Environmental Science Water Research & Technology

TUTORIAL REVIEW

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Environ. Sci.: Water Res. Technol.*, 2018, **4**, 346

Open storm: a complete framework for sensing and control of urban watersheds†

Matthew Bartos, * Brandon Wong and Branko Kerkez

Received 19th September 2017, Accepted 22nd December 2017

DOI: 10.1039/c7ew00374a

rsc.li/es-water

Leveraging recent advances in technologies surrounding the *Internet of Things*, "smart" water systems are poised to transform water resources management by enabling ubiquitous real-time sensing and control. Recent applications have demonstrated the potential to improve flood forecasting, enhance rainwater harvesting, and prevent combined sewer overflows. However, adoption of smart water systems has been hindered by a limited number of proven case studies, along with a lack of guidance on how smart water systems should be built. To this end, we review existing solutions, and introduce *open storm*—an open-source, end-to-end platform for real-time monitoring and control of watersheds. *Open storm* includes (i) a robust hardware stack for distributed sensing and control in harsh environments (ii) a cloud services platform that enables system-level supervision and coordination of water assets, and (iii) a comprehensive, web-based "how-to" guide, available on open-storm.org, that empowers newcomers to develop and deploy their own smart water networks. We illustrate the capabilities of the *open storm* platform through two ongoing deployments: (i) a high-resolution flash-flood monitoring network that detects and communicates flood hazards at the level of individual roadways and (ii) a real-time stormwater control network that actively modulates discharges from stormwater facilities to improve water quality and reduce stream erosion. Through these case studies, we demonstrate the real-world potential for smart water systems to enable sustainable management of water resources.

Water impact

A new generation of autonomous technologies promises to transform the way we manage watersheds. Real-time analytics and control will reduce stormwater pollution, protect aquatic ecosystems, and enable high-resolution flood forecasting. We present an end-to-end framework for building "smart" watersheds, including an open-source hardware/software platform, and a community-driven "how-to" guide that will empower newcomers to implement their own "smart" water systems.

1 Introduction

Advances in wireless communications and low-power sensing are enabling a new generation of "smart cities," which promise to improve the performance of municipal services and reduce operating costs through real-time analytics and control.[1] While some applications of "smart" infrastructure have received a great deal of attention—such as autonomous vehicles,[2,3] energy grid management,[3] and structural health monitoring[3,4]—integration of these technologies into water systems has lagged behind. However, "smart" water systems offer new inroads for dealing with many of our most pressing urban water challenges, including flash flooding, aquatic ecosystem degradation, and runoff pollution. The goal of this paper is to provide an end-to-end blueprint for the next generation of autonomous water systems, with a particular focus on managing urban stormwater. Towards this goal, we introduce *open storm*, an open source framework that combines sensing, real-time control, wireless communications, web services and domain-specific models. We illustrate the potential of *open storm* through two real-world case studies: 1) a 2200 km^2 wireless flood forecasting network in Texas, and 2) an 11 km^2 real-time stormwater control network in Michigan. Most importantly, to encourage broader adoption by the water resources community, this paper is accompanied by extensive ESI† on open-storm.org, including videos, photos, source code, hardware schematics, assembly guides, and deployment instructions. These materials make it possible for newcomers to implement their own "smart" stormwater systems, without extensive experience in programming or embedded systems design.

2350 Hayward, 2044 GG Brown, Ann Arbor, Michigan 48109-2125, USA. E-mail: mdbartos@umich.edu; Tel: +1 734 647 0727

† Electronic supplementary information (ESI) available: open-storm.org. See DOI: 10.1039/c7ew00374a.

2 Background

2.1 Motivation

Effective management of water supply and water excess are some of the largest engineering problems faced by cities today,[5] and in the wake of rapid urbanization, aging infrastructure, and a changing climate, these challenges are expected to intensify in the decades to come.[6,7] Floods are the leading cause of severe weather fatalities worldwide, accounting for roughly 540 000 deaths between 1980 and 2009.[8] Furthermore, large quantities of metals, nutrients, and other pollutants are released during storm events, making their way *via* streams and rivers into lakes and coastal zones.[9,10] The need to manage pollutant loads in stormwater has persistently been identified as one of our greatest environmental challenges.[11] To contend with these concerns, most communities maintain dedicated gray infrastructure (pipes, ponds, basins, wetlands, *etc.*) to convey and treat water during storm events. However, many of these systems are approaching the end of their design life.[12] At the same time, stormwater systems are being placed under greater stress due to larger urban populations, changes in land use, and the increasing frequency of extreme weather events.[5,7] In some communities, stormwater and wastewater systems are combined, meaning that they share the same pipes. For these systems, large storms often lead to combined sewer overflows, which release viruses, bacteria, nutrients, pharmaceuticals, and other pollutants into estuaries downstream.[13] When coupled with population stressors, it comes as little surprise that the current state of stormwater infrastructure in the United States has been given a near-failing grade by the American Society of Civil Engineers.[14]

Engineers have traditionally responded to increasing demands on stormwater systems by expanding and constructing new *gray* infrastructure. However, the upsizing of pipes and storage elements can prove expensive, time-consuming, and may even result in deleterious long-term side effects. Benefits from stormwater conveyance facilities can be diminished if individual sites are not designed in a global context. Even when best management practices are followed, discharges from individual sites may combine to induce downstream flows that are more intense than those produced under unregulated conditions.[15] Without system-level coordination, gray infrastructure expansion may lead to overdesigned solutions that adversely impact flooding, increase stream erosion, and impair water quality.[16] In response to these concerns, *green* infrastructure (GI) has been proposed as an alternative to traditional "steel and concrete" stormwater solutions. These systems use smaller, distributed assets—such as bioswales, green roofs and rain gardens—to condition flows and improve water quality. However, recent research has raised questions about the scalability and maintenance requirements of green infrastructure.[17] Regardless of the choice between "gray" or "green", new construction is limited by cost, and often cannot keep pace with evolving community needs. To preserve watershed and ecological stability, there is an urgent need to incorporate systems thinking into stormwater designs and to engineer solutions that can optimize stormwater system performance—not only for individual sites, but for entire watersheds.

2.2 The promise of sensing and control

"Smart" water systems promise to improve the study and management of water resources by extending monitoring and control beyond centralized facilities and into watersheds as a whole. With increased access to inexpensive sensors and wireless communications, the feasibility of deploying and maintaining large sensor networks across urban landscapes is now within reach for many public utilities and research groups. While many of the technologies have existed for some time, it was not until the integration of wireless sensor networks with web services (*i.e.* the *Internet of Things*) that large networks consisting of hundreds or thousands of heterogeneous devices could be managed reliably.[18] This in turn has enabled watersheds to be studied at spatial and temporal scales that were previously unattainable. By densely instrumenting urban watersheds, researchers can finally begin to understand the complex and spatially variable feedbacks that govern water flow and quality across the built environment. A system-level understanding of urban watershed dynamics will provide decision makers with actionable insights to alert the public, and improve stewardship of water resources.

Beyond new insight gained through sensing, the ability to dynamically regulate water levels across a watershed will reduce flooding, preserve riparian ecosystems, and allow for distributed treatment of stormwater. While these functions were previously achieved only through construction of static gray infrastructure or centralized treatment facilities, the addition of remotely-controlled valves and pumps promises to realize the same benefits while at the same time reducing costs, expanding coverage, and allowing system performance to scale flexibly with changing hydrologic conditions. Adding valves to existing stormwater facilities, for instance, can extend hydraulic retention time, thereby promoting the capture of sediment-bound pollutants.[19,20] Modulation of flows (hydrograph shaping) may reduce erosion at downstream locations by ensuring that discharges do not exceed critical levels.[19] More fundamentally, distributed control will enable operators to coordinate outflows from stormwater sites (tens to hundreds) across an entire city. Along with reducing flooding, this will allow water managers to utilize the latent treatment capacity of existing ponds and wetlands—effectively allowing a watershed to function as a distributed wastewater treatment plant.[20]

Such a vision for "smart" stormwater systems is no longer limited by technology. Rather, adoption of smart water systems has been hindered by (i) a reliance on proprietary technologies, (ii) a lack of proven case studies, and (iii) an absence of end-to-end solutions that are specifically designed and tested for water resources applications. To enable truly holistic management and control, there is an urgent need to combine modern technologies with domain knowledge from water sciences—something which present solutions do not address or make transparent. These solutions are reviewed next, after which the *open storm* framework is introduced as an end-to-end blueprint for "smart" water systems management. This open-source framework combines low-power wireless hardware with modern cloud computing services and domain-specific applications to enable scalable, real-time control of urban water systems.

3 Existing technologies

Real-time sensing and control of water infrastructure is not a new idea. Supervisory control and data acquisition (SCADA) systems have long been used to monitor and control critical water infrastructure.[21] In addition to traditional SCADA systems, there has been a recent explosion in the development of wireless sensor networks (WSNs) for water resources management. While these technologies have made great strides in enabling monitoring and control of water systems, a lack of end-to-end solutions has inhibited system-scale management of watersheds. In this section, we review existing technological solutions for water system monitoring and control, and describe how *open storm* advances the state of the art by providing the first open source, end-to-end solution for distributed water systems management.

3.1 SCADA systems

Most water utilities use SCADA systems to manage the conveyance, treatment and distribution of water.[21] These systems comprise collections of devices, communication protocols, and software that enable remote monitoring and control of water assets.[21] Most commonly applied in water distribution systems, SCADA systems typically monitor parameters that indicate service quality—such as flows, pressures, and chemical concentrations—and then use this information to control the operation of pumps and valves in real-time.[21] Control may be manual or automatic, and in some cases may integrate optimization algorithms, decision support systems and advanced control logic.[21] While legacy SCADA systems remain popular among water utilities, they suffer from limitations in three major areas: interoperability, scalability and security.

Perhaps the most critical limitation of legacy SCADA systems is the lack of interoperability between systems, reliance on proprietary protocols, and non-extensible software.[22] Traditional SCADA systems are often isolated and incapable of intercommunication.[22] Systems that manage water in one municipality, for instance, may be incapable of communicating with those in another municipality, despite sharing the same service area. Moreover, different SCADA systems within the same jurisdiction may also be isolated, meaning that management of stormwater systems may not in any way inform the operation of wastewater treatment facilities downstream. This lack of communication between water management architectures makes it difficult to coordinate control actions at the watershed scale. Proprietary SCADA systems are also often unable to interface with modern software layers, like Geographic Information Systems (GIS), network analysis software, or hydrologic models.[22] For this reason, SCADA-based control often cannot take advantage of modern domain-specific tools that would enable system-scale optimization of watershed resources.

The capacity of SCADA systems to implement watershed-scale control is also limited by a lack of spatial coverage. Due to their large power footprint and maintenance requirements, traditional SCADA systems are typically limited to centralized water assets with dedicated line power, such as drinking water distribution systems and wastewater treatment facilities.[23] Sensors are usually deployed at a select few locations within the network—like treatment plants, pump stations and boundaries with other systems—and in many cases plant and pump station discharges are not even recorded.[21] For decentralized applications, such as stormwater networks or natural river systems, the cost and power usage of traditional SCADA systems are prohibitive. As such, these distributed resources often go unmonitored and uncontrolled.

Recent studies have also raised concerns about the security of SCADA systems, many of which were designed and installed decades ago.[24,25] Many legacy SCADA systems rely on specialized protocols without built-in support for authentication, such as MODBUS/TCP, EtherNet/IP and DNP317.[24,25] The use of unsecured protocols means that it is possible for unauthorized parties to execute commands remotely on a device in the SCADA network.[24] To cope with this problem, SCADA networks are often isolated from public networks, such as the internet. However, remote attacks are still possible—particularly through the use of unsecured radio channels.[25] Moreover, isolation from public networks limits the use of modern web services such as cloud computing platforms. Reliance on closed networks and proprietary interfaces may also lend a false sense of security to legacy SCADA systems—a concept known as security through obscurity.[24] For these reasons, SCADA systems have gained the reputation of being relatively closed and only manageable by highly-trained operators or specialized local consultants. While SCADA systems remain the most popular platform for managing urban water systems, new tools are needed to improve security, expand coverage, and encourage integration with modern software.

3.2 Wireless sensor networks

The past decade has witnessed a large reduction in the cost and power consumption of wireless electronics; leveraging these advances, wireless sensor networks (WSNs) have opened up new frontiers in environmental monitoring, with applications ranging from biodiversity monitoring,[26] forest fire detection,[27,28] precision agriculture,[29] glacier research,[30] and structural health monitoring.[4] Unlike SCADA systems, WSNs are ideal for low-cost, low-power, and low-maintenance applications, making them well-suited for the monitoring of large water systems like rivers and watersheds. WSNs have been applied to great success in applications ranging from flood monitoring to real-time stormwater control; however, current implementations are generally experimental or proprietary, resulting in a lack of discoverability, limited interoperability, and duplication of effort among projects.

Within the water sciences, flood monitoring represents a particularly important application area for WSNs. While several groups have worked to expand the capabilities of existing legacy flood detection networks,[31–33] only a small number of groups have designed and deployed their own flood monitoring WSNs. Hughes *et al.* (2008) describe a 15-node riverine flood monitoring WSN in the United Kingdom, which interfaces with remote models, performs on-site computation, and sends location-specific flood warnings to stakeholders.[34,35] Other riverine flood monitoring networks include a 3-node river monitoring network in Massachusetts, a 4-node network in Honduras,[36] and—perhaps the largest unified flood monitoring network in the US—the Iowa Flood Information System (IFIS), which draws on a network of over 200 cellular-enabled sensor nodes.[37] While most existing flood-monitoring networks focus on large-scale river basins, flash-flooding has received considerably less attention in the WSN community. Marin-Perez *et al.* (2012) construct a 9-node WSN for flash flood monitoring in a 660 km^2 semiarid watershed in Spain,[38] while See *et al.* (2011) use a Zigbee-based WSN to monitor gully-pot overflows in an urban sewer system.[39] While most deployments are still pilot-scale, these projects demonstrate the potential of WSNs for distributed flood monitoring across a variety of scales and environments.

In addition to monitoring watershed hazards, a limited—but promising—number of projects are illustrating the potential of WSNs for real-time control. Web-enabled sensor nodes have been used to develop adaptive green infrastructure at a select number of pilot sites—for instance, by using weather forecasts to facilitate predictive rainwater harvesting and capture of sediment-bound pollutants.[40] At larger scales, a combined sewer network in South Bend, Indiana uses over 120 flow and depth sensors along with nine valves to actively modulate flows into the city's combined sewer system.[41] This network optimizes the use of existing in-line storage and has achieved a roughly five-fold reduction in combined sewer overflows from 2006–2014 (ref. 41)—all without the construction of additional infrastructure. While distributed control of storm and sewer systems shows promise, most existing implementations are proprietary. A lack of transparency makes these solutions inaccessible to decision makers and the water resources community at large.

Although many research groups have realized the potential for real-time watershed monitoring, existing WSN deployments are generally small-scale and experimental in nature. In order for these networks to be accepted as "critical infrastructure" by the water resources community at large, consistent standards for design, deployment and functionality are needed. In designing their own WSNs, researchers tend to look towards previous research projects.[36] However, research papers rarely include the detailed documentation needed to implement an end-to-end sensor platform.[36] As a result, researchers are often forced to design and deploy their own WSNs from scratch. To prevent duplication of effort and ensure best practices, a community-driven *how-to guide* is urgently needed. Moreover, while proprietary control networks have proven their effectiveness in improving the performance of stormwater systems, an open source alternative is needed to encourage transparency, interoperability, and extensibility. Without open software, standards, and documentation, these new technologies risk becoming like the SCADA systems of old: isolated, proprietary, and incapable of intercommunication.

4 The *open storm* platform

Open storm provides a transparent and unified framework for sensing and control of urban watersheds. To our knowledge, it is the only open-source, end-to-end platform that combines real-time sensing, control and cloud services for the purpose of water resources management. The project is designed to foster engagement by lowering the technological barriers for stakeholders, decision makers, and researchers. To this end, the *open storm* framework is accompanied by a body of reference material that aims to make it easy for non-experts to deploy their own sensors and controllers. This living document, available at open-storm.org, provides tutorials, documentation, supported hardware, and case studies for end-to-end sensor network management. In addition to documenting core features, this guide details the (literal) *nuts-and-bolts* of sensor network deployment, including information that is typically not available in journal articles—such as mounting hardware, assembly instructions and deployment techniques.

The *open storm* framework can broadly be divided into three layers: hardware, cloud services, and applications (Fig. 2). The *hardware* layer includes devices that are deployed in the field—such as sensors for collecting raw data, actuators for controlling water flows, microprocessors, and wireless transmitters. The *cloud services* layer includes processing utilities that receive, store and process data, and interact with field-deployed devices through user-defined applications. Finally, the *application* layer defines how users, algorithms, and real-time models interact with field-deployed devices. This three-tier architecture allows for applications to be developed at a high level, without the need for low-level firmware programming. Together, these layers comprise a scalable framework that can easily be adapted to the needs of a wide variety of users and applications.

4.1 Hardware

4.1.1 The sensor node. At its core, the *open storm* hardware layer (Fig. 1) is enabled by the sensor node—a custom low-power embedded computer with wireless capabilities. The sensor node collects measurements from attached sensors, transmits and receives data from a remote server, and executes control actions. A microcontroller (PSOC5-LP by Cypress Semiconductor) serves at the processing unit for the board. This microcontroller is programmed with a simple operating system that schedules the tasks to be executed, and interfaces with a series of device drivers that control the

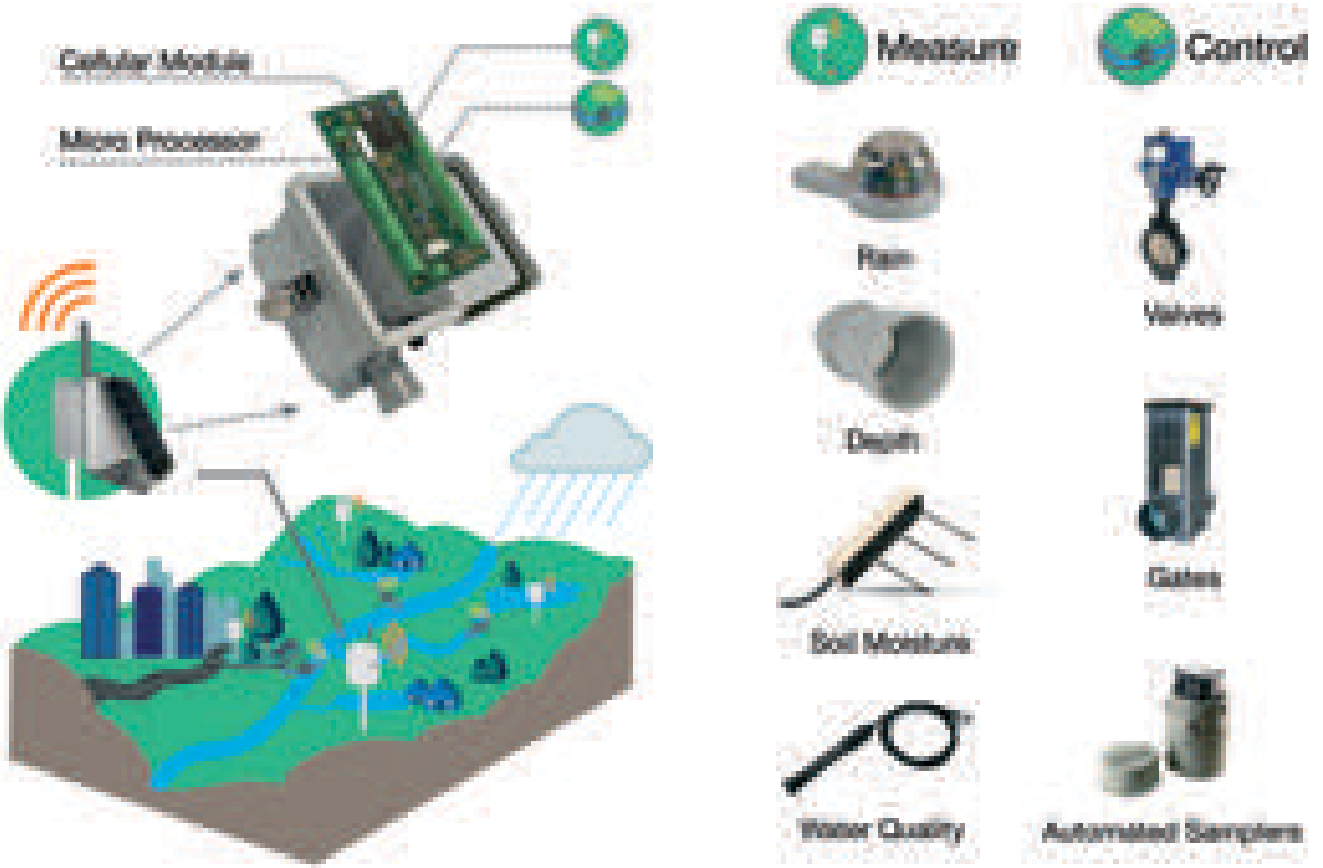


Fig. 1 The *open storm* hardware layer. The left panel shows the complete sensor node along with a representative schematic of its placement in an urban watershed. The right panel shows typical sensors and actuators used in *open storm* research projects.

behavior of attached sensors and actuators. The operating system is designed to minimize power use and consists of a single routine which (i) wakes the device from sleep mode, (ii) downloads pending instructions from the cloud server, (iii) takes sensor readings and triggers actuators, (iv) transmits sensor data to the server, and (v) puts the device back into sleep mode. The sensor node spends the majority of its deployment in sleep mode, allowing it to conserve battery power and remain in the field for an extended period of time.

