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 time-scales 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