the extent of the analysis area, this number was the same for all sites. Unlike in the case of the dry-weather model, which depends on readily-available dry weather observations, the number of available storms to train the wet-weather model can often be limited, which explains why the wet-weather model requires longer periods of training data compared to the dry-weather model. In fact, the number of storms in a particular time period can vary widely based on the time of year; for instance, for a nine month period, the number of storms included varied from 5 to 17 storm events. There appears to be a minimum data requirement for the procedure to yield satisfactory results. For most sites, the System Identification procedure required at least nine months of data to result in the best relative performance, which suggests that the value of sensor data increases with the deployment period. As such, investments into sensor networks should be cognizant of this requirement, especially if prediction of wet-weather flows is desired.

As in the case of the dry-weather flows, the biggest take-away relates to the discovery that including more data beyond a certain lookback point can actually degrade the performance of the model. While this optimal period varied for each site, this feature was evident across almost all sites. In our study, no physiograhic features of the urban environment or storm characteristics revealed why some sites require more or less training data than others. These features included position of sensors in the watershed (i.e., upstream, downstream), size of the noncontributing sewersheds,

as well as the precipitation intensity and peak rainfall of storm events. This again suggests that system uncertainty plays a large role in model performance, even when forecasting wet-weather flows. As such, the ability to predict future flows depends on re-calibration of the model. The toolchain presented herein offers the opportunity to autonomously adjust this optimal lookback period for each site, and thus ensure that each future forecast is based on the most relevant time record of historical measurements. The approach thus implicitly addresses the need to re-calibrate and adapt the model as the overall collection system changes due to changing land uses, climate, or infrastructure upgrades. This will be particularly important in locations where changes in climate, storms, and infrastructure need to be accounted for to ensure longevity and accuracy of a model.

Beyond sensor data availability, knowledge of the physical system will also play a large role in the long-term efficacy of the proposed toolchain. Large storms that cause pipe capacity exceedance will result in non-linear system behavior that may not be fully represented by our wet-weather model. As illustrated in Fig. 8b, if the wet-weather model is trained on a smaller storm, the linear transfer function will result in a predicted hydrograph that may exceed the capacity of conduits or sensing limits. In such rare instances, the model will still have utility, as it will be able to predict instances of overflows. Also, as mentioned in Section 4.2, the impact of control actions on flow dynamics must be considered in some systems. The identification approach used here provides the flexibility to learn a dynamical model with exogenous inputs (Ljung, 1999), which will be explored in future work.

As with all modeling approaches, sensor quality will play a large role in predictive outcomes. Sensor background noise will likely not be the major challenge, as our toolchain addresses noise via filtering and probabilistic estimation. Extensive studies on real-time data quality evaluations are available in literature (Campisano et al., 2013; Métadier and Bertrand-Krajewski, 2011; van Bijnen and Korving, 2008). Beyond formal methods there will still be a need to subjectively evaluate, on a case-by-case basis, whether data quality is sufficient for the model to provide adequate forecasts, especially for critical operations. Ultimately, the quality of the forecasts will be implicitly determined by the quality of the input data. Once calibrated on reliable data, if measurements deviate too much from forecasts, as measured by specific magnitude or probabilistic bounds, alerts could be sent to operators to maintain or repair field-deployed sensors.

Additionally, if the toolchain is to be entirely automated across large systems, a degree of sensor redundancy may be required, which will allow for a more accurate assessment of any given situation (e.g., broken sensor vs. clogged pipe). Redundancy will be critical across inputs (rain gauges) as well as outputs (flows), with the latter offering more flexibility because short-term flow forecasts can still be made given rainfall alone. Practical guidelines regarding the challenge of sensor quality will be best learned over time by observing how operators trust, interpret, and use the forecasts offered by toolchains such as the one presented here.