The sensor node uses wireless telemetry to transmit and receive data from a remote server. While internet connectivity can be achieved through a number of wireless protocols, *open storm* nodes currently use a cellular communications protocol, which enables telemetry through 2G, 3G and 4G LTE cellular networks. Cellular connectivity is implemented through the use of a cellular module (by Telit), along with a small antenna for broadcasting the wireless signal. Compared to other protocols (such as satellite or wi-fi), cellular telemetry is especially suitable for urban and suburban environments due to (i) consistent coverage, (ii) relatively low cost, and (iii) high data throughput. At the time of writing, IoT cellular data plans can be purchased for under $5 per month per node (1–10 MB), making it financially feasible for even small research groups to maintain large-scale networks.

The sensor node is equipped with a power regulation subsystem to provide power to the microcontroller and attached devices. The power supply system consists of four components: (i) a battery, (ii) a solar panel, (iii) a charge controller, and (iv) a voltage converter. The voltage converter permits the sensor node to be powered across a range of 3–40 V. While most sensor nodes are powered by a 3.7 V lithium ion battery, 12 V batteries can also be used for higher-voltage sensors and actuators. The solar panel and solar charger are used to recharge the battery, allowing the device to remain in the field without routine maintenance. At the time of writing, many field-deployed sensor nodes have reported data for over a year without loss of power.

Detailed technical information regarding the sensor node —including parts, schematics and programming instructions —are available online at open-storm.org/node. Excluding the cost of auxiliary sensors, the sensor node can currently be assembled from off-the-shelf parts for a price of approximately $350 per node.

4.1.2 Sensors and actuators. The *open storm* platform supports an extensive catalog of digital and analog environmental sensors. Typical sensors include (i) ultrasonic and pressure-based water level sensors, (ii) soil moisture sensors, (iii) tipping-bucket and optical rain gages, (iv) automated grab samplers for assessing pollutant loads, and (v) *in situ* water quality sensors, including probes for dissolved oxygen, pH, temperature, conductivity, dissolved solids, and oxidation–reduction potential. While many sensors are known to work "out of the box", new sensors can be quickly integrated by adding device drivers to the sensor node firmware. Support for nearly arbitrary sensors is provided by the microcontroller's system-on-chip (SoC), which allows for analog and digital peripherals—like analog-to-digital converters, multiplexers, and logic gates—to be generated using programmable blocks in the device firmware. In addition to environmental sensors, the sensor node also includes internal sensors that report device health statistics, including battery voltage, cellular reception strength, and connection attempts. These device health statistics help to diagnose network issues, and can be used as inputs to remote trigger routines. Sensors can be configured remotely using web services (see *cloud services* section). This capability allows users to turn sensors on or off, or to change the sampling frequency of a sensor without reprogramming the device in the field.

The *open storm* platform also supports an array of actuators that can be used to move mechanical devices in the field. These devices are used to guide the behavior of water systems in real-time, by controlling the flow of water in ponds, channels and pipes. Butterfly valves are one common type of actuating device, and are typically used to control discharge from storage elements such as retention basins. Valves can be opened, closed, or configured across any number of partially opened configurations to modulate flows. As with onboard sensors, these devices are operated remotely using commands sent from a server. Control signals can be specified manually, or through automated control algorithms.

Detailed technical information regarding supported sensors and actuators, along with guides for integrating new devices are provided online at open-storm.org/sensors.

4.2 Cloud services

While sensor nodes can function independently by storing data and making decisions on a local level, integration with cloud services enables system-scale supervision, configuration, and control of field-deployed devices. Like a traditional SCADA system, the cloud services layer facilitates telemetry and storage of sensor data, provides visualization capabilities, and enables remote control of devices—either through manual input or through automated routines. However, unlike a traditional SCADA system, the cloud services layer also allows sensor nodes to communicate with a wide variety of user-defined web applications—including advanced data visualization tools, control algorithms, GIS software, external data ingesters, alert systems, and real-time hydrologic models. By combining real-time supervision and control with domain-specific tools, this architecture enables flexible system-scale control of water assets.

In brief, the cloud services layer performs the following core functions: (1) stores and processes remotely-transmitted data, (2) simplifies management and maintenance of field-deployed sensor nodes, and (3) enables integration with a suite of real-time models, control algorithms, and visualizations. These services are environment-agnostic, meaning that they can be deployed on a local server or a virtual server in the cloud. In practice, however, current *open storm* projects are deployed on popular cloud services—such as Amazon Elastic Compute Cloud (EC2)[42] or Microsoft Azure[43]—to ensure that computational resources flexibly scale with demand. In the following section, we describe the basic architecture, and present example applications that are included with the *open storm* platform.

The cloud services layer follows a simple design pattern, in which applications communicate with sensor nodes through a central database. On the device side, sensor nodes push sensor measurements to the database, and then query the database to determine the latest desired control actions. On the server side, applications query the latest sensor readings from the database, feed these sensor readings into user-defined applications, and then write commands to the database to control the behavior of field hardware remotely. This architecture allows field-deployed sensors to be managed through a single endpoint, and also allows new applications to be developed without modifying critical device firmware.

The database serves dual purposes as both a storage engine for sensor data, and as a communication layer between field-deployed sensors and web applications. The primary purpose of the database is to store incoming measurements from field-deployed sensors. Sensor nodes report measurements directly to the database *via* a secure web connection—using the same protocol that one might use to access web pages in a browser (HTTPS). The database address (URL) is specified in the sensor node firmware, allowing the user to write data to an endpoint of their choosing. In addition to storing sensor measurements, the database also enables bidirectional communication between the node and cloud-based applications by storing device configuration data, command signals, and data from external sources. Server applications communicate with the sensor node by writing commands to the database. These commands are then downloaded by the sensor node on each wakeup cycle. For example, a real-time control application might adjust outflow from a storage basin by writing a sequence of valve positions to the database. At each sampling interval, the sensor node will query the latest desired valve position and enact the appropriate control action. This system enables bidirectional communication with field-deployed sensor nodes without the need for complex middleware.

For its database backend, the *open storm* project uses InfluxDB, a time-series database that is optimized for high availability and throughput of time-series data.[44] Communications with the database backend are secured through the use of basic authentication (*i.e.* a username and password), as well as Transport Layer Security encryption (TLS/SSL). The use of basic authentication prevents unauthorized parties from executing malicious commands on the network, while the use of encryption prevents attackers from intercepting sensitive data. Because applications communicate with the sensor node through the database, this means that applications are secured automatically against attackers as well. Altogether, this system comprises a data backend that is secure, maintainable, and extensible.

4.3 Applications

The *open storm* platform features a powerful application layer that enables users to process and analyze data, build user interfaces, and control sensor nodes remotely. Applications are implemented by creating a series of subscriptions on the central database. These subscriptions perform one of three actions: (i) *read* from the database, (ii) *write* new entries to the database, and (iii) *trigger* actions based on user-specified conditions. While seemingly simple, this system allows for the development of a wide range of applications. A data visualization platform, for instance, is implemented by continuously querying sensor streams from the database; similarly, automated control is implemented by writing a continuous stream of commands. In the following section, we demonstrate the potential of the open storm application platform by presenting example applications, including a data visualization portal, a push alert system, adaptive control, and real-time integration with hydrologic models.

4.3.1 Network supervision and maintenance tools. Much like a traditional SCADA system, the *open storm* platform provides a web-based graphical user interface for real-time visualization and device configuration. Fig. 2 shows an example dashboard, with time series of cellular connection strength (top), radial gauges for monitoring battery voltage (center), and real-time depth readings from two sensor nodes (bottom). Time series visualizations are implemented using the Grafana analytics platform,[45] which allows users to develop customized dashboards that suit their individual needs. To facilitate remote configuration of sensor nodes, *open storm* also includes a web portal that allows users to change device parameters (such as sampling frequency), control actuator behavior, and set event triggers using a web browser.

4.3.2 Automated alerts and adaptive control. In addition to enabling manual supervision and control, *open storm* also provides a rich interface for triggering automatic actions based on user-specified conditions. Push alerts are one common type of trigger event. Alerts can be used to notify stakeholders of hazardous field conditions, such as flooding, or to recommend control strategies to operators in real time. Alerts are also used to notify the user about the health of the network—for instance, by sending push warnings when node battery voltages drop below a threshold, or by emitting a critical alert when data throughput ceases. These system health alerts allow network outages to be promptly diagnosed and serviced. Alerts can be pushed to a variety of endpoints, including email, text messages, or to social media platforms such as Twitter and Slack.[46,47] The wide variety of available push notification formats means that the *open storm* platform is suited to handling both (i) confidential alerts for system operators, and (ii) public emergency broadcasts.

In addition to the alert system, subscriptions are also used to implement adaptive sampling and automatic control. Adaptive sampling allows the sampling frequency of the node to be changed remotely in response to weather forecasts, data anomalies, or manual user input.[48] This in turn allows hydrologically interesting events—such as storm events and dam releases—to be measured at an enhanced resolution. To manipulate sampling frequencies in response to changing weather conditions, for instance, weather forecasts are first downloaded into the *open storm* database using an external data ingester. Next, the subscription service parses the incoming data. If the service detects a probability of rain, the sampling frequency of a

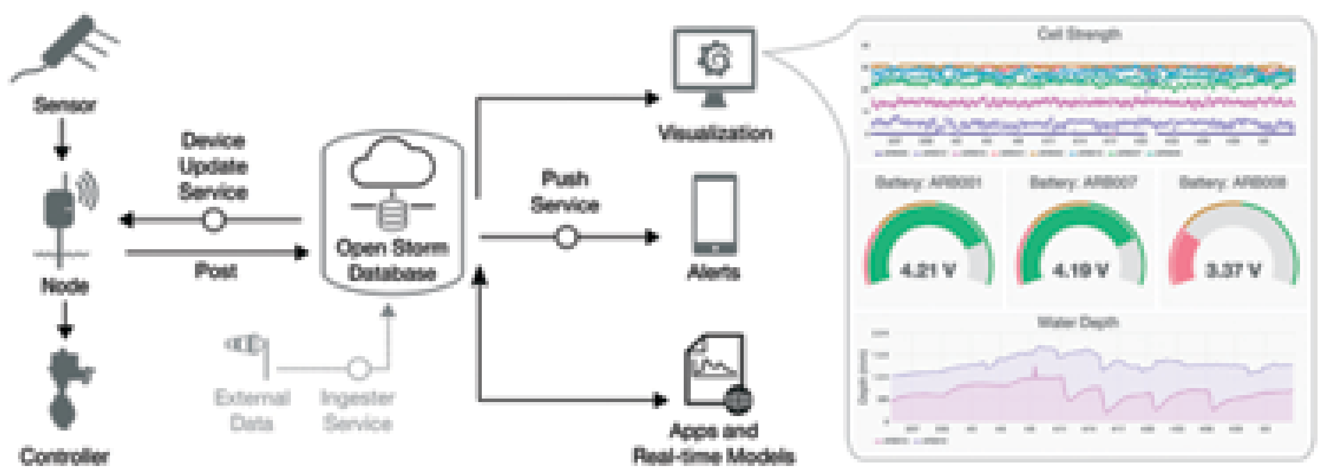


Fig. 2 The *open storm* stack. The hardware layer (left) comprises the sensor node along with auxiliary sensors and actuators. The cloud services layer (center) includes the database backend, along with a series of publication and subscription services for controlling sensor node behavior and interfacing with applications. The applications layer (right) allows for real-time supervision and control of field-deployed devices. The rightmost panel shows an example dashboard, including sensor feeds and network status visualizations.

node is increased. When no precipitation is anticipated, the sampling frequency is decreased, allowing the node to conserve battery power. The same principle is used to implement automated control. The subscription service can be configured as a simple set-point or PID controller, for instance, by computing a control signal based on an input data stream. This controller can in turn be used to optimize outflow from a retention pond, by controlling the position of an outlet valve. More sophisticated control schemes can be implemented by attaching the subscription service to an online model, which optimizes control strategies over an entire stormwater network, achieving system-level benefits. Examples include the MatSWMM and pySWMM software packages,[49,50] which are used to simulate real-time control strategies for urban drainage networks.

Detailed information regarding cloud services and applications can be found at open-storm.org/cloud. In addition to the cloud services platform described here, the *open storm* sensor node is also compatible with other cloud-based data management services, such as the CHORDS (Cloud Hosted Real-time Data Services for the Geosciences) portal.[51]

5 Case studies

To demonstrate the capabilities of the *open storm* platform, we present two ongoing case studies. The first is a real-time flash flood warning network for the Dallas–Fort Worth metroplex in Texas. This deployment detects flash floods at the level of individual roadways, allowing targeted alerts for motorists and improved routing of emergency services during storm events. The second case study is a "smart" stormwater control network in the City of Ann Arbor, Michigan. This deployment aims to improve water quality and mitigate stormwater damage by adaptively timing releases from retention basins across an entire watershed.

5.1 Case study 1: flood monitoring

Located in "flash-flood alley", the Dallas–Fort Worth (DFW) metroplex has historically been one of the most flood-prone areas in the United States.[52] Chronic flooding results in an average of 17 fatalities per year in the state of Texas, with a majority of deaths arising from flash floods.[52] Despite recent efforts to improve stormwater management,[53] lack of fine-scale runoff measurements inhibits prediction and communication of flash flood risks. To address this problem, we are using the *open storm* platform to build a real-time flash flood monitoring network. Drawing on the *open storm* real-time alert system, this network aims to improve disaster response by communicating flood risks to emergency managers in real-time, and by generating targeted alerts that will allow motorists to safely navigate around inundated roadways.

To date, urban flash flooding remains a poorly-understood phenomenon. There is currently no model that is capable of generating reliable flash flood estimates in urban areas.[54] Modeling of urban flash floods is complicated by an absence of natural flow paths and interaction of runoff with man-made structures.[54] However, lack of data at appropriate spatial and temporal scales also presents a major challenge. For reliable modeling of flash floods, Berne (2004) recommends using rainfall data at a minimum spatial resolution of 500 meters,[55] while a recommended temporal resolution of 1–15 minutes for rainfall is recommended by Smith (2007).[56] Existing rain gages and river stage monitors are often too sparsely distributed to meet these requirements. Within the DFW metroplex, NWS maintains 12 quality-controlled gages,[57] while USGS provides precipitation data at 30 sites.[58] This means that the current spatial resolution of validated rain gages within the DFW metroplex is roughly 1 gage per 600 km^2—too sparse for reliable prediction of flash floods. Likewise, current river stage monitors for the DFW region are largely deployed along mainstems of creeks and rivers with contributing areas ranging from 20 km^2 to 21 000 km^2 (and a median contributing area of 220 km^2). While these gages provide excellent coverage of riverine flooding, they offer limited potential for capturing flash floods.

To fill coverage gaps and enable real-time flash flood forecasting, we are building a wide-area flood monitoring network that is specifically tailored to monitoring flash floods over small-scale catchments (ranging from about 3 to 80 km^2 in size). Our approach is to leverage a large array of inexpensive depth sensors to capture runoff response at the scale of individual roadways, creeks, and culverts. By using inexpensive hardware, we are able to scale our network to a size that would be infeasible with state-of-the-art stage monitoring stations (such as those used by NOAA or USGS). At the time of writing, 40 sensor nodes have been allocated and built for the DFW flood monitoring project, with over 15 nodes currently deployed and reporting. These 40 sensor nodes have been built for a cost of $20 000 USD—less than the cost as a single USGS gaging station.‡[59]

Fig. 3 presents an overview of the DFW flood monitoring network. The left panel shows a map of the DFW metroplex, with current and proposed sensor node locations. The bottom-right panel shows a detail of a typical sensor node installation. Like most nodes in the network, this node is mounted to a bridge deck with an ultrasonic depth sensor pointed at the stream surface below. The sensor node records the depth to the water surface at a typical time interval of 3–15 minutes. The top-right plot shows a time series of stream depth during two distinct storm events for a sample of nodes on the network. From this plot, it can be seen that the runoff response varies widely between sensor locations, even in a relatively concentrated geographic area. During the second event, for instance, node 2 (yellow) reports a large increase in discharge, while node 9 (purple) reports no change in

‡ The installation cost for a USGS stage-discharge streamgaging station is roughly $20 000, with an annual recurring cost of approximately $16 000.

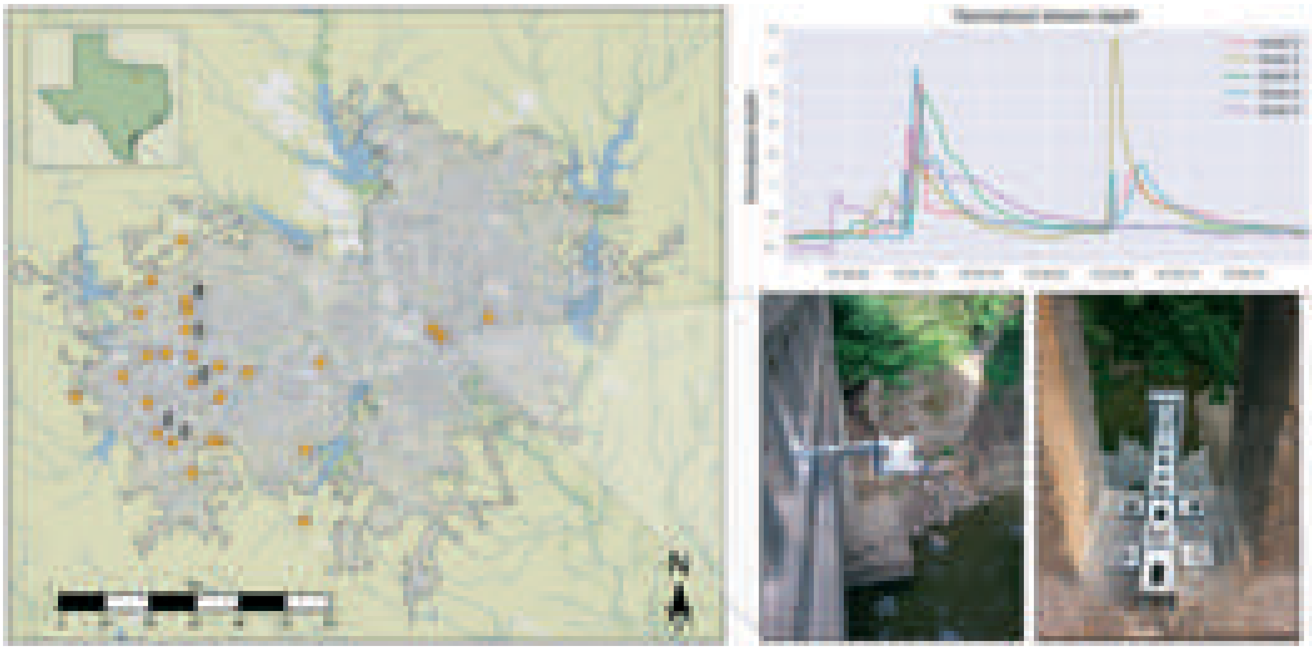


Fig. 3 Flood monitoring network in the Dallas–Fort Worth metroplex. The map (left) shows current and proposed sensor sites, while the detail photos (bottom-right) show an example bridge-mounted depth sensor node. Time series (top-right) show the response in stream depth to a series of storm events from August 5–6, 2016. From these stage hydrographs, it can be seen that the response varies widely even within a relatively small geographic area.

discharge. Comparison of the hydrographs with NEXRAD[60] radar data shows that the variability in stage is largely explained by spatial variability in the rainfall fields.§ This result confirms the need for increased spatial resolution in stream stage measurements for flash flood monitoring.

The *open storm* platform enables detection and communication of flood risks on spatial and temporal scales appropriate for real-time disaster response and control. Adaptive management of traffic during extreme weather events represents one important application of this technology. The Dallas–Fort Worth flood monitoring network could improve disaster response by communicating flood risks to motorists in real-time, thereby allowing them to safely navigate around flooded roadways. This is especially important given that in the US, roughly 74% of fatalities from flooding are motor-vehicle related,[8] and in Texas, as much as 93% of flood-related deaths result from walking or driving into floodwaters.[52] Current alert systems are to a large extent insensitive to spatial variability in flood response.[35] However, the *open storm* framework enables targeted alerts that can be integrated into existing mobile navigation apps. In a future that may be characterized by autonomous vehicles and vehicle-to-infrastructure communication,[61] this technology could one day be used to adaptively route traffic during extreme weather events.

5.2 Case study 2: controlling watersheds

As illustrated by the Dallas–Fort Worth flood-monitoring network, real-time measurements can play a pivotal role in providing alerts to stakeholders and improving our understanding of watershed dynamics. However, with the addition of active control, it is possible to not only monitor adverse events, but also to prevent them. The *open storm* platform is capable of enacting control on a watershed scale using distributed valve controllers, adaptive control schemes, and cloud-hosted hydrologic models. Instead of building bigger stormwater systems, operators may use real-time control to make better use of existing water infrastructure, mitigate flooding, and decrease contaminant loads into sensitive ecosystems.

The *open storm* framework is presently being used to control an urban watershed in the City of Ann Arbor, Michigan. The *Malletts Creek* watershed—a 26.7 km^2 tributary of the Huron River—has traditionally served as a major focal point in the city's strategy to combat flooding and reduce runoff-driven water quality impairments.[62] Given its proximity to the Great Lakes, water resource managers have placed an emphasis on reducing nutrient loads from urban runoff. A majority of the discharge in Malletts creek originates from the predominantly impervious upstream (southwestern) reach of the watershed, while a significant, but smaller portion of the discharge originates from the central reach of the watershed. For this reason, local water resource managers have constructed a number of flood-control basins in the upstream segments of the catchment. It is these basins that are now modified to allow for real-time control of the watershed.

The watershed is modified for control at two locations by retrofitting existing basin outlets with remotely-operated valves (Fig. 4). The first control point is a stormwater retention pond in the southern part of the watershed (shown in

§ See https://github.com/open-storm/docs.open-storm.org/wiki/Case-study:-Flood-Monitoring-in-Dallas-Fort-Worth.

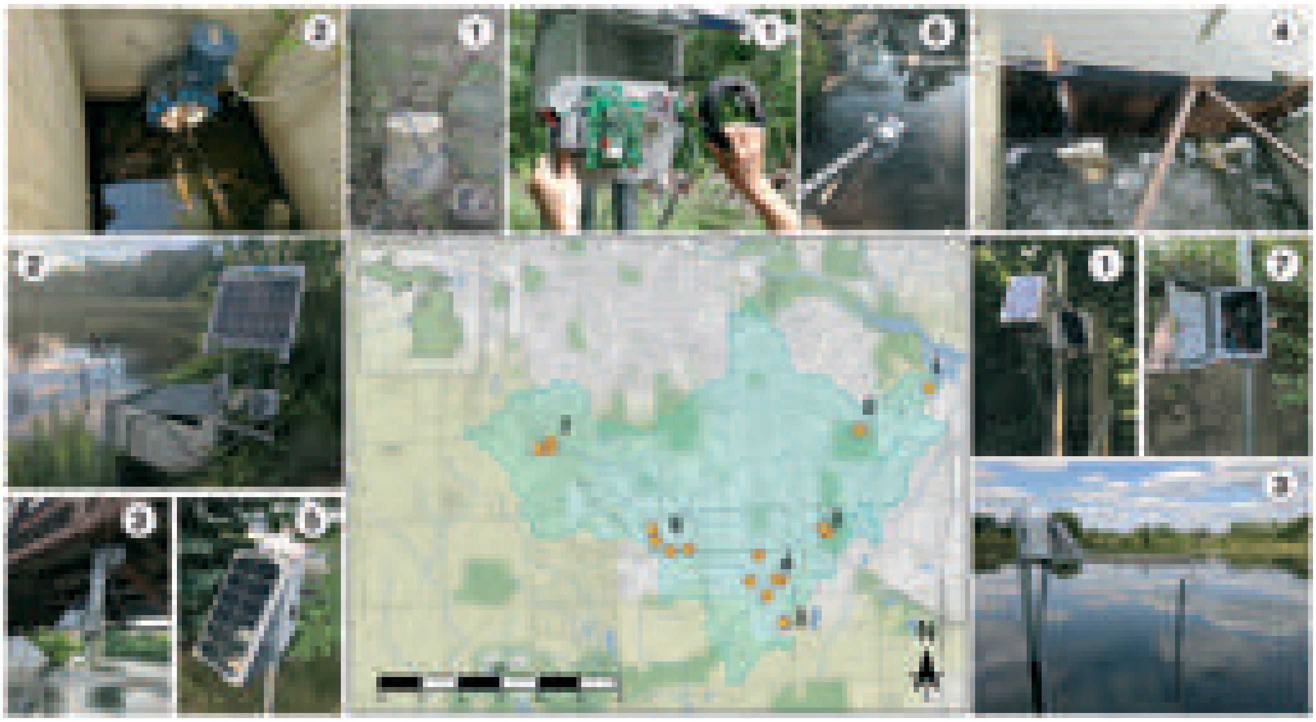


Fig. 4 The Ann Arbor stormwater control network with selected sites. The blue region of the map shows the boundary of the Malletts Creek watershed, while node locations are indicated by gold markers. The outlet of the watershed (1) is monitored by a USGS gage, along with an automated sampler (1, top-left) and a depth-sensing node (1, center-right). Active control is added through a butterfly valve (2), and dual gate valves (4). The watershed is monitored by an array of depth sensors (6, 7 and 3, bottom-left), soil moisture sensors (5) and *in situ* water quality sensors (3, bottom-right).

red in Fig. 5). While originally designed as a flow-through (detention) pond, the addition of two 30 cm diameter gate valves allows for an additional 19 million liters of water to be actively retained or released. The second control point is a smaller retention pond, located in the central reach of the watershed (shown in green in Fig. 5). This control site is retrofitted with a rugged 30 cm diameter butterfly valve. The position of each valve is controlled *via* an attached sensor node, which relays commands from a remote server. Each sensor node is equipped with a pair of ultrasonic sensors: one to measure the water depth at the pond, and one to measure the depth of the outflow stream. The control sites operate entirely on 12 V battery power, along with a solar panel to recharge the battery during daylight hours. This configuration allows the controller to remain in the field permanently, without the need for a dedicated external electricity source.¶

In addition to the two control sites, the Ann Arbor network is also instrumented with more than twenty sensor nodes that monitor system performance and characterize real-time site conditions. Using a combination of ultrasonic depth sensors, optical rain gages, and soil conductivity sensors, these nodes report stream stage, soil moisture, soil temperature, and precipitation accumulation approximately once every 10 minutes (with an increased resolution of 2–3 minutes during storms). An additional set of nodes is deployed to measure water quality—including dissolved oxygen, pH, temperature, oxidation reduction potential, conductivity, temperature—as well as an automated grab sampler for capturing contaminants of interest (such as heavy metals and microbes). These nodes are deployed at the inlet and outlet of constructed wetlands to determine how real-time control affects the removal of pollutants.

Measurements from the sensor network are validated using an external United States Geological Survey flow measurement station (USGS station 4174518), located at the watershed outlet. These federally-certified measurements are available freely on the web, making them relatively easy to ingest into the *open storm* framework as an external data source. Furthermore, localized weather forecasts are ingested from public forecasting services (darksky.net) to provide daily, hourly, and minute-level forecasts to inform the control of each site in the network.[63] These external data sources allow for near-instant validation of sensor data, and provide a holistic "snapshot" of system states.

We confirm the effectiveness of the control network through a simple experiment. In this experiment, stormwater is retained at an upstream control site, then released gradually to maximize sedimentation and reduce erosion downstream. While it is known that the addition of control valves affords many localized benefits—such as the ability to

¶ With two people, installation at each site takes approximately one day. This includes time dedicated to mounting valves, sensors, and remotely-testing the equipment.

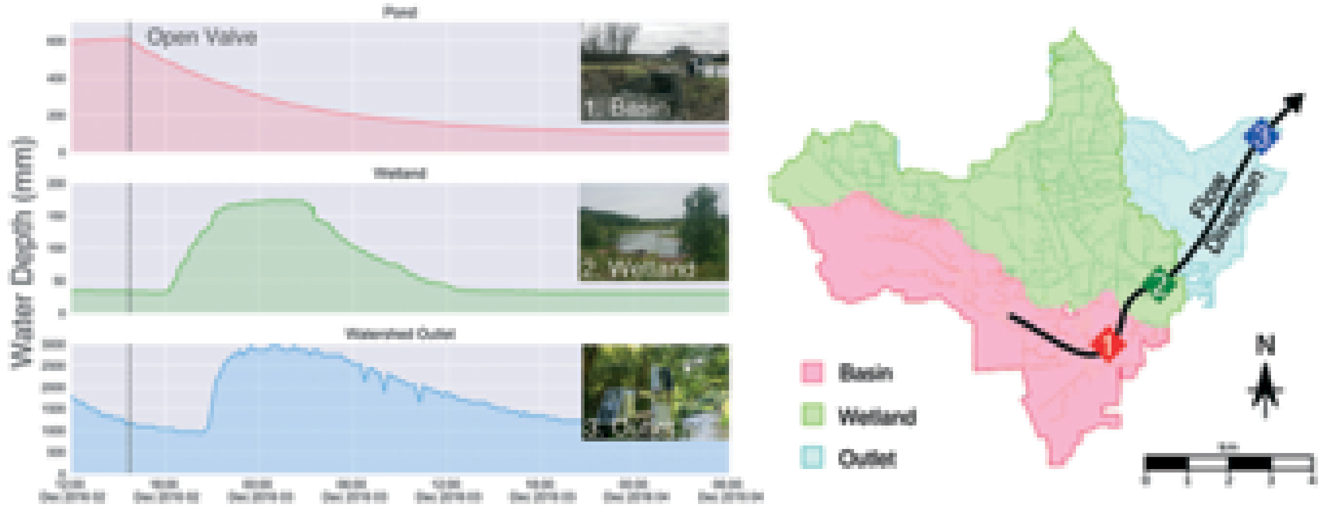


Fig. 5 Malletts Creek control experiment in Ann Arbor. The left panel shows time series of water depth from 12:00 pm on December 2 to 6:00 am on December 4, 2016. The right panel shows the location of the three sites in the watershed, with the partitioned contributing areas of each location corresponding to the colors of the time series plots.

increase retention and capture sediments[64]—the goal of this experiment is to test the extent to which control of individual sites can improve watershed-scale outcomes. The control experiment takes place on a river reach that stretches across three sites: a retention pond (upstream), a constructed wetland (center), and the watershed outlet. Fig. 5 (right) shows the three test sites within the watershed, with the fractional contributing area of each site indicated by color. In this system, runoff flows from the retention pond (red) to the watershed outlet (blue) by way of an end-of-line constructed wetland (green) designed to treat water, capture sediments, and limit downstream erosion. Erosion, in particular, has been shown to be primary source of phosphorus in the watershed,[48] thus emphasizing the need to reduce flashy flows. While the wetland serves a valuable purpose in improving water quality, it is sized for relatively small events. Specifically, the basin is designed to hold up to 57 million liters of stormwater but experiences as much as 760 million liters during a ten-year storm. Thus, it often overflows during storms, meaning that treatment benefits are bypassed. To maximize treatment capacity, a sensor node is placed into the wetland to measure the local water level and determine the optimal time to release from the retention pond upstream.

At the outset of the experiment, water is held in the upstream retention pond following a storm on December 1, 2016. Residual discharge from the original storm event can be observed as a falling hydrograph limb at the USGS gaging station (blue) during the first 10 hours of the experiment (Fig. 5). The sensor located at the wetland is used to determine the time at which it is safe to release upstream flows without overflowing the wetland (Fig. 5). Water is initially released from the pond at 4:00 pm on December 2, as indicated by a drop in the water level of the pond. Two hours later, the water level in the wetland begins to rise due to the discharge arriving from upstream. Finally, after another three hours, the discharge wave reaches the outlet, where it is detected by the USGS flow station. Over the course of the controlled release, the station registers roughly 19 million liters of cumulative discharge.

The control experiment shows demonstrable improvements in system performance compared to the uncontrolled case. While the water quality benefits will be measured in the coming year, a number of likely benefits can be posited. As measured, over 19 million liters were removed from the storm window and retained in the basin following the storm event. The residence time of the water in the pond increased by nearly 48 hours, increasing the potential for sedimentation.[64] The removal of stormwater flows also resulted in attenuation of the downstream hydrograph. The peak flows at the watershed outlet were measured to be 0.28 $m^3 s^{-1}$ during the storm, but would have been nearly 0.60 $m^3 s^{-1}$ had the valves in the basin not been closed. Based on prior studies in the watershed—which showed that flows in the stream correlate closely with suspended sediment concentrations—it can be estimated that the flows from the basin were discharged at roughly 60 mg L^{-1}, rather than 110 mg L^{-1}, thus nearly halving the concentration of suspended solids and total phosphorus in the flows originating from the controlled basin.[48] Moreover, the controlled experiment enhanced the effective treatment capacity at the wetland downstream, which would have overflowed during the storm, thus not treating the flows from the upstream pond. As such, the simple addition of one upstream valve provided additive benefits across a long chain of water assets, demonstrating firsthand how system-level benefits can be achieved beyond the scale of individual sites. While the water quality impacts of active control deserve further assessment, this study opens the door for adaptive stormwater control at the watershed scale. Rather than optimizing the performance of isolated sites, the *open* *storm* platform can be used to determine the optimal control strategy for an entire watershed, then enact it in real-time.

6 Conclusion

Open storm is an all-in-one, "batteries included" platform for monitoring and managing urban water systems. Its emphasis on extensive configurability, real-time response, and automated control make it an ideal choice for water system managers and environmental researchers alike. While many open hardware platforms exist, *open storm* is the first open-source, end-to-end platform that combines sensing, control and cloud computing in service of water resources management. Aside from providing a technological blueprint, *open storm* addresses the real-world requirements that can be expected in water resources applications, such as field-robustness, low-power operation and system-scale coordination. The *open storm* project has shown proven results in extending the capabilities of existing stormwater systems: both by increasing the spatiotemporal resolution of measurements, and by actively improving water quality through real-time control. However, *open storm* is not just a platform—it's also a community of researchers, stakeholders and decision-makers who are dedicated to realizing smarter water systems. To assist in the dissemination and development of smart water systems, we are creating a living document at open-storm.org in order to share standards, reference materials, architectures, use cases, evaluation metrics, and other helpful resources. We invite users to participate in this project by sharing their experiences with designing, deploying and maintaining smart water systems.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We would like to acknowledge the Cities of Ann Arbor, Dallas, Fort Worth and Arlington for their support in deploying and maintaining the wireless sensor networks described in this study. We would also like to thank the Huron River Watershed Council for their assistance in operating and maintaining the Ann Arbor stormwater control network.

References

1 A. Caragliu, C. D. Bo and P. Nijkamp, *J. Urban Technol.*, 2011, **18**, 65–82.
2 G. Dimitrakopoulos and P. Demestichas, *IEEE Vehicular Technology Magazine*, 2010, vol. 5, pp. 77–84.
3 A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, *IEEE Internet Things J.*, 2014, **1**, 22–32.
4 J. P. Lynch, *Struct. Control Health Monit.*, 2005, **12**, 405–423.
5 L. Mays, *Water Resources Engineering*, John Wiley & Sons, 2010.
6 A. Bronstert, D. Niehoff and G. Bürger, *Hydrol. Process.*, 2002, **16**, 509–529.
7 T. Stocker, *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, 2014.
8 S. Doocy, A. Daniels, S. Murray and T. D. Kirsch, *PLoS Currents*, 2013.
9 J. H. Ahn, S. B. Grant, C. Q. Surbeck, P. M. DiGiacomo, N. P. Nezlin and S. Jiang, *Environ. Sci. Technol.*, 2005, **39**, 5940–5953.
10 R. O. Carey, W. M. Wollheim, G. K. Mulukutla and M. M. Mineau, *Environ. Sci. Technol.*, 2014, **48**, 7756–7765.
11 C. J. Vörösmarty, P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann and P. M. Davies, *Nature*, 2010, **467**, 555–561.
12 *Clean Watersheds Needs Survey 2012*, Environmental protection agency technical report, 2016.
13 B. Sercu, L. C. V. D. Werfhorst, J. L. S. Murray and P. A. Holden, *Environ. Sci. Technol.*, 2011, **45**, 7151–7157.
14 *2013 Report Card for America's Infrastructure*, American Society of Civil Engineers, 2013.
15 C. H. Emerson, C. Welty and R. G. Traver, *J. Hydrol. Eng.*, 2005, **10**, 237–242.
16 R. J. Hawley and G. J. Vietz, *Freshw. Sci.*, 2016, **35**, 278–292.
17 *The Importance of Operation and Maintenance for the Long-Term Success of Green Infrastructure: A Review of Green Infrastructure O&M Practices in ARRA Clean Water State Revolving Fund Projects*, Environmental protection agency technical report, 2013.
18 B. P. Wong and B. Kerkez, *Environ. Model Softw.*, 2016, **84**, 505–517.
19 B. Kerkez, C. Gruden, M. Lewis, L. Montestruque, M. Quigley, B. Wong, A. Bedig, R. Kertesz, T. Braun, O. Cadwalader, A. Poresky and C. Pak, *Environ. Sci. Technol.*, 2016, **50**, 7267–7273.
20 A. Mullapudi, B. P. Wong and B. Kerkez, *Environ. Sci.: Water Res. Technol.*, 2017, **3**, 66–77.
21 L. W. Mays, *et al.*, *Water distribution systems handbook*, McGraw-Hill New York, 2000, vol. 17.
22 R. Powell and K. Hindi, *Computing and Control for the Water Industry*, Research Studies Press, 1999.
23 A. Association, *Instrumentation and Control*, American Water Works Association, 3rd edn (M2), 2001.
24 V. M. Igure, S. A. Laughter and R. D. Williams, *Computers & Security*, 2006, **25**, 498–506.
25 L. Mays, *Water supply systems security*, McGraw-Hill, 2004.
26 A. Cerpa, J. Elson, M. Hamilton, J. Zhao, D. Estrin and L. Girod, *Workshop on Data communication in Latin America and the Caribbean - SIGCOMM LA 01*, 2001.
27 M. Hefeeda and M. Bagheri, *2007 IEEE Internatonal Conference on Mobile Adhoc and Sensor Systems*, 2007.
28 H. Soliman, K. Sudan and A. Mishra, *2010 IEEE Sensors*, 2010.

29 Y. Kim, R. G. Evans and W. M. Iversen, *IEEE Trans. Instrum. Meas.*, 2008, **57**, 1379–1387.
30 K. Martinez, J. K. Hart and R. Ong, *Computer*, 2004, **37**, 50–56.
31 P. Bonnet, J. Gehrke and P. Seshadri, *IEEE Personal Communications*, 2000, vol. 7, pp. 10–15.
32 T. Imielinski and S. Goel, *IEEE Personal Communications*, 2000, vol. 7, pp. 4–9.
33 N. Chen, K. Wang, C. Xiao and J. Gong, *Environ. Model Softw.*, 2014, **54**, 222–237.
34 D. Hughes, P. Greenwood, G. Blair, G. Coulson, P. Grace, F. Pappenberger, P. Smith and K. Beven, *Concurr. Comput.*, 2008, **20**, 1303–1316.
35 P. J. Smith, D. Hughes, K. J. Beven, P. Cross, W. Tych, G. Coulson and G. Blair, *Meteorol. Appl.*, 2009, **16**, 57–64.
36 E. A. Basha, S. Ravela and D. Rus, *Proceedings of the 6th ACM conference on Embedded network sensor systems - SenSys 08*, 2008.
37 I. Demir and W. F. Krajewski, *Environ. Model Softw.*, 2013, **50**, 77–84.
38 R. Marin-Perez, J. García-Pintado and A. S. Gómez, *Sensors*, 2012, **12**, 4213–4236.
39 C. See, K. Horoshenkov, R. Abd alhmeed, Y. Hu and S. Tait, *IEEE Sens. J.*, 2011, 1545–1553.
40 M. Quigley, *EWRI Currents*, 2015, vol. 17, year.
41 L. Montestruque and M. D. Lemmon, *Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks - CySWater 15*, 2015.
42 *Elastic Compute Cloud (EC2): Cloud Server & Hosting, AWS*, 2017, https://aws.amazon.com/ec2/.
43 *Microsoft Azure: Cloud Computing Platform & Services*, 2017, https://azure.microsoft.com.
44 *InfluxDB: Scalable datastore for metrics, events, and real-time analytics*, 2017, https://www.influxdata.com/.
45 *Grafana – the open platform for analytics and monitoring*, 2017, https://grafana.com.
46 *Twitter. It's what's happening*, 2017, https://twitter.com.
47 *Slack: Where work happens*, 2017, https://slack.com.
48 B. P. Wong and B. Kerkez, *Water Resour. Res.*, 2016, **52**, 8986–9000.
49 G. Riaño-Briceño, J. Barreiro-Gomez, A. Ramirez-Jaime, N. Quijano and C. Ocampo-Martinez, *Environ. Model Softw.*, 2016, **83**, 143–154.
50 *pySWMM: Python Wrappers for SWMM*, 2017, https://github.com/OpenWaterAnalytics/pyswmm.
51 B. Kerkez, M. Daniels, S. Graves, V. Chandrasekar, K. Keiser, C. Martin, M. Dye, M. Maskey and F. Vernon, *Geosci. Data J.*, 2016, **3**, 4–8.
52 H. O. Sharif, T. L. Jackson, M. M. Hossain and D. Zane, *Nat. Hazards Rev.*, 2014, **16**, 04014016.
53 L. Lee, *txH2O*, 2016.
54 H. A. P. Hapuarachchi, Q. J. Wang and T. C. Pagano, *Hydrol. Process.*, 2011, **25**, 2771–2784.
55 A. Berne, G. Delrieu, J.-D. Creutin and C. Obled, *J. Hydrol.*, 2004, **299**, 166–179.
56 J. A. Smith, M. L. Baeck, K. L. Meierdiercks, A. J. Miller and W. F. Krajewski, *Adv. Water Resour.*, 2007, **30**, 2087–2097.
57 *F6 Climate Data*, 2017, https://www.weather.gov/fwd/f6.
58 *Texas Water Dashboard*, 2017, https://txpub.usgs.gov/txwaterdashboard.
59 H. Sheehan, Personal communication.
60 T. D. Crum and R. L. Alberty, *Bull. Am. Meteorol. Soc.*, 1993, **74**, 1669–1687.
61 D. Jiang and L. Delgrossi, *VTC Spring 2008 - IEEE Vehicular Technology Conference*, 2008, pp. 2036–2040.
62 R. Lawson, E. Riggs, D. Weiker and J. Doubek, *Total Suspended Solids Reduction and Implementation Plan for Malletts Creek: October 2011 - September 2016*, Huron river watershed council technical report, 2016.
63 *Dark Sky API*, 2017, https://darksky.net/dev.
64 E. Gaborit, D. Muschalla, B. Vallet, P. A. Vanrolleghem and F. Anctil, *Urban Water J.*, 2013, **10**, 230–246.

sensors

MDPI

Article

Shaping Streamflow Using a Real-Time Stormwater Control Network

Abhiram Mullapudi [†], Matthew Bartos [†], Brandon Wong [†] and Branko Kerkez *,[†]

Department of Civil & Environmental Engineering, University of Michigan; Ann Arbor, MI 48109, USA; abhiramm@umich.edu (A.M.); mdbartos@umich.edu (M.B.); bpwong@umich.edu (B.W.)