The need to understand urban rainfall variability will play, perhaps, the largest role in improving model predictions. This is not only true for the models presented in this paper, but for any modeling toolchain. This will be particularly true for large collection systems, such as the one studied in this paper. Since there was no clear correlation between model performance and the distance of a flow measurement to its corresponding rain gauge, the role of micro-climate thus needs to be accounted for as we develop the next generation of high-resolution urban water models. Much of this is expected to improve as more rain gauges are deployed and as high-resolution radar data become available.

6. Conclusions

Our study underscores the need to understand the uncertain and dynamic nature of urban water systems. The various components that drive flows in urban sewer and storm networks change across their own unique timescale. To capture these changes more measurements are needed to characterize the spatiotemporal variability in flows. This will be enabled by the rapid rise in sensors that are now being deployed throughout cities. Sufficient historical observations must be available to train the models, while also ensuring that data that are *"too old"* are not used in the identification of the model. More importantly, the duration of the training periods was unique for each site and each sub-model. The need to train the dry-weather model on a different amount of data than the wet-weather model suggests that a conventional, broad-sweeping calibration approach would struggle to accomplish such a task. As such, individual models of each flow component should be uniquely calibrated and re-calibrated for each site, which is well within the flexibility and computational efficiency of the proposed toolchain. The accuracy of predicted flows may thus depend on the ability to autonomously re-calibrate a model, perhaps more so than the model itself. This may permit the use of simpler models to predict flows, as opposed to more complex models that try to capture higher degrees of complexity.

The proposed toolchain has the potential to serve as an operational tool for sewer system control, overflow prediction, and treatment control. Since the model explicitly separated dry-weather and wet-weather flows, it will also be applicable in the modeling of separated stormwater or sewer systems. More importantly, the formulation of the toolchain as a dynamical system opens up the possibility of applying a suite of robust control algorithms that could be used to guide the real-time operation of the sewer system. While this is beyond the scope of this paper, it was a major motivation for the specific formulation presented herein. Future work will thus focus on extending the utility of our proposed models into a holistic and real-time control toolchain.

Acknowledgements

Funding for this study was provided by the REFRESCH program at the University of Michigan. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1256260. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2017.08.065.

References

Andreas, A., 1993. Digital Filters, Analysis, Design, and Applications, 2nd Edition. McGraw-Hill, Inc.

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. Environ. Model. Softw. 54, 88–107.

Banik, B.K., Alfonso, L., Di Cristo, C., Leopardi, A., Mynett, A., 2017. Evaluation of different formulations to optimally locate sensors in sewer systems. J. Water Resour. Plan. Manag. 143 (7), 04017026.

Beeneken, T., Erbe, V., Messmer, A., Reder, C., Rohlfing, R., Scheer, M., Schuetze, M., Schumacher, B., Weilandt, M., Weyand, M., 2013. Real time control (RTC) of urban drainage systems - a discussion of the additional efforts compared to conventionally operated systems. Urban Water J. 10 (5), 293–299.

Björck, Å., 1996. Numerical Methods for Least Squares Problems. SIAM.

Borsanyi, P., Benedetti, L., Dirckx, G., De Keyser, W., Muschalla, D., Solvi, A.-M.,