* Correspondence: bkerkez@umich.edu; Tel.: +1-734-647-0727

† These authors contributed equally to this work.

Received: 15 June 2018; Accepted: 11 July 2018; Published: 13 July 2018

check for updates

Abstract: "Smart" water systems are transforming the field of stormwater management by enabling real-time monitoring and control of previously static infrastructure. While the localized benefits of active control are well-established, the potential for system-scale control of watersheds is poorly understood. This study shows how a real-world smart stormwater system can be leveraged to shape streamflow within an urban watershed. Specifically, we coordinate releases from two internet-controlled stormwater basins to achieve desired control objectives downstream—such as maintaining the flow at a set-point, and generating interleaved waves. In the first part of the study, we describe the construction of the control network using a low-cost, open-source hardware stack and a cloud-based controller scheduling application. Next, we characterize the system's control capabilities by determining the travel times, decay times, and magnitudes of various waves released from the upstream retention basins. With this characterization in hand, we use the system to generate two desired responses at a critical downstream junction. First, we generate a set-point hydrograph, in which flow is maintained at an approximately constant rate. Next, we generate a series of overlapping and interleaved waves using timed releases from both retention basins. We discuss how these control strategies can be used to stabilize flows, thereby mitigating streambed erosion and reducing contaminant loads into downstream waterbodies.

Keywords: smart cities; smart water systems; wireless sensor networks; stormwater; real time control

1. Introduction

Burdened by aging infrastructure, growing populations and changing hydrologic conditions, many municipalities struggle to adequately manage stormwater [1]. Flash flooding can occur when stormwater infrastructure is unable to convey runoff away from developed areas [2]. At the same time, pollutants from urban runoff—such as nutrients, heavy metals and microbes—can contaminate downstream waterbodies, damaging aquatic habitats and resulting in toxic algal blooms [1]. Traditionally, civil engineers have addressed these challenges by building larger storage and conveyance infrastructure (e.g., basins and pipes). However, this approach suffers from a number of important disadvantages. First, new construction is expensive, and is often unfeasible for chronically underfunded stormwater departments [3]. Second, static designs are inflexible to future changes in weather, population growth, and regulatory requirements [2]. Third, overdesigned conveyance systems can cause flooding, erosion and damage to downstream property and ecosystems, which ultimately necessitates further remediation and construction [1]. In the face of increasing urbanization and more frequent extreme weather events [4,5], new strategies are needed to ensure effective management of stormwater.

In contrast to traditional *steel-and-concrete* solutions, real-time control has emerged as a novel means to improve the performance of stormwater systems at minimal expense. Drawing on wireless communications, low-power microcontrollers, and modern advances in control theory, these systems achieve performance benefits by reconfiguring water infrastructure in real time [1,6]. Real-time control of stormwater basins, for instance, can improve water quality following a storm event by enhancing removal of contaminants [1]. Similarly, active regulation of discharges through constructed wetlands can improve water quality and rehabilitate aquatic habitats [6,7]. More broadly, by controlling flows over a large network, operators can harness the latent treatment capacity of many distributed stormwater assets, effectively turning urban watersheds into distributed wastewater treatment plants [1,6].

A small number of studies have evaluated the benefits of real-time stormwater control. Most of these studies describe retrofits of isolated sites for rainwater capture and on-site pollutant treatment. Middleton and Barrett (2008) show that equipping existing retention basins with real-time controllers can reduce stormwater pollutant loads downstream by increasing the retention time of captured stormwater [8]. Roman et al. (2017) describe an adaptively-controlled rainwater harvesting system in New York City that captures 35–60% more rainwater than conventional systems [9]. Similarly, Klenzendorf et al. (2015) describe a rainwater harvesting pilot project and a retention basin retrofitted for real-time control in Austin, Texas [10]. The authors show that the controlled retention basin reduces deposition of nitrogen and total suspended solids (TSS) into the downstream system. These studies demonstrate that active control can significantly improve the performance of existing sites at a lower cost than new construction. However, benefits are only examined at a local scale. This distinction is important, given that localized practices do not necessarily achieve the best system-scale outcomes. Indeed, some research indicates that when local best management practices are implemented without accounting for global outcomes, they can produce adverse flow conditions at the watershed scale [11].

Currently, the benefits of coordinated stormwater control are poorly understood. Inspiration for the benefits of system-level control can be taken from sewer operations. While most sewer systems still only rely on local control logic, such as water level setpoints [12], recent work has demonstrated how wider benefits can be achieved through the cooperative action of multiple controllers working in tandem. The cities of Copenhagen and Barcelona, for instance, implement a combination of local rule-based control, and some higher-level optimization that jointly coordinates actions between groups of actuators [13]. Montestruque and Lemmon (2015) describe CSOnet, a sewer control network consisting of 120 sensors and 12 actuators in the city of South Bend, Indiana [3]. This network uses dynamic control algorithms to adaptively balance hydraulic loads throughout the sewer's interceptor lines, ultimately reducing combined sewer overflows (CSOs) by as much as 25%. While these systems achieve impressive system-scale control of a large sewer networks, it is still unclear how lessons learned from these proprietary sewer control approaches may translate to the broader control of urban watersheds and separated stormwater systems.

In this study, we describe an approach for managing stormwater discharges across an urban watershed using internet-connected valves and sensors. We show that by actively coordinating releases from two parallel retention basins, we can produce desirable flow regimes at a target location downstream, which would not be possible with passive infrastructure alone. This study takes place in four phases. In the first phase, we describe the development of a real-time stormwater control system in the city of Ann Arbor, Michigan. Building on an existing wireless sensing and control network described in Bartos et al. (2018) [6], we demonstrate how static retention basins can be retrofitted with internet-controlled valves, and present a new method for controlling these basins using a controller scheduling application. In the second phase, we characterize the ability of the control network to shape the downstream hydrograph by releasing impulses of different sizes from two retention basins and determining the magnitude, travel time, and decay envelope of the resulting waves. In the third phase, we use the data gathered from this exploratory analysis to determine the control input needed to produce a flat hydrograph at the outlet of the watershed. We discuss how this control strategy can be used to prevent erosion and reduce phosphorus loads into downstream waterbodies. Finally, in the fourth phase, we show how control inputs can be timed to produce synchronized and de-synchronized pulses at a downstream target location. In addition to demonstrating the precision of the control system, this experiment shows how interleaving pulses can be used to free up capacity in upstream retention basins without inducing synchronized flashy flows downstream. We discuss how these simple control "building blocks" can be used by system operators to achieve more sophisticated stormwater management targets. Unlike most existing systems, our control network uses an open-source hardware and software stack, making it freely available to municipalities that are interested in implementing their own smart stormwater control systems. Thus, when combined with supplementary *how-to* documentation on `open-storm.org`, this study provides the foundation for an "operator's manual" for real-time control of urban watersheds.

2. Study Area and Technologies

2.1. Study Area

This study focuses on a wireless control network in the Mallets creek watershed—an urbanized creekshed located in the city of Ann Arbor, Michigan. This creekshed has been the focus of ongoing efforts to reduce peak flows and improve water quality [14]. The creekshed has an area of about 26.7 km^2 and contains streams that altogether exceed 16 km in length. These streams drain into the Huron River and ultimately the Great Lakes. With high areas of development and over 33% imperviousness, little natural land is available for infiltration and uptake, resulting in flashy flows that erode stream banks and result in unstable habitats. These rapid flows drive stream erosion and increased transport of sediments and nutrients out of the watershed [14]. While there are no lakes in the creekshed, there are several natural and manmade stormwater basins that have been constructed to help stabilize flows throughout the creekshed and mitigate the impacts of non-point source runoff.

To investigate the effects of real-time control on the creekshed, we deploy a control network that measures and regulates flows from two large stormwater basins. The control network consists of four sites centered around the main stem of the creek. Figure 1 shows the locations of each of these four sites in the control network. Water first flows into a large retention basin with a storage capacity of 19 ML (site A), located at the most upstream point in the control network. From this retention basin, water travels 1.4 km downstream to a constructed wetland (site C), designed to slow the flow of water and remove contaminants. After passing through the wetland, water travels another 3 km until it is joined by flows arriving from a smaller retention basin with a storage capacity of 7.5 ML (site B). The combined flows exit the creek at the outlet of Mallet's creek (site D), after which they enter the Huron River. Internet-controlled valves are deployed at the two stormwater basins at sites A and B. These valves are used in subsequent experiments to regulate flows at the outlet of the creek.

2.2. Technologies and Architecture

Flows throughout the creekshed are measured and controlled using a custom wireless sensing and control network. This network is built using the `open storm` hardware and software stack, which has been described and documented in Bartos et al. (2018) [6]. The hardware layer uses an ultra-low power ARM Cortex-M3 microcontroller (Cypress PSoC), which implements the sensing and control logic in its firmware. Internet connectivity is achieved using a CDMA cellular modem (Telit DE910), which facilitates wireless bi-directional communication between the field device and a remote server. The full unit is powered using a solar-rechargeable 3.7 V lithium-ion battery. To measure the hydrologic response of the system, wireless sensor nodes are deployed along the main stem of the creek. Each sensor node is equipped with an ultrasonic depth sensor (Maxbotix MB7384) to measure water levels (shown in Figure 1, site C). At the time of writing, sensor nodes can be constructed using less than $500 USD of parts.

To control discharges throughout the creekshed, stormwater basins are retrofitted with one of two valves: (i) a 0.3 m diameter butterfly valve (Dynaquip MA44) (Figure 1, site B) or (ii) a 0.3 m gate valve (Valterra 6912) mated to a linear actuator (AEI 6112CH) (Figure 1, site A). Each control valve is connected to a sensor node. The valves are actuated by the microcontroller and powered by rechargeable 12 V sealed lead-acid batteries. Solar panels allow the control sites to operate without line power. Assuming that the valve can be attached to a basin's outlet without structural modification, each control site can be constructed using less than $3500 USD of parts at the time of writing.

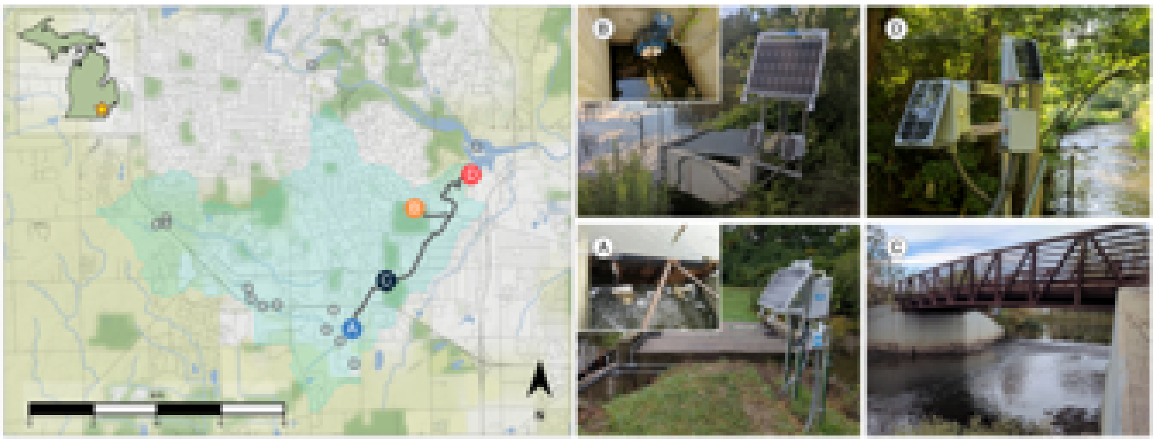


Figure 1. Overview of the study area. The map (**left**) shows the location of relevant control and sensor sites, additional sensor sites (light grey), flow paths between each site (dark grey), and the contributing area of the watershed (light blue). Site images (**right**) show the two control sites (**A** & **B**) along with two downstream sensor locations (**C** & **D**).

Remote control of valves and sensors is implemented using a *polling* scheme, in which field-deployed nodes request commands from a remote server (Figure 2). To conserve power, nodes spend most of their time in a deep sleep state, consuming only 1–10 µA of current. Upon waking up, each node takes sensor readings and transmits the readings to a cloud-hosted time series database (InfluxDB) via authenticated (and optionally encrypted) *HTTP* requests. Before going back to sleep, the node polls a set of commands from a dedicated feed in the same database. The commands may include, but are not limited to, changing the sampling frequency, triggering additional sensor readings, or opening a valve. Operations can be cancelled and rescheduled either by the application or by an operator. This is useful if, for example, the application detects that a control action was not successfully executed and that pending operations need to be rescheduled. Most importantly, the database supports modern web service standards and application programming interfaces (APIs), which allow the control logic to be quickly implemented via simple web applications. These applications can be written in any number of popular programming languages (Python, Matlab, etc.). This feature improves flexibility, reduces reliance on low-level firmware updates, and allows for the seamless integration of external data sources, such as public weather forecasts [6,15].

For the experiments described in this study, field devices in the creekshed are controlled using a simple Python web application. This application can be executed in either automatic or manual mode. In automatic mode, the application queries water level sensor feeds, rainfall forecasts, and external flow measurements from a publicly-listed measurement station at the outlet of the creekshed (USGS 04174518 (https://waterdata.usgs.gov/usa/nwis/uv?04174518)). Based on these sensor readings, new commands are then written to the database to open and close valves. In manual operation, a predefined set of commands is written to the database, then subsequently executed by the field device. For this study, the manual operation mode is used. The application toolchain is implemented on an Amazon Web Services (AWS) medium-sized linux Elastic Compute Cloud (EC2) instance.

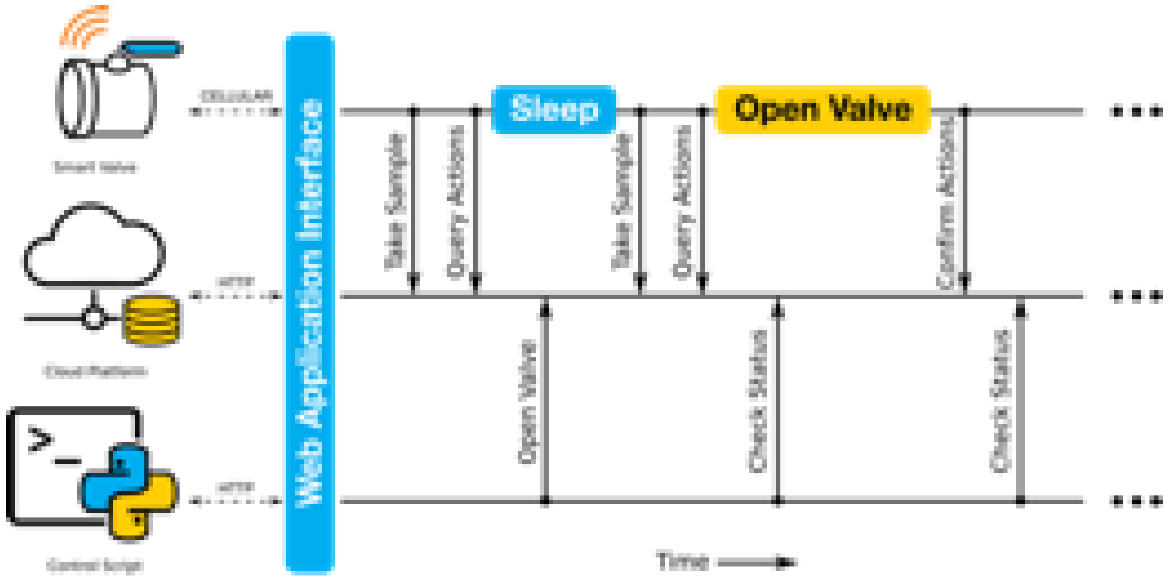


Figure 2. Control system architecture. Field-deployed nodes use a polling system to download and execute commands issued from a remote server. Control actions can be specified manually, or through automated web applications and scripts.

3. Characterizing Control Actions

Before evaluating potential control strategies, we first characterize the ability of each control site to shape downstream flows. Specifically, we quantify the travel time P and decay time D of various waves as they move between the originating control site and the outlet of the watershed. The characterization is accomplished by releasing pulses of different durations from each stormwater basin and then observing the resulting waves that these pulses generate downstream. To limit confounding effects caused by rainfall, these experiments are carried out during dry conditions (at least four days following a storm). Figure 3 shows a 1-h release, 4-h release, and 48-h release from retention basin A (shown left to right, respectively). The 48-h release empties the retention basin, meaning that this release characterizes the maximum possible output from site A. The travel times for each wave from site A to site C are approximately 3.5 h (time to start of rise) and 6–8 h (time to peak), with faster rise times for the larger releases due to nonlinearities in the speed of wave propagation. The decay times for each release are 6 h, 18 h and 44 h, respectively. From this experiment, it can be seen that the maximum change in flow that site A can generate at the outlet is roughly 0.17 m^3/s. Similar experiments are used to characterize site B. From these experiments, we estimate average travel times from site B to the outlet of 1.5 h (time to start of rise) and 1.8 h (time to peak), with an average decay time of 3 h, and a maximum change in flow of approximately 0.2 m^3/s.

In addition to release duration, sites are also characterized with respect to the hydraulic head (water level) of the originating retention basin. Figure 4 shows the result of releasing three 1-h pulses from site B, without allowing the basin to refill between releases. While the same duration is used for each release, the hydraulic head (stored volume) of the retention basin decreases with each pulse. Thus, the resulting wave becomes smaller with each successive opening of the valve, even though the same input signal is used. In spite of this difference, the travel times and decay times of the waves remain consistent between each release. The magnitude of the resulting wave varies from roughly 0.2 m^3/s to 0.13 m^3/s, depending on the water level in the basin.

Although retention basin B is significantly smaller than retention basin A, it can produce a comparable change in flow at the watershed outlet (approximately 0.2 m^3/s). This effect can be attributed to two main factors. First, site B is located closer to the outlet (3.0 km as opposed to 5.9 km for site A), meaning that the wave is subject to less hydraulic dispersion. Second, the retention basin at site B is elevated higher above the receiving stream, meaning that flows exit the control structure more rapidly than flows released from site A. Thus, compared to site A, site B produces short pulses with a rapid onset and large peak. Despite its relatively smaller volume, control actions from site B must thus be tailored to avoid generating flashy flows at the outlet.

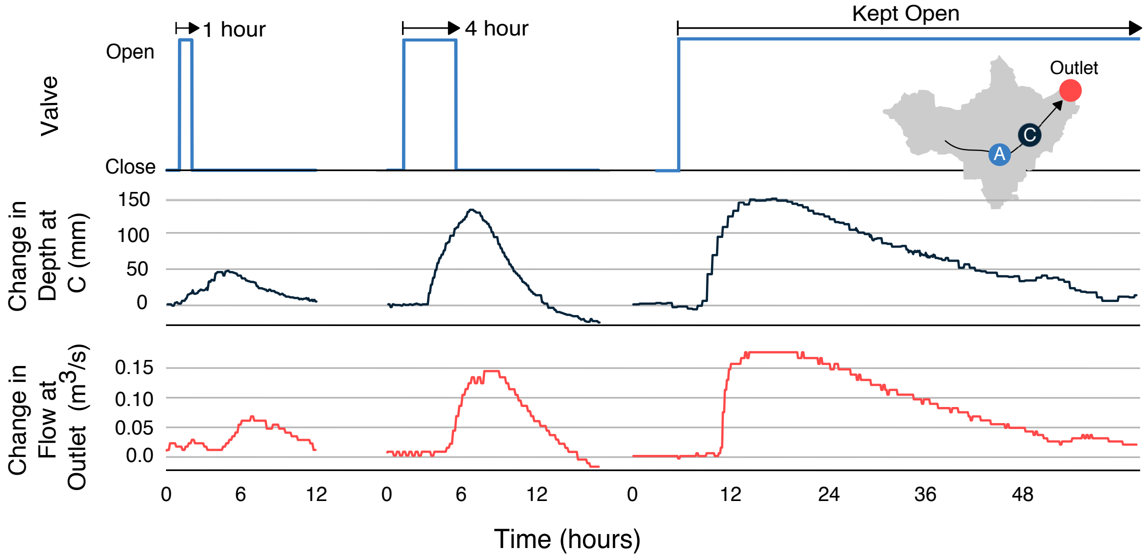


Figure 3. Characterization of control actions from site A. In the first two experiments, the valve at site A is opened for 1-h and 4-h durations. For the third experiment, the valve is held open indefinitely. The resulting waves travel through a constructed wetland (site C) before arriving at the outlet of the watershed. Wave depth (black line) is measured at the wetland, while flow rate (red line) is measured at the outlet.

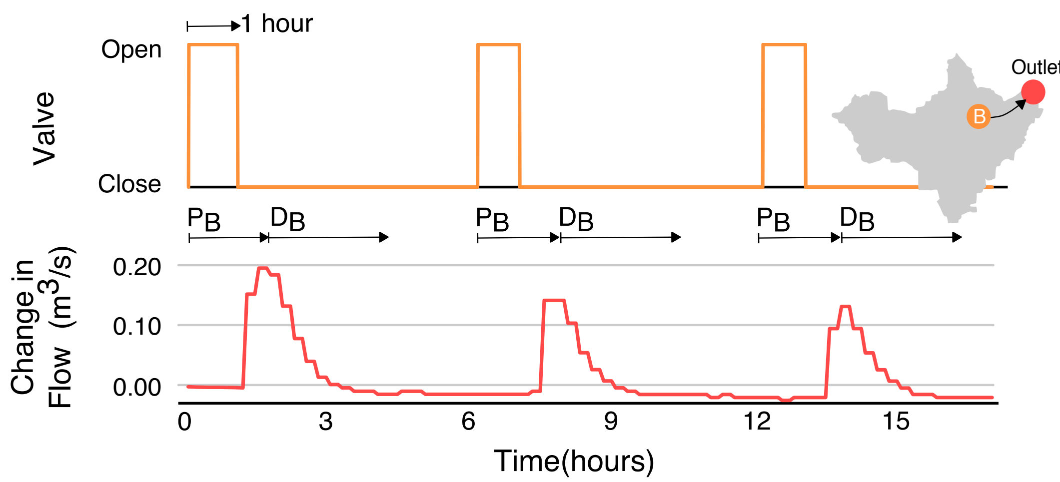


Figure 4. Characterization of control actions originating from site B. Three subsequent pulses are released. While the duration of each control pulse is the same (1 h), the magnitude of the flow at the outlet decreases because the hydraulic head (pressure) in the basin is reduced with each release.

One crucial result of these experiments is that for the purposes of control, nonlinearities in wave propagation can be safely ignored. Shallow-water waves exhibit a nonlinear relationship between wave height and wave speed, meaning that larger waves propagate faster [16]. If these nonlinearities were significant, then control strategies would need to account for changes in travel time due to (i) variations in release durations; (ii) variations in basin head; and (iii) superposition of waves originating from different locations. For the system examined in this study, the effect of these nonlinearities is small. Namely, while nonlinearities in wave propagation affect the shape of the resulting hydrograph (skewing the peak toward the left), they do not significantly affect the bulk travel time of an isolated wave. Specifically, the travel times for site A and site B remain consistent (3.5 h and 1.5 h, respectively) despite scheduling releases of different durations and magnitudes. This result is consistent with findings from previous studies that use linear dynamics for stormwater system control [17–19]. Thus, for the scale of our creekshed the travel time of a wave originating at an upstream stormwater basin can be considered independent of both the amount of water released and the water level of the originating basin. Moreover, superposition of two waves from two parallel sources does not effect a noticeable change in bulk wave speed. This result suggests that for the purposes of control, the channel network may be approximated as a linear system in which waves originating from each retention basin can be superimposed in order to produce a desired output hydrograph downstream.

By characterizing the downstream response to various impulsive inputs, these initial experiments yield a set of "building blocks" that are subsequently used to achieve more complex control objectives at the watershed outlet. While the propagation of waves within a channel network is described by nonlinear equations, we find that a linear system approximation adequately describes the dynamics needed to generate control strategies. Thus, the characterization experiments described in this section are conceptually analogous to quantifying the unit impulse response of a linear system. This framework suggests that desired waveforms can be generated via simple linear combinations of known input signals. With this conceptual model in hand, we carry out a number of control experiments to showcase the utility of the stormwater control network. First, we show how pulse-width modulation of a valve can be used to produce a flat hydrograph that meets but does not exceed a given flow threshold. Next, we show how valve releases can be timed to generate synchronized and desynchronized waves at the outlet. These experiments provide recipes for managing releases from upstream retention basins while simultaneously fostering desirable flow conditions downstream.

4. Set-Point Hydrographs

Real-time control can be used to flatten downstream hydrographs, helping to reduce erosion and maintain healthy aquatic ecosystems. In passive stormwater systems, hydrographs often exhibit a distinct peak, preceded by a rapid rise and followed by a slower decay. While typically associated with rain events, this phenomenon can also be observed when water is released from a retention basin (see Figures 3 and 4). Peak flows that exceed downstream capacity will often lead to flooding. Furthermore, urban streams can become unstable if a critical flow velocity or flow rate is reached [20]. Exceedance of these thresholds may lead to ecological damage and stream erosion, as well as the mobilization of sediments. These sediments in turn may carry nutrients, metals and other pollutants downstream, impairing water quality and promoting the growth of algal blooms [21]. This particular impairment underpins the major challenge of "urban stream syndrome", forcing many cities to spend millions of dollars to reduce downstream flow rates [22,23]. While active control has been proposed as a means to condition stormwater flows, the specific control strategies needed to achieve stable flow conditions within an urban watershed are currently not well understood.

To address this challenge, a sequence of control actions is designed to yield a constant set-point condition at the outlet of the watershed. Specifically, we aim to create a flat hydrograph, for which the flow rate remains close to (but does not exceed) a specified value. While the set-point used in this experiment is chosen arbitrarily, this threshold may be chosen to control for objectives related to downstream flooding and water quality—for instance, ensuring that the critical flow threshold for sediment transport is not surpassed. To achieve a constant set-point flow rate, we derive inspiration from *pulse-width modulation*—a method used in electrical systems to generate analog signals from discrete digital pulses. Isolated pulses of water are emitted from the control site, spaced apart such that the arrival time of each wave overlaps with the receding limb of the prior wave. As the pulses travel through the channel network, they disperse, causing the individual waves to overlap and combine. The resulting superposition of partly-dispersed waves results in an approximately constant flow rate.

As seen in the hydrograph response (Figure 5), the "flat hydrograph" objective is achieved by modulating the valve position in successive 30-min pulses. The flows at the outlet remain approximately flat, without significantly exceeding a setpoint of 0.04 m^3/s. Of course, the shape is not perfectly flat, given the large distance between the two sites and nonlinearities inherent in wave propagation. However, these experimental results show that active modulation of a valve can produce highly stable flow conditions downstream that would not be possible using passive infrastructure alone. In a real-world scenario, this control strategy could be used to drain a watershed as fast as possible without exceeding critical flood conditions downstream. Minimizing the change in flows downstream also reduces the likelihood of stream erosion. From our prior studies in this creekshed that were not affected by real-time control [24], it can be estimated that pollutant concentrations during this flat stage were no greater than 127 mg/L for sediment and 0.209 mg/L for total phosphorus. For comparison, keeping the valve open would have resulted in concentrations of at least 390 mg/L for sediment and 0.618 mg/L for total phosphorus. By modulating the valve position to achieve a relatively flat and steady outflow, the control actions likely reduced the total mass of solids and phosphorus that would otherwise contribute to ecological damage and harmful algal blooms. Future studies will confirm and refine these estimates by measuring real-time water quality changes that result from control.

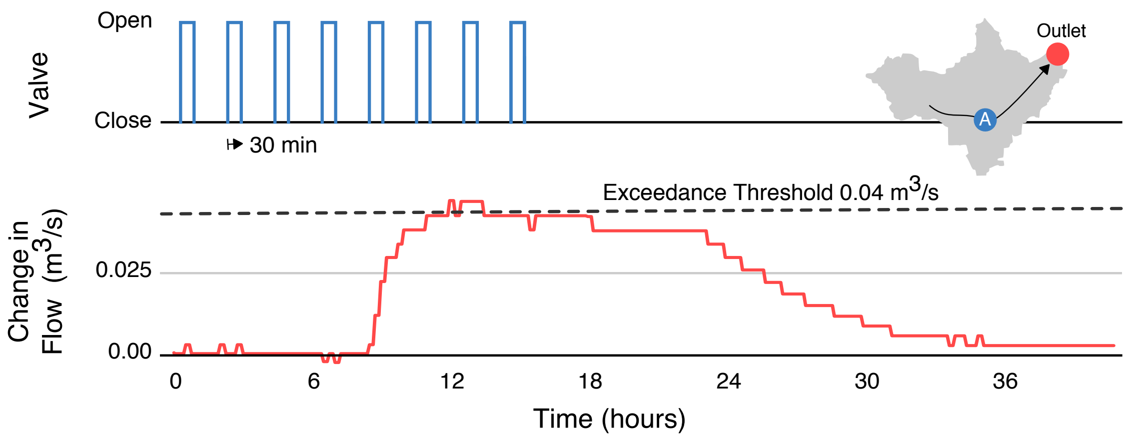


Figure 5. Generating a set-point hydrograph. Small, evenly spaced pulses (30-min duration) are released from the controlled basin. The pulses disperse as they travel through the 6 km-long stream, leading to a relatively flat response at the outlet of the watershed.

5. Coordinated Releases between Multiple Control Sites

Motivated by the larger goal of watershed-scale control, a final experiment is devised to evaluate the level of precision that can be achieved when coordinating releases from multiple sites. Namely, we schedule releases from the two controlled basins in order to produce synchronized and interleaved pulses at the outlet. Before running the experiment, we first determine the control signals needed to generate the combined and interleaved waves, respectively, by assessing the travel time and decay time of waves released from each retention basin. Figure 6 shows the hydrographs resulting from 1-h pulses released simultaneously from site A and site B. Based on the travel times of each wave, it can be seen that in order to achieve a synchronized wave at the outlet, a 1-h release from site B must be scheduled approximately six hours after a 1-h release from site A. Conversely, to achieve an interleaved pattern at the outlet, the following pulse train can be used: (i) release a 1-h pulse from site A; (ii) release a pulse from site B approximately 12 h later; (iii) release a pulse from site A after waiting an additional four hours; and (iv) repeat the pattern starting at step (ii).

Once the input signals required to produce each desired shape are known, we schedule a series of commands to be executed by each valve. The experiment is divided into two stages. During the first stage, flows from the control sites are released such that the peaks of the hydrographs overlap. In the second stage of the experiment, the flows are released off-phase, such that the flows arriving from one site begin exactly when the flows from the other site recede. Figure 7 shows the result of this experiment, with the overlapping waves occurring from hours 6 to 15, and the interleaved waves occurring from hours 15 to 44. As hypothesized earlier, the superposition of waves is approximately linear. In other words, the maximum change in flow is approximately equal to the sum of the maximum flow of each component wave. Moreover, the superposition of the two waves does not appear to appreciably change the bulk travel time.

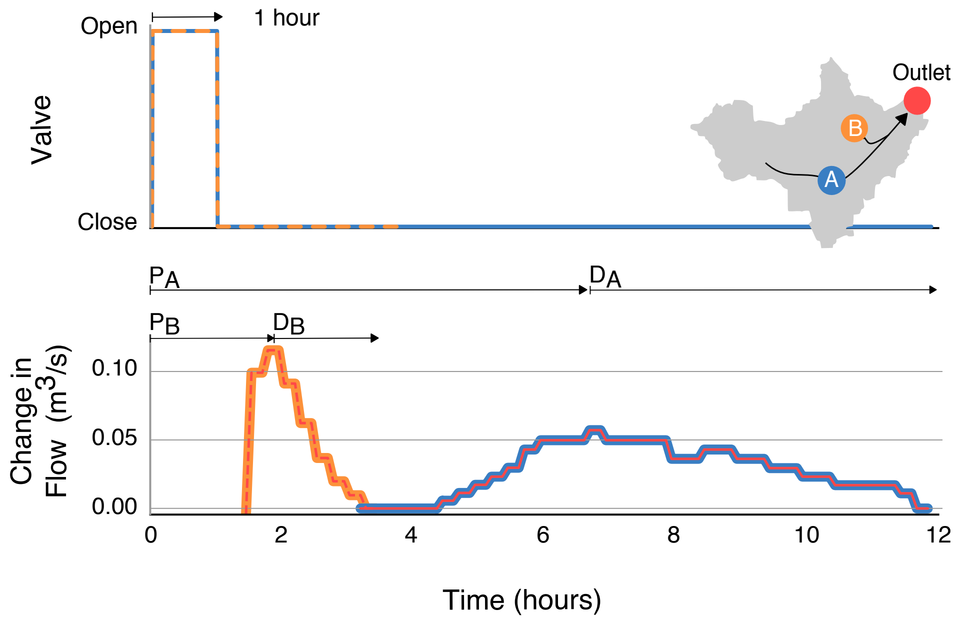


Figure 6. Flows at outlet of watershed resulting from 1 h releases from each control site. Time to peak *P*, magnitude, and decay time *D* for each release are labeled.

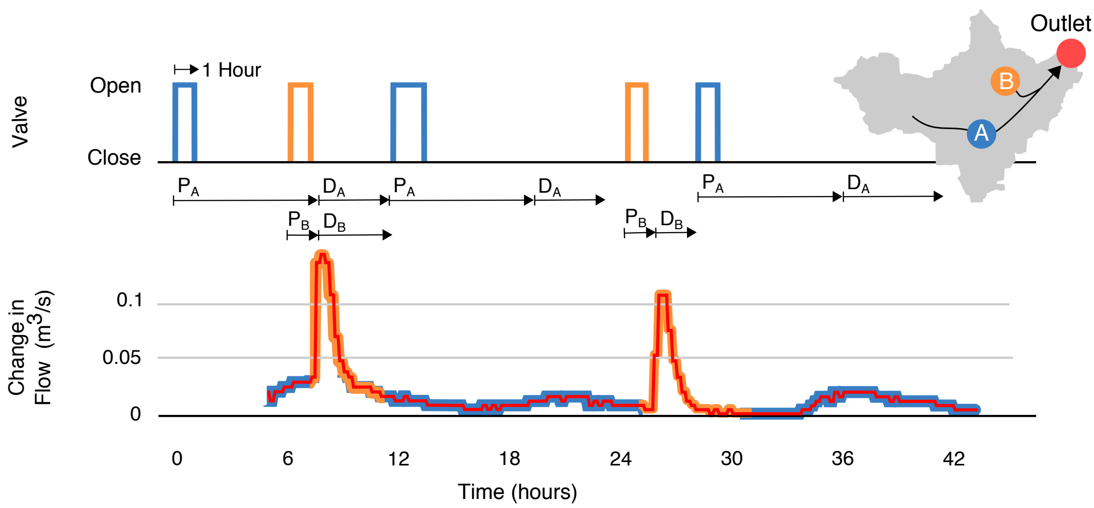


Figure 7. Superposition and interleaving of waves from retention basins A and B. Overlapping waves (coincident peaks) are generated from hours 6 to 12. Interleaved waves (off-phase peaks) are generated from hours 18 to 44.

This experiment shows that real-time control of stormwater systems can achieve precise control over downstream flow conditions, and it also suggests a strategy for coordinating releases in order to remove stormwater from retention basins while simultaneously achieving target flow conditions downstream. Like the set-point experiment, an interleaving control pattern can be used to de-water upstream retention basins without exceeding a particular flow threshold downstream. When waves generated by several upstream retention basins combine, they can generate large, flashy flows at a downstream location. This in turn can contribute to erosion of the surrounding channel. For this reason, it is desirable to avoid the collision of waves from two different upstream sources. By interleaving flows from upstream retention basins, one can free up capacity in the system without generating adverse flow conditions downstream. More broadly, the results of this experiment demonstrate the fine level of flow control that can be achieved across urban watersheds using a low-cost sensor and control network. While the underlying control logic only uses rudimentary time-of-travel metrics, it nonetheless produces desirable flow regimes that would be difficult to achieve with passive infrastructure alone. As such, this experiment builds a foundation for more complex control strategies by verifying that the watershed responds consistently and predictably to individual control actions. This result suggests that future studies may one day demonstrate more complex, possibly near-arbitrary, hydrograph shapes. Time of travel may not be sufficient for such approaches, however, and more complex and analytical control techniques should be considered.

6. Conclusions

This study shows how internet-connected stormwater control valves can be used to shape streamflows within a large urban watershed. To our knowledge, this study is the first to document how coordinated releases between multiple stormwater control sites can satisfy system-scale watershed performance goals—such as maintaining downstream flow at a constant rate or preventing sediment transport. Building on an existing wireless sensor network, we demonstrate how static stormwater retention basins can be retrofitted with internet-controlled valves to enable active control at a low cost. Characterizing the system in a series of exploratory experiments, we find that a linear approximation is sufficient to describe the downstream response associated with a given input. Next, we use the system to generate two flow conditions downstream: (i) a set-point hydrograph in which flow is maintained at a roughly constant rate; and (ii) a series of overlapping and interleaved waves. We find that pulse-width modulation of upstream valves generates a flat downstream response. Similarly, interleaving of discharges provides an effective tool for emptying upstream retention basins without inducing flashy flows downstream. In addition to demonstrating the precision of the control system, these experiments suggest strategies for managing stormwater transfers across a watershed while maintaining desired flow conditions. To make the smart stormwater system described in this paper accessible to water managers worldwide, all hardware, software and documentation for this project are made available at `open-storm.org`.

Author Contributions: A.M., M.B., B.W. and B.K. jointly conceived the study, designed the experiments, analyzed the data, created the figures, and wrote the paper.

Funding: This research was funded by the Great Lakes Protection Fund (Grant 1035), the US National Science Foundation (Grant 1737432) and the University of Michigan.

Acknowledgments: The authors acknowledge the support of the Great Lakes Protection Fund for making the technologies and experiments possible. They also acknowledge the support of the US National Science Foundation and the University of Michigan. We would like to thank the Washtenaw County Water Resources Commissioner's office, the city of Ann Arbor, and the Huron River Watershed council. Finally we would like to thank Rachel Menge, Nick Jeng, Adarash Mishra, Gerardo Longoria, Jackie Fortin, and the Real-Time Water Systems Lab.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kerkez, B.; Gruden, C.; Lewis, M.; Montestruque, L.; Quigley, M.; Wong, B.; Bedig, A.; Kertesz, R.; Braun, T.; Cadwalader, O.; et al. Smarter Stormwater Systems. *Environ. Sci. Technol.* **2016**, *50*, 7267–7273. [CrossRef] [PubMed]
2. Wright, J.; Marchese, D. Briefing: Continuous monitoring and adaptive control: The 'smart' storm water management solution. *Proc. Inst. Civ. Eng. Smart Infrastruct. Constr.* **2017**, *170*, 86–89. [CrossRef]

3. Montestruque, L.; Lemmon, M.D. Globally Coordinated Distributed Storm Water Management System. In Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks (CySWater'15), Seattle, WA, USA, 13–16 April 2015; ACM Press: New York, NY, USA, 2015.
4. Bronstert, A.; Niehoff, D.; Bürger, G. Effects of climate and land-use change on storm runoff generation: Present knowledge and modelling capabilities. *Hydrol. Process.* **2002**, *16*, 509–529. [CrossRef]
5. Stocker, T. *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2014.
6. Bartos, M.; Wong, B.; Kerkez, B. Open storm: A complete framework for sensing and control of urban watersheds. *Environ. Sci. Water Res. Technol.* **2018**, *4*, 346–358. [CrossRef]
7. Mullapudi, A.; Wong, B.P.; Kerkez, B. Emerging investigators series: Building a theory for smart stormwater systems. *Environ. Sci. Water Res.* **2017**, *3*, 66–77. [CrossRef]
8. Middleton, J.R.; Barrett, M.E. Water Quality Performance of a Batch-Type Stormwater Detention Basin. *Water Environ. Res.* **2008**, *80*, 172–178. [CrossRef] [PubMed]
9. Roman, D.; Braga, A.; Shetty, N.; Culligan, P. Design and Modeling of an Adaptively Controlled Rainwater Harvesting System. *Water* **2017**, *9*, 974. [CrossRef]
10. Klenzendorf, B.; Barrett, M.; Christman, M.; Quigley, M. *Water Quality and Conservation Benefits Achieved via Real Time Control Retrofit of Stormwater Management Facilities near Austin, Texas*; Technical Report; OptiRTC: Boston, MA, USA, 2015.
11. Emerson, C.H.; Welty, C.; Traver, R.G. Watershed-Scale Evaluation of a System of Storm Water Detention Basins. *J. Hydrol. Eng.* **2005**, *10*, 237–242. [CrossRef]
12. Schütze, M.; Campisano, A.; Colas, H.; Schilling, W.; Vanrolleghem, P.A. Real time control of urban wastewater systems—Where do we stand today? *J. Hydrol.* **2004**, *299*, 335–348. [CrossRef]
13. Mollerup, A.L.; Mikkelsen, P.S.; Thornberg, D.; Sin, G. Controlling sewer systems—A critical review based on systems in three EU cities. *Urban Water J.* **2016**, *14*, 435–442. [CrossRef]
14. Lawson, R.; Riggs, E.; Weiker, D.; Doubek, J. *Total Suspended Solids Reduction and Implementation Plan for Malletts Creek: October 2011–September 2016*; Technical Report; Huron River Watershed Council: Ann Arbor, MI, USA, 2016.
15. Wong, B.P.; Kerkez, B. Real-time environmental sensor data: An application to water quality using web services. *Environ. Model. Softw.* **2016**, *84*, 505–517. [CrossRef]
16. Kinnmark, I. *The Shallow Water Wave Equations: Formulation, Analysis and Application*; Springer Science & Business Media: Berlin, Germany, 2012; Volume 15.
17. Litrico, X.; Fromion, V. Simplified Modeling of Irrigation Canals for Controller Design. *J. Irrig. Drain. Eng.* **2004**, *130*, 373–383. [CrossRef]
18. Marinaki, M.; Papageorgiou, M. Linear-quadratic regulators applied to sewer network flow control. In Proceedings of the 2003 European Control Conference (ECC), Cambridge, UK, 1–4 September 2003.
19. Garcia, L.; Barreiro-Gomez, J.; Escobar, E.; Tellez, D.; Quijano, N.; Ocampo-Martinez, C. Modeling and real-time control of urban drainage systems: A review. *Adv. Water Res.* **2015**, *85*, 120–132. [CrossRef]
20. Bledsoe, B.P. Stream erosion potential and stormwater management strategies. *J. Water Res. Plan. Manag.* **2002**, *128*, 451–455. [CrossRef]
21. Michalak, A.M.; Anderson, E.J.; Beletsky, D.; Boland, S.; Bosch, N.S.; Bridgeman, T.B.; Chaffin, J.D.; Cho, K.; Confesor, R.; Daloglu, I.; et al. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6448–6452. [CrossRef] [PubMed]
22. Schilling, J.; Logan, J. Greening the rust belt: A green infrastructure model for right sizing America's shrinking cities. *J. Am. Plan. Assoc.* **2008**, *74*, 451–466. [CrossRef]
23. Wise, S. Green infrastructure rising. *Planning* **2008**, *74*, 14–19.
24. Wong, B.P.; Kerkez, B. Adaptive measurements of urban runoff quality. *Water Resour. Res.* **2016**, *52*, 8986–9000. [CrossRef]

Spring 2018

INTERNATIONAL FRONTIERS OF ENGINEERING

The BRIDGE

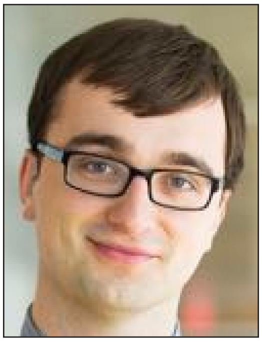
LINKING ENGINEERING AND SOCIETY

NATIONAL ACADEMY OF ENGINEERING

The mission of the National Academy of Engineering is to advance the well-being of the nation by promoting a vibrant engineering profession and by marshalling the expertise and insights of eminent engineers to provide independent advice to the federal government on matters involving engineering and technology.