Vandenberghe, V., Weyand, M., Vanrolleghem, P.A., 2008. Modelling real-time control options on virtual sewer systems. J. Environ. Eng. Sci. 7 (4), 395–410.
Butler, D., Schütze, M., 2005. Integrating simulation models with a view to optimal control of urban wastewater systems. Environ. Model. Softw. 20 (4), 415–426.
Campisano, A., Cabot Ple, J., Muschalla, D., Pleau, M., Vanrolleghem, P., 2013. Potential and limitations of modern equipment for real time control of urban wastewater systems. Urban Water J. 10 (5), 300–311.
Dawson, C., Wilby, R., 2001. Hydrological modelling using artificial neural networks. Prog. Phys. Geogr. 25 (1), 80–108.
Doglioni, A., Primativo, F., Laucelli, D., Monno, V., Khu, S.-T., Giustolisi, O., 2009. An integrated modelling approach for the assessment of land use change effects on wastewater infrastructures. Environ. Model. Softw. 24 (12), 1522–1528.
El-Din, A.G., Smith, D.W., 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. Water Res. 36 (5), 1115–1126.
Ellis, J.B., 2001. Sewer infiltration/exfiltration and interactions with sewer flows and groundwater quality. In: 2nd International Conference Interactions between Sewers, Treatment Plants and Receiving Waters in Urban Areas–Interurba II, pp. 19–22.
Fletcher, T., Andrieu, H., Hamel, P., 2013. Understanding, management and modelling of urban hydrology and its consequences for receiving waters: a state of the art. Adv. Water Resour. 51, 261–279.
Gernaey, K.V., Flores-Alsina, X., Rosen, C., Benedetti, L., Jeppsson, U., 2011. Dynamic influent pollutant disturbance scenario generation using a phenomenological modelling approach. Environ. Model. Softw. 26 (11), 1255–1267.
Golnaraghi, F., Kuo, B.C., 2009. Automatic Control Systems, 9th Edition, Vol. 2. Wiley.
Google Maps, 2017. URL http://maps.google.com.
Grafana, 2017. URL http://grafana.org/.
Häck, M., Wiese, J., 2006. Trends in instrumentation, control and automation and the consequences on urban water systems. Water Sci. Technol. 54 (11-12), 265–272.
Haykin, S.S., 1999. Neural Networks: a Comprehensive Foundation, 2nd Edition. Prentice Hall.
Hill, D., Kerkez, B., Rasekh, A., Ostfeld, A., Minsker, B., Banks, M.K., 2014. Sensing and cyberinfrastructure for smarter water management: the promise and challenge of ubiquity. J. Water Resour. Plan. Manag. 140 (7), 1–3.
InfluxDB, 2017. URL https://www.influxdata.com/time-series-platform/influxdb/.
Karpf, C., Krebs, P., 2011. Quantification of groundwater infiltration and surface water inflows in urban sewer networks based on a multiple model approach. Water Res. 45 (10), 3129–3136. http://www.sciencedirect.com/science/article/pii/S0043135411001333.
Kisi, O., Shiri, J., Tombul, M., 2013. Modeling rainfall-runoff process using soft computing techniques. Comput. Geosci. 51, 108–117. http://www.sciencedirect.com/science/article/pii/S0098300412002257.
Kurth, A., Saul, A., Mounce, S., Shepherd, W., Hanson, D., 2008. Application of artificial neural networks (ANNs) for the prediction of CSO discharges. In: 11th International Conference on Urban Drainage.
Li, X., Zhou, F., Lodewyk, S., 2010. Applications of artificial neural networks in urban water system. In: Proceedings of Watershed Management 2010: Innovations in Watershed Management under Land Use and Climate Change.
Liu, Y., Bralts, V.F., Engel, B.A., 2015. Evaluating the effectiveness of management practices on hydrology and water quality at watershed scale with a rainfall-runoff model. Sci. Total Environ. 511, 298–308.
Ljung, L., 1999. System Identification: Theory for the User, 2nd Edition. Prentice Hall.
Löwe, R., Vezzaro, L., Mikkelsen, P.S., Grum, M., Madsen, H., 2016. Probabilistic runoff volume forecasting in risk-based optimization for RTC of urban drainage systems. Environ. Model. Softw. 80, 143–158.
Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Model. Softw. 15 (1), 101–124.