Instead of costly construction, new technologies may make it possible to use existing water systems much more effectively.

Building Smarter Water Systems



Branko Kerkez is an assistant professor of civil and environmental engineering at the University of Michigan and founder of Open-Storm.org.

Branko Kerkez

In the era of self-driving cars, digital assistants, and other smart things, can the same level of autonomy and "intelligence" be embedded in water systems? Such technologies have the potential to dramatically reshape adaptation to some of the greatest water challenges, such as floods and droughts. Software-updatable water systems are well within reach, promising to enable highly cost-effective water infrastructure that dynamically redesigns itself in response to changing needs and uncertain inputs.

Background

Recent news coverage has concerned droughts in the American West, contaminated tap water in the Midwestern United States, and floods in California, New England, and the southeastern coastal states. In fact, flooding is the leading cause of extreme weather fatalities in the United States (Vörösmarty et al. 2010).

At the same time, dry regions struggle to find new and clean sources of water. By many estimates, the capture of stormwater in Los Angeles could offset 10 billion gallons per storm (Monte 2015), a large portion of the city's annual water budget. Unfortunately, most of this water is washed into the Pacific Ocean within hours after a storm. Capturing it requires distributed storage and treatment, both of which are limited.

The scarcity of water in the West and other parts of the United States is not helped by the age of the infrastructure that conveys it. By many estimates, nearly 20 percent of treated water is lost through old and leaky pipes, some of which were constructed at the turn of the last century (US EPA 2013).

Aside from water losses, leaks present significant revenue and energy challenges—treatment and conveyance across water systems account for almost 2 percent of the US energy budget (US EPA 2017). Thus, the need to better track and address challenges to water supplies and water distribution is imperative to maintain safe, leak-free, and energy-efficient water systems.

Status of US Stormwater Systems

Recent flash floods (e.g., Hunter 2016) are an all too common, dramatic example that aging infrastructure is struggling to keep up with changing and increasingly severe weather patterns. In addition, nutrients, metals, and many other pollutants are washed from urban surfaces when it rains, eventually ending up in streams, lakes, and oceans (Barco et al. 2008; Finkebine et al. 2000; Wang et al. 2001). And many parts of the country are dealing with chronically impaired coastlines due to algal blooms, which are driven, in part, by urban stormwater runoff (Carey et al. 2014; Doughton 2015; Wines 2014).

Infrastructure Challenges

To address flooding and water quality challenges, cities build and maintain complex networks of distributed stormwater assets such as pipes, canals, and basins. This infrastructure reduces flooding by moving water away from roads and buildings during storms. It also includes natural elements, such as wetlands and raingardens, to capture sediments and dissolved pollutants before they can be discharged to downstream ecosystems.

In some communities, however, stormwater and wastewater are combined, sharing the same pipes. With these systems, large storms can lead to sewer overflows into natural waterways, introducing viruses, bacteria, nutrients, pharmaceuticals, and other pollutants.

Many US stormwater systems were designed at times of less stringent regulations and for populations different from those they now serve. Most are approaching, or already exceed, their design life and face problems similar to those of America's other ailing infrastructure. The American Society of Civil Engineers (ASCE) report card has given US stormwater infrastructure a near failing grade (ASCE 2017), and the problem is echoed in the National Academy of Engineering's Grand Challenge to "restore and improve urban infrastructure" (NAE 2008).

Costly and Piecemeal Fixes

The distributed nature and massive size of most municipal stormwater infrastructure makes it impossible to dig up and resize the entire system. Instead, problems are often fixed one by one through expensive construction projects that are difficult to change afterward. As such, most stormwater systems comprise an amalgam of distributed fixes that have been constructed over decades and rarely add up to an optimized whole.

> *Most stormwater systems are an amalgam of distributed fixes over decades that rarely add up to an optimized whole.*

One of the country's largest stormwater and sewer tunnels was recently built in Chicago, and a few other cities are following this billion-dollar trend (Evans 2015). But these impressive, large construction projects are a luxury for most communities, many of which face challenges in simply maintaining their existing systems.

Even in small communities, stormwater systems can quickly add up to hundreds of linear miles of assets that must be maintained and repaired. In many US cities low or no fees are charged for stormwater services, in comparison to drinking water or sanitary sewer systems. With highly limited revenue streams, solutions that rely on new construction cannot keep pace with evolving community needs and uncertain weather. As cash-strapped communities seek more resilient infrastructure solutions, novel alternatives must be explored.

Technologies for Smarter Water Systems

The role of information technology in managing water supplies and drinking water distribution systems has been made evident in a number of applications. These include sensors for the management of large hydropower and agricultural basins (Kerkez et al. 2012; Rheinheimer et al. 2016), real-time pump optimization and leak localization systems (Stoianov et al. 2007; Whittle et al. 2010), drinking water contamination detection (Ostfeld and Salomons 2004; Storey et al. 2011), and even residential WiFi irrigation widgets.

A burgeoning smart water industry is beginning to fill the needs of modern water utilities and municipalities. The overall technology outlook for water supplies and drinking water systems is promising, notwithstanding much work that must still be done in the development of water quality sensors for important contaminants, such as lead, bacteria, and viruses.

Not all water sectors are embracing technology at the same pace. This is particularly true across stormwater systems, an often overlooked and possibly the most poorly funded subset of urban water infrastructure (Kea et al. 2016). Instead of relying on costly construction, new technologies may make it possible to use existing systems much more effectively.

Industrial and academic efforts are under way to demonstrate the benefits of real-time sensing, computation, and wireless connectivity for the management of urban watersheds. The driving hypothesis behind this work is that smart stormwater systems will vastly shrink the size of infrastructure required to manage runoff pollution and other impacts of changing weather. This approach will dynamically repurpose existing stormwater systems by adapting them on a storm-by-storm basis.

Researchers at the University of Michigan have been spearheading the development of open source technologies that enable watersheds to be retrofitted for sensing and real-time control (figure 1). These technologies and associated case studies are being shared through Open-Storm.org, an open source consortium of academic, industry, and municipal partners that seeks to provide a complete, "batteries included" template for the development of smart watersheds to combat flood-

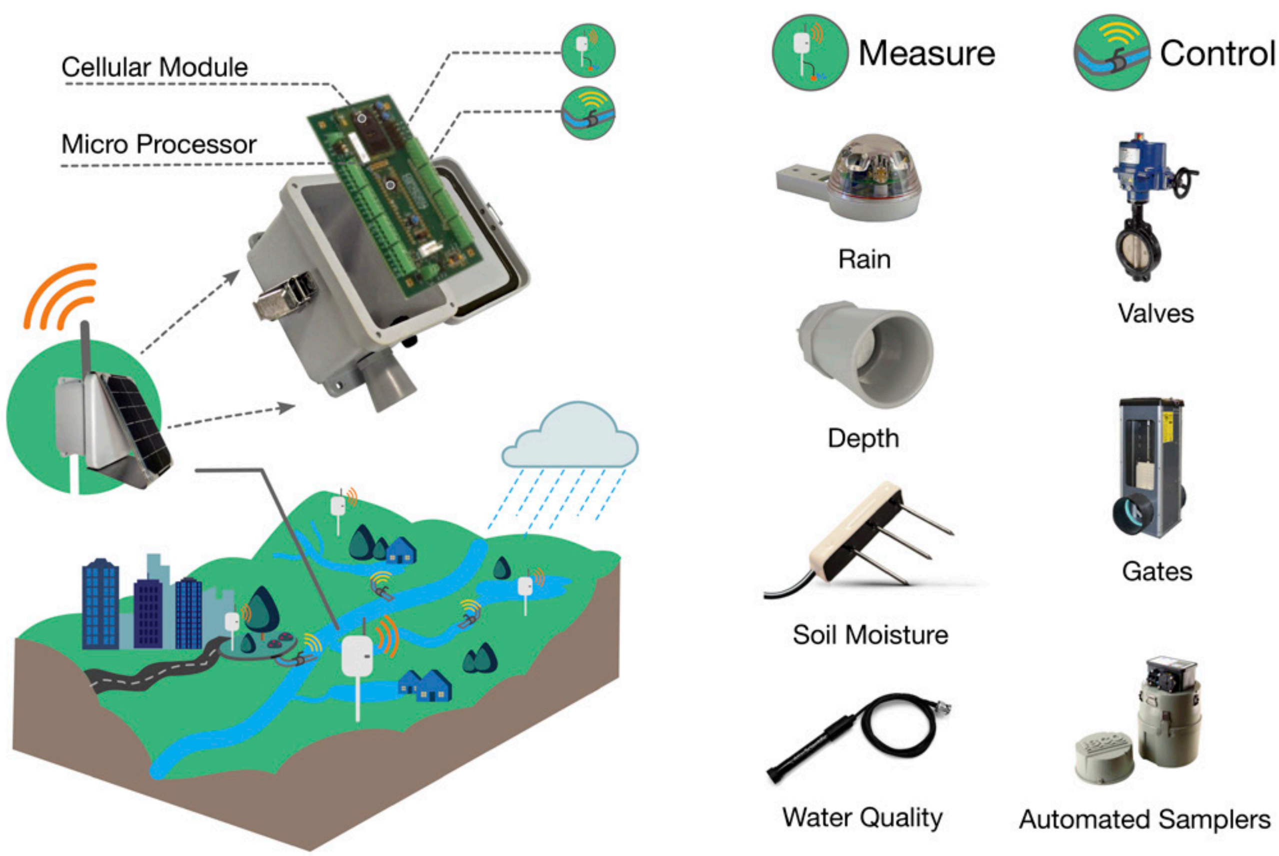


FIGURE 1 Technologies for the sensing and control of watersheds. A full wireless network is composed of many individual sensor nodes (left column) dispersed throughout a watershed, typically at least one per square mile. Each sensor node contains a low-power microprocessor, which collects measurements from a variety of connected sensors. The same unit can be used to control flows using gates, valves, or pumps (right column). A wireless radio, often a cellular module, allows sensor measurements or commands to be transmitted using cloud-hosted services in real time. Reprinted from Bartos et al. (2018) with permission from the Royal Society of Chemistry.

ing and improve urban water quality. The technologies include rapidly deployable wireless sensor nodes for the measurement of urban water flows, water quality, soils, and weather. These are complemented by cloud-hosted data services that allow measurements to be analyzed immediately, providing real-time information on the "health" of both the watershed and the infrastructure.

Commands can also be transmitted from the cloud to the watershed to change the configuration of infrastructure in real time. This is enabled by wirelessly controlled valves, gates, and pumps that can be quickly and easily attached to existing stormwater infrastructure, such as pipes, retention basins, or wetlands. Water levels at retrofitted sites can be safely controlled to release water based on sensor measurements or real-time weather forecasts. By dynamically controlling flows across sites, system-level storage can be adapted to the unique nature of any given storm. This allows infrastructure to be "redesigned" or updated in near real time without the need for new construction.

Even a single remotely controlled valve can provide major benefits. At a basin or pond, a valve can be used to control flooding and reduce stress on downstream systems—a simple control algorithm closes the valve during a storm and opens it before the next storm starts. Storing water locally during a storm temporarily removes it from downstream areas that may otherwise be prone to flooding. The captured water can be directly used for irrigation in adjacent neighborhoods or injected into underlying aquifers to replenish the groundwater table. Some level of treatment can even be achieved at the controlled site by promoting the capture of sediment-bound or dissolved pollutants through settling or natural treatment. Many of these benefits can already be explored by cities and stormwater utilities through commercial real-time control solutions.

Perhaps the biggest benefit of real-time control is the ability to coordinate flows across entire infrastructure systems. For example, the upgraded large combined sewer system in South Bend, Indiana, features over 100 sensors and 10 control points working in tandem to reduce sewer overflows over an area of some 40 square miles.

System-level control promises effective coordination of the many distributed parts of urban stormwater infrastructure at the scale of entire watersheds.

Testbed Implementation

Smart stormwater control networks are being deployed across the Midwestern United States as testbeds for system-level control. They were developed and deployed through the support of the Great Lakes Protection Fund, in partnership with the cities of Ann Arbor and Toledo, Washtenaw County, the Universities of Toledo and Michigan, and Michigan Aerospace.

By dynamically controlling flows across sites, system-level storage can be adapted on a storm-by-storm basis.

The largest Open Storm testbed is in Ann Arbor (figure 2). The sensor network covers a 10 square mile urban watershed, in which some portions have a density of over 15 sensors per square mile. Multiple basins and wetlands have been retrofitted for control, sometimes in just one day by a group of university and high school students. Now in its second year of operation, the relatively inexpensive testbed has demonstrated the following:

- Measurements of soil moisture, water flows, rain, and water quality are transmitted in real time, analyzed, and made available to researchers and city engineers. This provides continuous performance insights that can be used to maintain or upgrade the existing infrastructure.
- Some of the largest controlled basins can store nearly 5 million gallons per storm, making it possible to control the majority of flows across the watershed.
- Real-time coordination of multiple valves has reduced flooding risk. Controlling how long water is held in basins after a storm has also reduced the output of sediments and nutrients from the watershed.

All of these benefits were achieved without new construction, but rather by using existing infrastructure more effectively.

Many scientific questions must now be addressed, but the testbed has proven to be effective and can serve as a blueprint for future smart watersheds. Researchers are also engaging with residents, city managers, and regulators on the value of these technologies to the community.

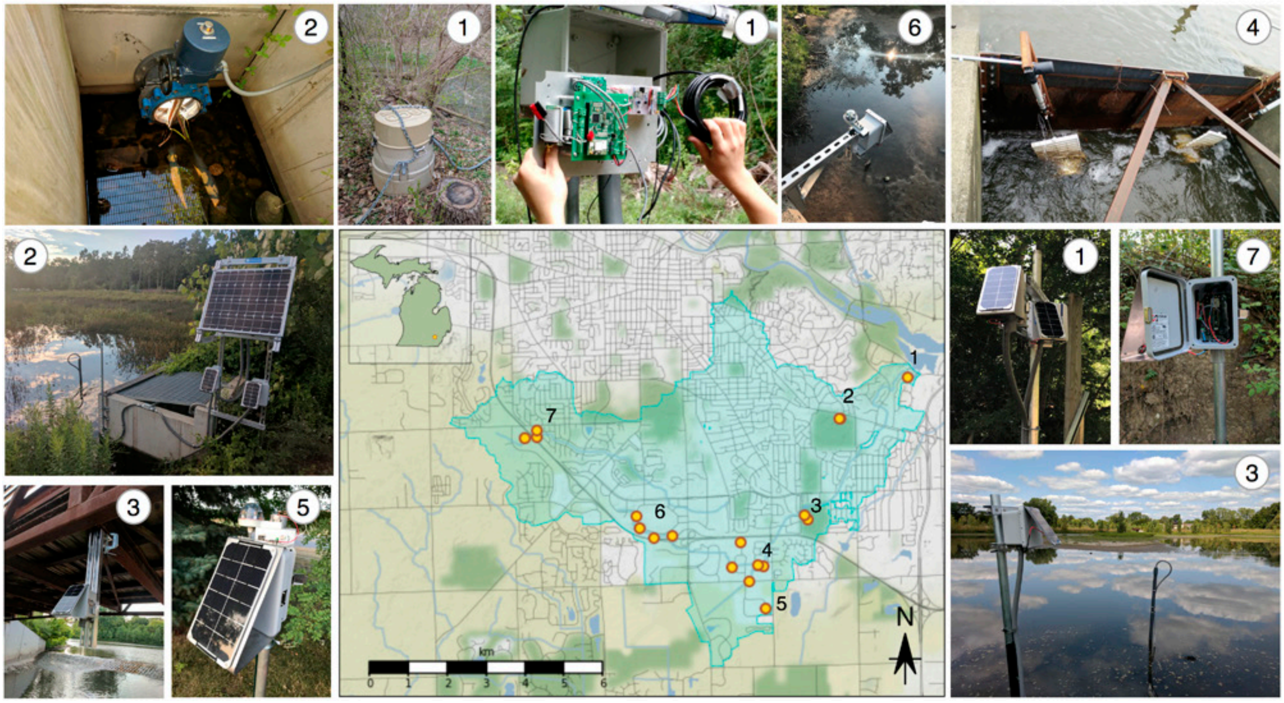


FIGURE 2 The real-time controlled watershed (blue region) in Ann Arbor, Michigan. Starting from the top right corner of the region, the numbers indicate water flow (in reverse order). (1) designates the outlet of the watershed, where urban stormwater flows enter the Huron River. Sensor nodes at each location measure multiple parameters: flows and water levels (locations 3, 4, 6, and 7), soil moisture and rainfall (location 5), and water quality (locations 1 and 3). The watershed is also controlled at a number of points using remotely controllable valves (locations 2 and 4). The current placement and density of sensors follow major infrastructure assets; the sensor network is continuously being expanded to capture the remaining parts of the watershed. Reprinted from Bartos et al. (2018) with permission from the Royal Society of Chemistry.

Simulating for Safety

Public safety demands that control algorithms be exhaustively validated and verified before control of watersheds is delegated to an autopilot. Efforts are focusing on achieving this through new simulation frameworks that allow large control networks to be modeled using state-of-the-art hydraulic solvers and control algorithms (Mullapudi et al. 2017).

Rather than running continuously, as is done in most water simulations, the water model can be halted after every step so that an external algorithm can "make decisions" on how to control valves or other assets. This allows a variety of algorithms to be evaluated in a physically realistic manner before being deployed on real-world systems (figure 3). These efforts are being freely shared through open source toolboxes to reach engineers in other disciplinary communities (e.g., control theory, machine learning) and encourage them to apply their own algorithms to combat flooding and improve water quality.

This simulation framework has already been used to begin investigating algorithms ranging across dynamical feedback control, market-based optimization, and reinforcement learning controllers. It can also help municipal managers decide on investments in real-time control.

Initial results obtained by University of Michigan researchers show that the addition of even a few control valves across urban watersheds allows the infrastructure to handle storms more than twice as large as those it was designed for. These findings also suggest that it may be possible to construct some new infrastructure at half the size when using real-time control, promising significant cost savings when compared to traditional methods.

Outlook

The adoption of smart water systems is no longer limited by technology (Kerkez et al. 2016), which has matured to the point that it can be ubiquitously deployed.

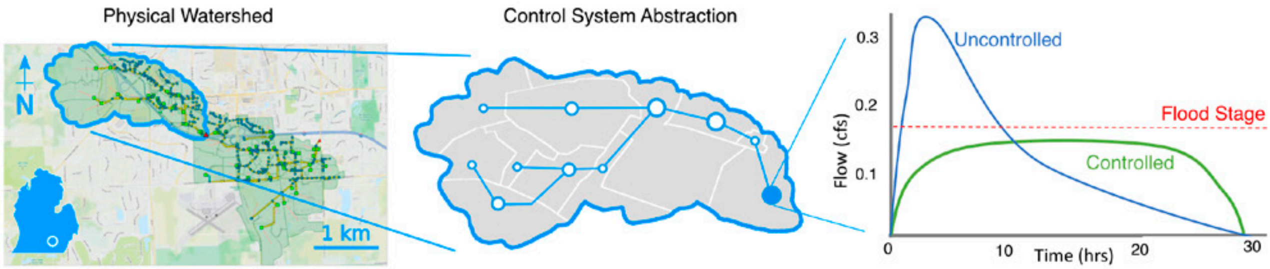


FIGURE 3 Control algorithms developed as part of the Open-Storm.org project allow watersheds to be abstracted into control models, which can then be used to simulate performance before the control systems are ported to the real world.

Rather, other barriers are becoming apparent and must be overcome to encourage and enable broader adoption:

- A new generation of control algorithms needs to be engineered, to ensure safe and autonomous operation at the scale of entire cities.
- More important, perhaps, social barriers to adoption need to be addressed. Doing so will require an understanding of how residents and decision makers perceive the benefits of smart versus traditional water systems.
- Finally, a new cross-disciplinary workforce needs to be trained and educated to be able to build, operate, and maintain this new generation of water systems.

New Control Algorithms

From a fundamental engineering perspective, there is a need to investigate how large water systems can be safely and autonomously controlled across the scales of entire cities and regions, requiring hundreds or thousands of control sites. To that end, extensive knowledge of water systems must be embedded in robust control algorithms.

While many control techniques for distributed systems have been successfully developed and applied in other engineering domains, an application-agnostic, one-size-fits-all approach may not be appropriate for water—what works for one type of water system may not work for another. For example, guidance on controlling flows across large spatial scales may come from research on reservoir operations (Wardlaw and Sharif 1999), water distribution systems (Cembrano et al. 2000), and control sewer systems (Marinaki and Papageorgiou 2005). However, the application of the same algorithms to stormwater may quickly reach limitations because of the need to account for challenges such as noisy sensor data, weather uncertainty, or the types of timescales, complexities, and feedbacks inherent in urban watersheds. Research is needed to determine which real-time control techniques will meet performance goals without risking the safety of nearby residents, property, and downstream ecosystems.

The need to solve these challenges presents an exciting opportunity for civil and environmental engineers to collaborate with colleagues in systems and computer science, electrical engineering, and other fields.