Meirlaen, J., Van Assel, J., Vanrolleghem, P., 2002. Real time control of the integrated urban wastewater system using simultaneously simulating surrogate models. Water Sci. Technol. 45 (3), 109–116.
Métadier, M., Bertrand-Krajewski, J.-L., 2011. From mess to mass: a methodology for calculating storm event pollutant loads with their uncertainties, from continuous raw data time series. Water Sci. Technol. 63 (3), 369–376.
Mounce, S., Shepherd, W., Sailor, G., Shucksmith, J., Saul, A., 2014. Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data. Water Sci. Technol. 69 (6), 1326–1333.
Neshaei, S., Ahmadnejad, A., Yousefi, F., Ghanbarpour, F., 2017. Estimating groundwater and rainfall infiltration into sewerage. Int. J. Sustain. Dev. Plan. 12 (1), 185–193.
Pawlowski, C., Rhea, L., Shuster, W., Barden, G., 2013. Some factors affecting inflow and infiltration from residential sources in a core urban area: case study in a Columbus, Ohio, neighborhood. J. Hydraulic Eng. 140 (1), 105–114.
Risholt, L., Schilling, W., Erbe, V., Alex, J., 2002. Pollution based real time control of wastewater systems. Water Sci. Technol. 45 (3), 219–228.
Rosén, C., 2001. A Chemometric Approach to Process Monitoring and Control - with Applications to Wastewater Treatment Operation (Ph.D. thesis). Lund University.
Ruano, A.E., 2005. Intelligent Control Systems Using Computational Intelligence Techniques, Vol. 70. IET.
Saagi, R., Flores-Alsina, X., Fu, G., Butler, D., Gernaey, K.V., Jeppsson, U., 2016. Catchment & sewer network simulation model to benchmark control strategies within urban wastewater systems. Environ. Model. Softw. 78, 16–30.
Schütze, M., Butler, D., Beck, B.M., 2011. Modelling, Simulation and Control of Urban Wastewater Systems. Springer Science & Business Media.
Schütze, M., Campisano, A., Colas, H., Schilling, W., Vanrolleghem, P.A., 2004. Real time control of urban wastewater systems - where do we stand today? J. Hydrol. 299 (3), 335–348.
Seggelke, K., Löwe, R., Beeneken, T., Fuchs, L., 2013. Implementation of an integrated real-time control system of sewer system and waste water treatment plant in the city of Wilhelmshaven. Urban Water J. 10 (5), 330–341.
Seggelke, K., Rosenwinkel, K.-H., 2002. Online-simulation of the WWTP to minimise the total emission of WWTP and sewer system. Water Sci. Technol. 45 (3), 101–108.
Smith, J., Eli, R.N., 1995. Neural-network models of rainfall-runoff process. J. Water Resour. Plan. Manag. 121 (6), 499–508.
Solomatine, D.P., Dulal, K.N., 2003. Model trees as an alternative to neural networks in rainfall-runoff modelling. Hydrol. Sci. J. 48 (3), 399–411. http://dx.doi.org/10.1623/hysj.48.3.399.45291.
Todini, E., 2007. Hydrological catchment modelling: past, present and future. Hydrol. Earth Syst. Sci. 11 (1), 468–482.
van Bijnen, M., Korving, H., 2008. Application and results of automatic validation of sewer monitoring data. In: Proceedings on 11th International Conference on Urban Drainage. Edinburgh, Scotland, UK.
van Daal, P., Gruber, G., Langeveld, J., Muschalla, D., Clemens, F., 2017. Performance evaluation of real time control in urban wastewater systems in practice: review and perspective. Environ. Model. Softw. 95, 90–101.
Vanrolleghem, P.A., Benedetti, L., Meirlaen, J., 2005. Modelling and real-time control of the integrated urban wastewater system. Environ. Model. Softw. 20 (4), 427–442.
Vaze, J., Post, D., Chiew, F., Perraud, J.-M., Viney, N., Teng, J., 2010. Climate non-stationarity–validity of calibrated rainfall–runoff models for use in climate change studies. J. Hydrol. 394 (3–4), 447–457.
Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning. The MIT Press, 2(3).
Yang, Z., Han, D., 1 2006. Derivation of unit hydrograph using a transfer function approach. Water Resour. Res. 42 (1) http://dx.doi.org/10.1029/2005WR004227.