Public Buy-in

Even perfectly engineered smart systems may not be adopted if public trust is not secured. As autonomy moves into the sphere of residential and commercial water systems, it must be accompanied by a deep appreciation of how residents, decision makers, and regulators perceive its benefits and risks.

The vast majority of water utilities and districts in the United States are publicly owned. An understandable level of risk aversion has developed over the decades since even small changes in operations can have large implications. The close connection between water and public safety, health, and other local priorities means that decisions are not always based on economics or efficiency. Engineers' limited understanding of the perceptions of decision makers and the public regarding smart water technologies may be the biggest barrier to adoption.

The engineering disciplines must begin engaging with the social and economic sciences to develop a more holistic appreciation of the opportunities afforded by modern technologies for water management.

Workforce Development

Growing interest in smart cities promises exciting opportunities for a new workforce of engineers and technicians who will respond to a multitude of cross-disciplinary challenges. In the context of water, this will require educating a new generation of students whose knowledge of water systems will be combined with mastery of other disciplines, such as computer science and electrical engineering. Novel graduate and undergraduate educational initiatives are forming to fill this need, such as the University of Michigan's Intelligent Systems graduate program in the Department of Civil and Environmental Engineering.

This new generation of students must be embraced by an industry willing to value skills that have not traditionally been part of its core. The economic and efficiency gains resulting from smart water systems will need to be translated into commensurate salaries for appropriately educated, technologically savvy engineers, who may otherwise be lured into much higher-paying tech careers.

Economic and efficiency gains from smart water systems will need to translate to commensurate salaries for appropriately educated, tech-savvy engineers.

Similarly, there is an opportunity to begin training a new workforce of smart water technicians and maintenance experts, who will not need college degrees to benefit from a tech career. This new workforce of engineering technicians may also help to advance equity and inclusion (Kuehn 2017).

Conclusions

Smart and autonomous water systems are well within reach, driven by pioneering efforts in a growing number of US communities and utilities. Early implementation of these systems indicates that, before resorting to costly new construction, it may already be possible to use existing infrastructure much more effectively.

Lessons learned by early adopters are being shared through new cross-sector groups and consortiums of regulators, companies, utilities, and academics, such as the international Smart Water Networks Forum (SWAN), and initiatives of established water organizations, such as ASCE's Environmental Water Resources Institute (EWRI) and the Water Environment Foundation (WEF). In addition, the National Science Foundation, through its new Smart and Connected Communities program, is funding efforts to encourage fundamental discovery and meaningful engagement with cities and communities; one recently funded project will work across a number of US cities to investigate and overcome major barriers to adoption of smart stormwater systems (NSF 2017).

Anyone interested in efforts to enhance water management is encouraged to engage with these groups or to learn and contribute through open source efforts such as WaterAnalytics.org and Open-Storm.org.

References

ASCE [American Society of Civil Engineers]. 2017. American Infrastructure Report Card. Available at www.infrastructurereportcard.org/.

Barco J, Hogue TS, Curto V, Rademacher L. 2008. Linking hydrology and stream geochemistry in urban fringe watersheds. Journal of Hydrology 360(1–4):31–47.

Bartos M, Wong B, Kerkez B. 2018. Open Storm: A complete framework for sensing and control of urban watersheds. Environmental Science: Water Research and Technology. doi:10.1039/C7EW00374A.

Carey RO, Wollheim WM, Mulukutla GK, Mineau MM. 2014. Characterizing storm-event nitrate fluxes in a fifth order suburbanizing watershed using in situ sensors. Environmental Science and Technology 48(14):7756–7765.

Cembrano G, Wells G, Quevedo J, Pérez R, Argelaguet R. 2000. Optimal control of a water distribution network in a supervisory control system. Control Engineering Practice 8(10):1177–1188.

Doughton S. 2015. Toxic runoff kills adult Coho salmon, study finds. Seattle Times, October 8.

Evans MA. 2015. The sewage crisis in America: Flushing the toilet has never been riskier. The Atlantic, September 17.

Finkebine JK, Atwater JW, Mavnic DS. 2000. Stream health after urbanization. Journal of the American Water Resources Association 36(5):1149–1160.

Hunter C. 2016. 6 dead in Texas floods, and more rain is coming. New York Times, May 30.

Kea K, Dymond R, Campbell W. 2016. An analysis of patterns and trends in United States stormwater utility systems. Journal of the American Water Resources Association 52(6):1433–1449.

Kerkez B, Glaser SD, Bales RC, Meadows MW. 2012. Design and performance of a wireless sensor network for catchment-scale snow and soil moisture measurements. Water Resources Research 48(9).

Kerkez B, Gruden C, Lewis M, Montestruque L, Quigley M, Wong B, Bedig A, Kertesz R, Braun T, Cadwalader O, and 2 others. 2016. Smarter stormwater systems. Environmental Science and Technology 50(14):7267–7273.

Kuehn D. 2017. Analyzing the engineering technician and technologist workforce: Data coverage and gaps. The Bridge 47(2):11–18.

Marinaki M, Papageorgiou M. 2005. Optimal Real-Time Control of Sewer Networks. London: Springer Science & Business Media.

Monte M. 2015. DWP to unveil plan to capture storm runoff. LA Times, June 25.

Mullapudi A, Wong BP, Kerkez B. 2017. Emerging investigators series: Building a theory for smart stormwater systems. Environmental Science: Water Research and Technology 3(1):66–77.

NAE [National Academy of Engineering]. 2008. NAE Grand Challenges for Engineering. Washington: National Academy Press.

NSF [National Science Foundation]. 2017. NSF announces $19.5M in awards to support fundamental research to advance the nation's local cities and communities. Alexandria VA.

Ostfeld A, Salomons E. 2004. Optimal layout of early warning detection stations for water distribution systems security. Journal of Water Resources Planning and Management 130(5):377–385.

Rheinheimer DE, Bales RC, Oroza CA, Lund JR, Viers JH. 2016. Valuing year-to-go hydrologic forecast improvements for a peaking hydropower system in the Sierra Nevada. Water Resources Research 52(5):3815–3828.

Stoianov I, Nachman L, Madden S, Tokmouline T, Csail M. 2007. PIPENET: A wireless sensor network for pipeline monitoring. IEE 6th International Symposium on Information Processing in Sensor Networks, April 25–27, Cambridge MA, pp. 264–273.

Storey MV, van der Gaag B, Burns BP. 2011. Advances in online drinking water quality monitoring and early warning systems. Water Research 45(2):741–747.

US EPA [Environmental Protection Agency]. 2013. Water Audits and Water Loss Control for Public Water Systems. Washington.

US EPA. 2017. Energy Efficiency for Water Utilities. Washington. Available at https://www.epa.gov/sustainable-water-infrastructure/energy-efficiency-water-utilities.

Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich A, Green P, Glidden S, Bunn SE, Sullivan CA, Reidy Liermann C, Davies PM. 2010. Global threats to human water security and river biodiversity. Nature 467(7315):555–561.

Wang L, Lyons J, Kanehl P, Bannerman R. 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environmental Management 28(2):255–266.

Wardlaw R, Sharif M. 1999. Evaluation of genetic algorithms for optimal reservoir system operation. Journal of Water Resources Planning and Management 125(1):25–33.

Whittle AJ, Girod L, Preis A, Allen M, Lim HB, Iqbal M, Srirangarajan S, Fu C, Wong KJ, Goldsmith D. 2010. WaterWiSe@SG: A testbed for continuous monitoring of the water distribution system in Singapore. In: Water Distribution Systems Analysis 2010, eds. Lansey KE, Choi CY, Ostfeld A, Pepper IL, pp. 1362–1378. Reston VA: ASCE. Available at http://cedb.asce.org/cgi/WWWdisplay.cgi?284708.

Wines M. 2014. Behind Toledo's water crisis, a long-troubled Lake Erie. New York Times, August 5.

The BRIDGE

(USPS 551-240)

National Academy of Engineering
2101 Constitution Avenue NW
Washington, DC 20418

Periodicals
Postage
Paid

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR022657

Real-Time Control of Urban Headwater Catchments Through Linear Feedback: Performance, Analysis, and Site Selection

B. P. Wong[1] and B. Kerkez[1]

[1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA

Key Points:
- A linear feedback controller is formulated for the control of urban catchments (1-5 km^2) using physical catchment properties
- Simulations show strong flow control performance when compared to uncontrolled catchment
- Desired hydraulic outcomes can be achieved by retrofitting a small number of control sites

Supporting Information:
- Supporting Information S1
- Figure S1
- Figure S2
- Figure S3

Correspondence to:
B. P. Wong,
bpwong@umich.edu

Citation:
Wong, B. P., & Kerkez, B. (2018). Real-time control of urban headwater catchments through linear feedback: Performance, analysis, and site selection. *Water Resources Research*, *54*, 7309–7330. https://doi.org/10.1029/2018WR022657

Received 29 JAN 2018
Accepted 29 AUG 2018
Accepted article online 4 SEP 2018
Published online 4 OCT 2018

Abstract The real-time control of urban watersheds is now being enabled by a new generation of "smart" and connected technologies. By retrofitting stormwater systems with sensors and valves, it becomes possible to adapt entire watersheds dynamically to individual storms. A catchment-scale control algorithm is introduced, which abstracts an urban watershed as a linear integrator delay dynamical system, parameterizes it using physical watershed characteristics, and then controls network flows using a *Linear Quadratic Regulator*. The approach is simulated on a 4-km^2 urban headwater catchment in Ann Arbor, Michigan, demonstrating the gains of a stormwater system that can adaptively balance between flood mitigation and flow reduction. We introduce an equivalence analysis and illustrate the performance of the controlled watershed across large events (30-year storms) to show the uncontrolled *passive* watershed can only match it during smaller events (10-year storm). For these smaller events, the storage volume of the controlled storage nodes (ponds, basins, and wetlands) could be reduced as much as 50% and still achieve the same performance of the controlled watershed. A controller placement analysis is also carried out, whereby all possible combinations of controlled sites are simulated across a wide spectrum of design storms. We show that the control of every storage node may not be needed in a watershed, but rather that in our case study a small subset (30%) of the overall watershed can be controlled in coordination to achieve outcomes that match a fully controlled system, even when tested across a long-term rainfall record and under noisy sensor measurements.

Plain Language Summary Internet-connected sensors are changing how we study and manage water systems. By adding valves, gates, and pumps to stormwater systems, it will be possible to adaptively control urban watershed in response to individual storms. Without requiring new and expensive construction, this will allow existing infrastructure to be used more effectively. In this paper, we introduce a control algorithm that can be used to control urban watersheds. We also investigate how many control valves are needed to retrofit an urban watershed and where, which may serve as tools for city managers to reduce flooding and manage flows.

1. Introduction

Population pressures continue to drive land use changes, often resulting in more paved and impermeable urban landscapes. Stormwater runoff has become more flashy and polluted, leading to flooding, erosion, and ecosystem impairments (Boyer & Kieser, 2012). Often referred to as the *urban stream syndrome* (Walsh et al., 2005), this collection of challenges is compounded by a changing climate that drives storms of increasing intensity and frequency (Alexander et al., 2006; Cheng & AghaKouchak, 2014). At a time of declining infrastructure funding (Grigg, 2015; Pagano & Perry, 2008), pressure is mounting on urban watershed managers to do more with less.

Traditionally, flooding and stream erosion have been mitigated through the expansion of constructed stormwater infrastructure, which conveys runoff from buildings and roads through a complex system of belowground and aboveground infrastructure, such as pipes, detention basins, and constructed wetlands. Most recently, *green infrastructure* has risen to prominence in the form of many smaller and distributed assets, such as bioswales, rain gardens, and green roofs (Benedict et al., 2012; Fletcher et al., 2014). Watershed managers thus have a large portfolio of stormwater options, ranging from large centralized assets to smaller distributed solutions, most of which are very expensive. Cost are dependent on location but can vary from hundreds of thousands of dollars per city block for green infrastructure, millions of dollars for residential basins, and upward of billions of dollars for large tunnels (Keeley et al., 2013; Schütze et al., 2002). These solutions are oftentimes designed to meet specific performance criteria; however, once constructed, passive stormwater systems become costly and difficult to adapt to changes in land use and weather patterns. While many of these solutions have shown to have very positive outcomes (Askarizadeh et al., 2015; Rosenberg et al., 2010; Strecker et al., 2001), recent studies have also shown that aggressive adoption via many large or distributed stormwater assets can actually lead to worse watershed outcomes if individual elements are not tuned to system-level outcomes (Goff & Gentry, 2006; Petrucci et al., 2013). In such instances, while a neighborhood-scale fix may attenuate local peak flows and thus give the impression of a favorable outcome, its hydrograph peak may be pushed to align more closely with a downstream peak, thus contributing more to downstream flooding and erosion than before the measure was put in place (Emerson et al., 2005). As such, there is an urgent need to find new adaptive solutions that are *aware* of the larger watershed.

"Smart" stormwater systems have recently been proposed to achieve real-time adaptation and system-level control (Eggimann et al., 2017; Kerkez et al., 2016). In lieu of new construction, this paradigm proposes to use many distributed, low-cost sensors and controllers (valves, gates, pumps, etc.) to coordinate flows across the scales of entire watersheds. While the use of controlled assets—such as pumps—is not necessarily new, they are rarely used since many stormwater systems simply rely on gravity for conveyance. If deployed, they are often operated with simple control logic (e.g., level control). More importantly, their operation is not coordinated across the broader watershed, which reduces their utility as an adaptive tool for watershed-scale management. The ability to connect such assets to the Internet using new and inexpensive wireless technologies allows for their joint coordinated operation, effectively allowing an entire watershed to be reconfigured on a storm-by-storm basis. For systems that do not have any controlled assets, cheap and open source wireless valves can now readily be used to retrofit basins, ponds, wetlands, and green infrastructure (Bartos et al., 2017).

Sensors and controllers have been applied across various *smart* water systems, such as drinking water distribution systems, wastewater treatment and conveyance, and irrigation (Lemos & Pinto, 2012; Montestruque & Lemmon, 2015; Ocampo-Martinez et al., 2013; Van Overloop et al., 2010). While it would seem intuitive that such benefits could be translated to stormwater systems, it is still unclear how. Unlike strictly man-made system, stormwater systems comprise a mix of man-made and natural components, leaving much research to be conducted to determine *how* they should be controlled safely and reliably across the scale of entire watersheds. This requires an interdisciplinary knowledge of domains spanning hydrology, infrastructure, data science, and control theory.

In this paper, we take a step toward the coordinated, real-time control of urban watersheds (Figure 1, Top) by asking the question: Where should urban catchments be retrofitted for real-time control and what performance gains can be achieved compared to *passive* alternatives? More specifically, we seek to investigate how control algorithms can be synthesized to reduce flooding and erosion in urban environments. The fundamental contributions of this paper are the following:

1. A feedback control methodology, which uses physical parameters of a watershed to mathematically formulate a dynamical, integrator-delay linear system and then uses a linear-quadratic (LQ) regulator (LQR) controller to coordinate the operation of valves and gates in the watershed.
2. A simulation-based approach to help identify how many distributed control valves are needed and where they should be placed to achieve the best real-time control outcomes, focusing specifically on reducing flooding and erosion.
3. A holistic equivalence analysis, which compares the real-time controlled system to passive solutions across many storms of varying intensities and durations.

Given that smart stormwater systems have yet to be constructed at large scales, this analysis will be carried out in simulation, which will allow for a variety of scenarios to be evaluated before results can be used to build real-world control networks. Furthermore, the analysis will focus on the scale of urban headwater catchments (1–5 km^2), which will serve as building blocks to inform the control of larger watersheds in the future.

2. Background

Wireless technologies are becoming increasingly more affordable and accessible (Gubbi et al., 2013). When coupled with measurements, such as low-cost water depth and in situ water quality sensors (Hart &

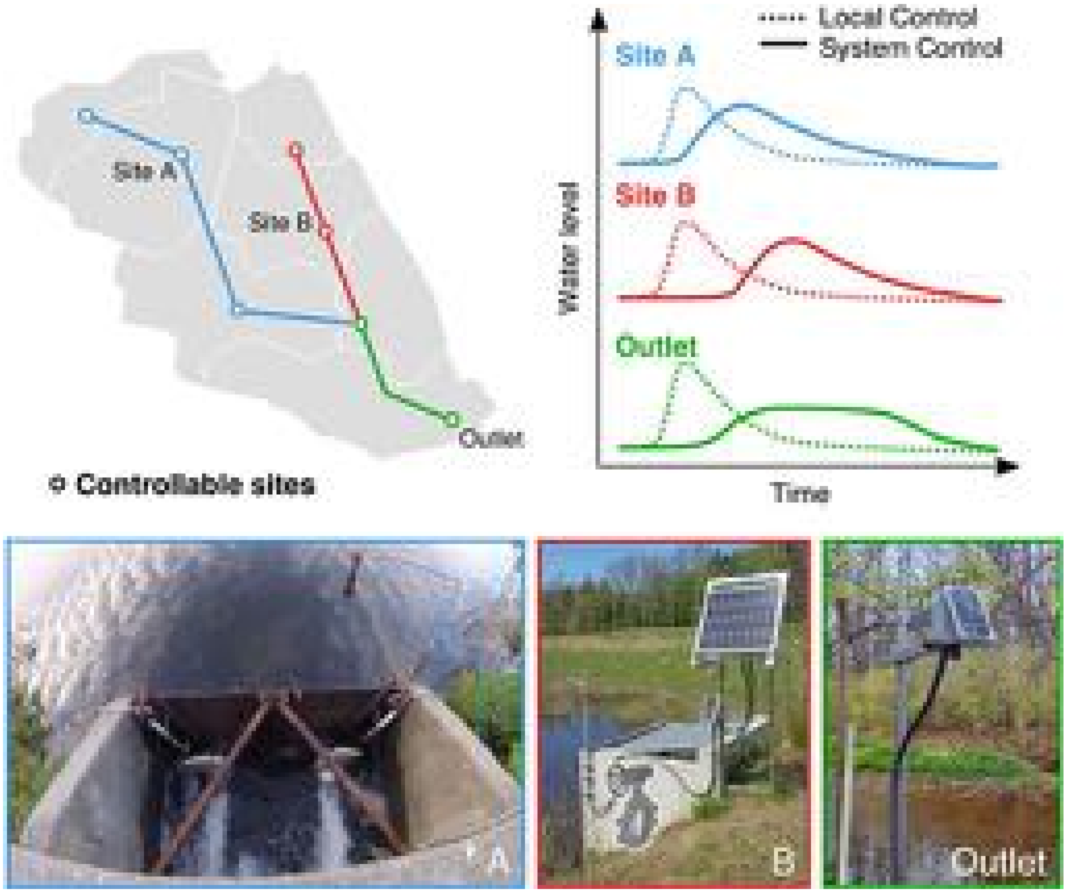


Figure 1. Conceptual representation of the study watershed, with and without controlled flows (top). Photos of sites instrumented with control valves (a and b) and sensors (a, b, and outlet) are shown on the bottom row.

Martinez, 2006; Wong & Kerkez, 2014), field-deployed devices are able to stream unprecedented amounts of real-time measurements about the health and performance of large watersheds (Bartos et al., 2017; Wong & Kerkez, 2016b). The resulting real-time data streams become particularly important when used in a bidirectional fashion. Namely, rather than simply receiving measurements, commands may be transmitted back to watersheds to change flows and hydrologic behavior. A simple example involves the addition of an inexpensive control valve (Figures 1a and 1b) to the outlet of a stormwater basin, such as those currently being retrofitted by the authors in the Midwestern United States (Bartos et al., 2017; Wong & Kerkez, 2016b). Compared to passive solutions, where the outflows are determined by the fixed geometry of an outlet, real-time control provides the ability to actively modulate runoff and adapt site behavior based on real-time hydrologic states and rainfall forecasts (Marsalek, 2005).

Even a single remotely controlled valve can yield immediate benefits (Bryant & Wadzuk, 2017; Gaborit et al., 2013; Muschalla et al., 2014). For example, a valve can be used to extend hydraulic retention time, thus promoting the capture of sediment-bound pollutants (Carpenter et al., 2014; Gaborit et al., 2016; Gilpin & Barrett, 2014). When a storm is forecasted or detected, water can then be released to create additional storage capacity. By extension, modulation of flows (hydrograph shaping) from a site could reduce erosion by ensuring that outflows do not exceed critical levels. Real-time control can thus adaptively balance water quality and flooding benefits, which is difficult to accomplish using passive solutions. Given the recent advent of these technologies, research studies addressing the benefits of real-time stormwater control have primarily focused on benefits that can be obtained at the scale of individual sites (Carpenter et al., 2014; Gaborit et al., 2013; Middleton & Barrett, 2008). Many of these studies rely on rule-based controls that are tailored and parameterized for each individual site (Gaborit et al., 2013). Thus, it becomes difficult to posit how these lessons may scale to the coordinated control of a watershed with many control sites.

Presently, the synthesis of real-time control strategies for separated stormwater systems remains largely unexplored. Guidance may however be taken from work on other water systems, including reservoir operations, open-channel irrigation systems, water distribution systems, and sewer systems (Hashemy et al., 2013; Montestruque & Lemmon, 2015; van Overloop, 2006; van Overloop et al., 2008). While these prior approaches show great promise, they do not explicitly account for the types of time scales and feedbacks inherent in urban watersheds. These time-dependent complexities, which span hydrology, rapidly changing weather, flashy flows, hydraulics, and even water quality, pose a pressing challenge to even the latest studies on networked-system control (Liu et al., 2011). In their work in controlling canal systems, van Overloop (2006) used a time-varying Integrator Delay model as the model for a model predictive controller that controlled the system while ensuring stable flows. While these results are promising, the studied canal systems were not subjected to rapid storm impulses, which makes it unclear how stable their operation would be in flashy urban watershed.

While control of sewer systems has received attention, it has primarily focused on the control of man-made infrastructure, with limited attention to hydrologic features. For example, the successful control of an 11-reservoir sewer network considered flows but neglected backwater effects, which occur as water levels rise from closing valves (Marinaki & Papageorgiou, 2003). Optimization-based control approaches, such as population dynamics (Barreiro-Gomez et al., 2015; Ramírez-Llanos & Quijano, 2010), game theoretic, and agent-based control (Barreiro-Gomez et al., 2017), can bypass the use of a physical hydraulic-hydrologic model altogether and instead rely on a mass-balance relationship to stabilize flows throughout the network. As a result, optimization-based controllers require complex parameterizations that are dependent on the availability and accuracy of data, which does not always exist (García et al., 2015). When data are limited, heuristics and expert knowledge about general system behavior can be leveraged by alternatives like fuzzy logic controllers (Aulinas et al., 2011; Klepiszewski & Schmitt, 2002; Marinaki & Papageorgiou, 2003). However, there is no general convention for these approaches, and they may require a large number of rules or scenarios (Gaborit et al., 2013; Muschalla et al., 2014).

Urban control for one site, no less an entire stormwater system, is still an emerging area of research. There is an opportunity, however, to build upon prior discoveries in other water domains to synthesize a system-level controller that factors in hydrologic processes but can also be parameterized by easily measurable and readily available sensor data. Here we present such a formulation, which is based on a linear integrator delay model, which can be parameterized by watershed features, such as storage curves and the distance between storage nodes (SNs). The system is then controlled by a linear quadratic regulator, which only requires water levels as measurements. We apply this approach to a simulated case study to an urban headwater catchment in the Midwestern United States and evaluate the performance of the approach under flooding and flow objectives.

3. Methods

3.1. Approach

When retrofitting urban watersheds for real-time control, a choice must be made in regard to the spatial scales at which these technologies will first be implemented and analyzed. We contend that the analysis of real-time control strategies should begin at the scale of urban headwater catchments. These subcatchments are as large as 5 km^2 (2 mi^2) and can be found in most urban and suburban communities, small and large (Emerson et al., 2005; J. G. Lee et al., 2012; Zhen et al., 2004). Overall, the choice to focus on this scale is motivated by a number of fundamental and practical factors. Fundamentally, the scalability of real-time watershed control requires smaller-scale systems to be analyzed and understood first (J. H. Lee & Bang, 2000). If feasible at these scales, the control of smaller catchments will ultimately underpin the control of larger watersheds.

Practically, it is unlikely that entire cities will be retrofitted with control valves all at once. Rather, valves will be added one-by-one or as controlled clusters. In the United States, decisions to build or upgrade stormwater infrastructure are often driven by new residential or commercial development projects, which impact flows at the scale of local pipes and stream networks (Grigg, 2012; Kessler, 2011). Given the recent emphasis on distributed stormwater management, these measures often include ponds, basins, and wetlands at commercial complexes, subdivisions, neighborhoods, and precincts. Urban flash flooding occurs at the scale of local road networks, which suggests that control strategies should operate to prevent flooding even as far upstream as first-order catchments. In fact, the U.S. Federal Emergency Management Agency provides flood advisory and insurance information at scale of 5 km^2 (2 mi^2) subwatersheds (https://msc.fema.gov/portal/search), which makes them of particular interest for analysis. Most existing radar and gage rainfall products are offered at 1–5-km^2 resolution as well, which is relevant if rainfall forecasts are to be integrated with real-time control. As such, both fundamental and practical considerations suggest that the scale of headwater catchments (1–5 km^2) provides a good starting point to answer the questions posed in this paper. Future studies can then analyze how the control of larger watersheds can be achieved through lessons learned at the catchment scale.

3.2. Dynamical System Representation

Most modern physically based models of urban watersheds, such as Environmental Protection Agency's *Stormwater Management Model* (SWMM; Gironás et al., 2010), are based on a coupled hydrologic-hydraulic approach, where hydrologic dynamics, such as runoff and infiltration, are represented via physical or empirical submodels. Flows are subsequently routed using a hydraulic engine, typically based on nonlinear *Saint-Venant* equations for shallow water flow (Rossman, 2010). Given the high degree of detail, complexity, and nonlinearities inherent in these models, the application of formal control and optimization approaches becomes intractable. Fortunately, for many complex control systems, such as those used on autopilots and factory processes, perfect models are not necessary to achieve desirable control outcomes (Schütze et al., 2002). Rather, for model-based controllers like linear quadratic regulators and model predictive controllers, a control model that approximates the dynamics of the underlying system is often sufficient as long it satisfies stability criteria since the actual control actions will often steer the system back into domains where the approximations hold true (Francis & Wonham, 1976). In feedback control, this is often accomplished by linearizing the system dynamics around desired setpoints (e.g., flows and flood stages), after which modern control techniques can be applied. For the specific control of water flows in pipes and canals, examples of approximated models have included the integrator delay (Schuurmans et al., 1995), integrator delay zero models (Litrico & Fromion, 2009), reduced Saint-Venant (Xu et al., 2011), Muskingum (Gill, 1978), and linear tank models (Ocampo-Martinez et al., 2013).

For our approach, the control model is based on a state-space representation of the hydraulic dynamics as an *integrator delay model* (Schuurmans et al., 1995). In recent studies, this representation has been used for the control of water levels in irrigation canals that are connected in series (Xu et al., 2011). However, the use of this formulation for urban watersheds presents additional complexities, making it unclear how well it will work in the control for stormwater systems, if at all. These include the need to accommodate hydrologic effects (runoff, antecedent moisture, etc.) and rainfall, as well as complex and interconnected infrastructure topologies (parallel SNs or tree-like networks). Our choice to adopt this approach is based on our expectation that it will sufficiently capture hydrologic and shallow-water flow dynamics to enable feedback control. Most importantly, however, the matrix-based representation can be fully parameterized from the watershed physical features, including the distance between SNs and their storage curves. This formulation also only relies on water level measurements for implementation, which our prior studies have shown to be very ubiquitous and cost effective (Bartos et al., 2017; Wong & Kerkez, 2014).

3.2.1. State-Space Representation of an Urban Watershed

Our integrator delay model conceptualizes an urban watershed as a system of interconnected SNs (Figure 2a) as a linearized state-space representation (1). The full time-varying state-space representation is given by

$$\begin{aligned} x(k+1) &= \mathbf{A}(k)x(k) + \mathbf{B_u}(k)u(k) + \mathbf{B_d}d(k) \\ y(k) &= \mathbf{C}x(k), \end{aligned} \tag{1}$$

where the state vector $x\,(k)$ is composed of the heights $h_i\,(k)$ and outflows $Q_i\,(k)$ of all SNs in the system at the kth time step; the control vector $u\,(k)$ contains the controlled outflow $Q_{\text{control},\,i}\,(k)$ for each controllable orifice and valve in the system; and the disturbance vector $d\,(k)$ is a vector of the total runoff $Q_{\text{runoff},\,i}\,(k)$ from all local subcatchments that flows into the ith SN. The output vector $y(k)$ is composed of the height of each SN. In this study, all necessary states are assumed observable (measured by sensors), allowing us to focus solely on controlling the evolution of the state vector $x\,(k)$.

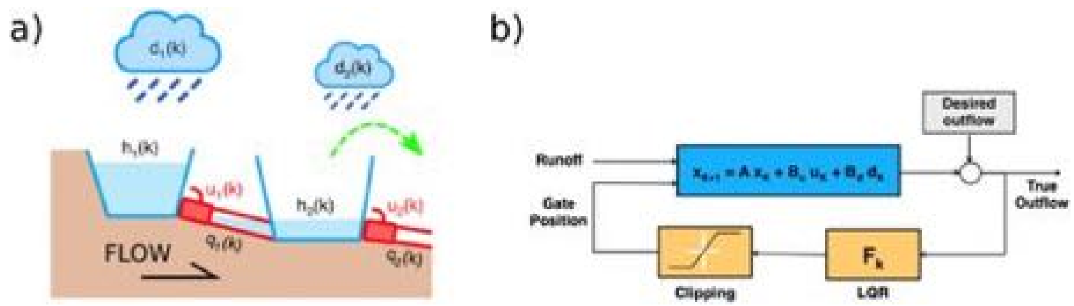


Figure 2. Graphical representation of (a) an integrator-delay model and (b) the block diagram of the feedback controller.

The state matrix **A**(k) relates the heights and flows from the current time step to the next; the control matrix $\boldsymbol{B}_u(k)$ links the control outflow to its associated SN; and the disturbance matrix $\boldsymbol{B}_d$ routes the rainfall-generated runoff from a given subcatchment to its associated SN. The output matrix $\boldsymbol{C}$ relates the state vector to the output vector. The matrices in (1) are composed of submatrices that are dependent on physical properties of the system, where each matrix is constructed from integrator and delay block components that correspond to each individual SN and channel, respectively.

The dynamics for the integrator delay model can be decomposed into two main components—integrators and delays.

3.2.2. The Integrator Component

$x_{\text{integrator } i}(k+1) = [h_i(k+1)]$, where

$$h_i(k+1) = h_i(k) \tag{2}$$

$$+\left[\frac{-T}{A_{S,i}(k)}\right]\mathbf{1}_{\{\text{orifices}\}}(i)Q_{\text{control}}(k) + \left[\frac{T}{A_{s,i}(k)}\right]\mathbf{1}_{\{\text{subcatchments}\}}(i)Q_{\text{runoff}}(k)$$
$$= \boldsymbol{A}_{\text{integrator } i}(k)h_i(k) + \boldsymbol{B}_{u,\text{ integrator } i}(k)Q_{\text{control}}(k) + \boldsymbol{B}_{d,\text{ integrator } i}\ Q_{\text{runoff}}(k). \tag{3}$$

The *integrator* component 2 models the change in height of the ith SN as a function of the measured height $h_i(k)$, inflows $q_i(k)$ from one or multiple upstream nodes, runoff inflows from local catchment $d_i(k)$, and the controlled outflows u_i (k). Each integrator component also includes the sampling period T and the cross-sectional area of the ith SN at the current time step $A_{s,\ i}(k)$, which is obtained from the storage curve using the measured height $h_i(k)$. $\boldsymbol{A}_{\text{integrator } i}(k)$ is a submatrix of **A**(k), while $\boldsymbol{B}_{u,\text{ integrator } i}(k)$ and $\boldsymbol{B}_{d,\text{ integrator } i}$ are submatrices of $\boldsymbol{B}_u(k)$ and $\boldsymbol{B}_d$, respectively.

3.2.3. The Delay Component

$$x_{\text{delay } i}(k+1) = \begin{bmatrix} Q_i(k+1) \\ Q_i(k) \\ Q_i(k-1) \\ \vdots \\ Q_i(k-n) \end{bmatrix}, \text{ where}$$

$$\begin{bmatrix} Q_i(k+1) \\ Q_i(k) \\ Q_i(k-1) \\ \vdots \\ Q_i(k-n) \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & & & 0 \\ 0 & 1 & & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} Q_i(k) \\ Q_i(k-1) \\ Q_i(k-2) \\ \vdots \\ Q_i(k-n-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \mathbf{1}_{\{\text{orifices}\}}(i)\, Q_{\text{control } i}(k) \tag{4}$$

$$= \boldsymbol{A}_{\text{delay } i}(k)\begin{bmatrix} Q_i(k) \\ Q_i(k-1) \\ Q_i(k-2) \\ \vdots \\ Q_i(k-n-1) \end{bmatrix} + \boldsymbol{B}_{u,\text{delay } i}(k)\, Q_{\text{control } i}(k). \tag{5}$$

The *delay* component 4 represents the travel time of water from one SN through a channel to the next node. The number of delay terms n is determined by the distance between SNs, the slope, and the sampling period T. For each delay component, $\boldsymbol{B}_{u,\text{ delay } i}(k)$ routes the control outflow from a given orifice to the ith channel, while $\boldsymbol{A}_{\text{delay } i}(k)$ delays this control outflow in time until it reaches the downstream SN. $\boldsymbol{A}_{\text{delay } i}(k)$ is a submatrix of $\mathbf{A}(k)$, while $\boldsymbol{B}_{u,\text{ delay } i}(k)$ is a submatrix of $\boldsymbol{B}_u(k)$. Together with the integrator components, these submatrices combine to form the linearized state-space representation of an urban watershed.

3.2.4. The Indicator Function

$$\mathbf{1}_{\{\xi\}}(j) = [x_1 \;\; x_2 \;\; \cdots \;\; x_n], \text{where } x_i = \begin{cases} 1, & \text{Element } \xi_i \text{ flows into channel } j \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

An indicator function 6 is used in the formulation of each integrator component 2 to indicate which control inputs and runoff disturbances are associated with the ith SN. An indicator function is also used to in the formulation of each delay component 4 to indicate which control outflows are routed into the ith channel.

3.3. Constructing the System Matrices From a Physical Representation

The link component 7 is a block matrix with a scalar value in the upper-right corner that fills in the nondiagonal submatrices of $\boldsymbol{A}(k)$ to effectively link the dynamics of the integrator and delay components.

$$\boldsymbol{A}_{\text{link } i,j}(k) = \begin{bmatrix} 0 & \cdots & a(k) \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}, \text{where } a(k) = \begin{cases} 1, & \boldsymbol{A}_i, \boldsymbol{A}_j \text{ are delay blocks} \\ \dfrac{T}{A_{s,i}(k)}, & \boldsymbol{A}_i \text{ is an integrator block} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

To begin constructing a state-space representation from a physically based model or real-world system, the integrator 2, delay 4, and link components 7 can be combined to form the state matrix $\boldsymbol{A}(k)$:

$$\boldsymbol{A}(k) = \begin{bmatrix} \boldsymbol{A}_{\text{integrator } 1}(k) & \boldsymbol{A}_{\text{link } 12}(k) & \cdots & & & & \\ \boldsymbol{A}_{\text{link } 21}(k) & \ddots & & & & & \\ \vdots & & \boldsymbol{A}_{\text{integrator } m}(k) & & & & \\ & & & \boldsymbol{A}_{\text{delay } 1}(k) & & \\ & & & & \ddots & \\ & & & & & \boldsymbol{A}_{\text{delay } n}(k) \end{bmatrix}. \tag{8}$$

Here T is the sampling period, and $A_{s,\, i}(k)$ is the cross-sectional area, which is obtained from the storage curve for the ith node using the measured height $h_i\,(k)$ at the current time step.

The control matrix $\boldsymbol{B}_u(k)$ is similarly assembled by iterating through all SNs and vertically concatenating the integrator components $\boldsymbol{B}_{u,\text{ integrator } i}(k)$ followed by concatenating the delay components $\boldsymbol{B}_{\text{u, delay } i}(k)$. The disturbance matrix $\boldsymbol{B}_d$ is assembled by iterating through all SNs and vertically concatenating the integrator components $\boldsymbol{B}_{d,\text{ integrator } i}(k)$ and padding the remaining rows with zeros for the delay components:

$$\boldsymbol{B}_u(k) = \begin{bmatrix} \boldsymbol{B}_{u,\text{integrator } 1}(k) \\ \vdots \\ \boldsymbol{B}_{u,\text{integrator } m}(k) \\ \boldsymbol{B}_{u,\text{delay } i}(k) \\ \vdots \\ \boldsymbol{B}_{u,\text{delay } n}(k) \end{bmatrix}. \tag{9}$$

$$\boldsymbol{B}_d(k) = \begin{bmatrix} \boldsymbol{B}_{d,\text{integrator } 1}(k) \\ \vdots \\ \boldsymbol{B}_{d,\text{integrator } m}(k) \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{10}$$

3.4. Estimating Outflow for Partial Valve Opening

To model the outflow for each controlled node $Q_{\text{control } i}$ (k) from the control vector u (k), the effective cross section of each valve (how far the valve is opened), $A_{\text{valve}}(k)$, is obtained by inverting the equation for flow through a free-flowing undershot gate 11 (Schuurmans, 1997).

$$A_{\text{valve } i}(k) = \frac{Q_{\text{control } i}(k)}{C\mu\sqrt{2gh_i(k)}}. \tag{11}$$

Here $Q_{\text{control } i}$ (k) is the desired controlled outflow; C is a calibration coefficient; μ is the contraction coefficient; g is the gravitational acceleration; and h_i (k) is the water level of the SN. To approximate the outflow for an uncontrolled orifice, the equation for flow through a free-flowing undershot gate 11 is linearized about the height of the SN 12. This is necessary to linearly model the outflow from uncontrolled SNs to the connected links at each time step.

$$Q(k) = C\mu h_i(k)\cdot \text{Width}_{\text{valve}}\sqrt{2g\ (h(k) - \mu\cdot h_i(k))}, \tag{12}$$

$$Q(k+1) - Q(k) \cong \lambda(k)\cdot(h_i(k) - h_0(k)), \text{where} \tag{13a}$$

$$\lambda(k) = \frac{gC\mu h_i(k)\cdot\text{Width}_{\text{valve}}}{\sqrt{2g(h_0(k) - \mu\cdot h_i(k))}}, \tag{13b}$$

and $h_0(k)$ is the linearization point. Changes in valve position are assumed to be small, so linearization terms about $h_i(k)$ are neglected.

3.5. Control Algorithm

Once a linear representation of the catchment is formulated, our approach uses an LQR to set the outflows from each controllable SN for any given time step (Figure 2b). The approach takes the desired system states (e.g., heights at each SN) as an input, compares these to the corresponding states throughout the actual system, and then calculates the necessary outflows at each controlled node to push the system toward these desired setpoints.

LQ control is a matrix-based, closed-loop feedback control method that incorporates open-loop dynamics to achieve desired setpoints (Malaterre & Baume, 1998). LQR is suitable for real-time control since the matrix computations are relatively fast, making it possible to run even on modern microcontrollers. LQR controls a linear dynamical system 1 by minimizing a quadratic cost function 14 using a control input u (k) based on the latest sensor measurements x (k) (Dorato et al., 1995).

$$J = x^T(N)\mathbf{Q}x(N) + \sum_{k=0}^{N-1}\left(\rho \cdot x^T(k)\mathbf{Q}x(k) + u^T(k)\mathbf{R}u(k)\right). \tag{14}$$

Cost matrices, **Q** and **R**, are tuned to generate a cost function J that produces results aligned with the desired setpoints. The matrix **Q** represents the cost of all measured levels throughout the network that deviate from the desired setpoints, while **R** signifies the collective cost of all the control outflows, specifically their magnitude. The parameter ρ shifts the relative weight of the cost between states (e.g., measured vs. desired levels) and control actions (e.g., outflows). Over the period which the system dynamics are constant, the performance cost of LQ control for linear, time-invariant systems, if it exists, is minimal, bounded, and optimal (Lin, 2007; Sinha, 2007).

Given the cost matrices and assuming the state and input matrices ($\mathbf{A}(k)$, $\mathbf{B}_u(k)$) from the state-space representation are stabilizable, the controlled outflow for each SN is then given by the control vector:

$$u(k) = -\mathbf{K}(k)\,\mathrm{x}\,(\mathrm{k}), \tag{15}$$

where $\mathbf{K}(k)$ is the gain matrix:

$$\mathbf{K}(k) = \left(\mathbf{B}_\mathbf{u}^\mathbf{T}(k)\mathbf{P}(k)\,\mathbf{B}_\mathbf{u}(k) + \mathbf{R}\right)^{-1}\left(\mathbf{B}_\mathbf{u}^\mathbf{T}(k)\mathbf{P}(k)\,\mathbf{A}(k)\right), \tag{16}$$

and the *cost-to-go* matrix $\mathbf{P}(k)$ is the solution to the discrete time *Riccati* equation (Lancaster & Rodman, 1995):

$$\mathbf{P}(k-1) = \mathbf{A}^\mathbf{T}(k)\mathbf{P}(k)\mathbf{A}(k) - \left(\mathbf{A}^\mathbf{T}(k)\mathbf{P}\mathbf{B}_\mathbf{U}^\mathbf{T}(k)\right)\left(\mathbf{B}_\mathbf{u}^\mathbf{T}(k)\mathbf{P}(k)\mathbf{B}_\mathbf{u}(k) + \mathbf{R}\right)^{-1}\left(\mathbf{B}_\mathbf{u}^\mathbf{T}(k)\mathbf{P}(k)\mathbf{A}(k)\right) + \mathbf{Q}. \tag{17}$$

To ensure a gain matrix could be computed for every valve placement combination, the system is factored into *controllable* and *uncontrollable* components using *Kalman Decompositio*n (Dahleh et al., 2004). The transformed system is obtained by applying a linear transformation **T**:

$$\hat{x}(k) = \begin{bmatrix} \hat{x}_C(k) \\ \hat{x}_{NC}(k) \end{bmatrix} = \mathbf{T}x(k), \tag{18}$$

$$\hat{\mathbf{A}}(k) = \begin{bmatrix} \hat{\mathbf{A}}_C(k) & \hat{\mathbf{A}}_{12}(k) \\ \hat{\mathbf{A}}_{21}(k) & \hat{\mathbf{A}}_{22}(k) \end{bmatrix} = \mathbf{T}\mathbf{A}(k)\mathbf{T}^{-1}, \tag{19}$$

$$\hat{\mathbf{B}}_u(k) = \begin{bmatrix} \hat{\mathbf{B}}_{u,C}(k) \\ \mathbf{0} \end{bmatrix} = \mathbf{T}\mathbf{B}_\mathbf{u}(k), \tag{20}$$

where matrices $\widehat{\boldsymbol{A}}_\boldsymbol{C}(k)$ and $\widehat{\boldsymbol{B}}_{u,C}(k)$ are the controllable components. Applying the same transformation to the cost matrices yields

$$\hat{\mathbf{Q}} = \mathbf{T}\mathbf{Q}\mathbf{T}^{-1}, \tag{21}$$

$$\hat{\mathbf{R}} = \mathbf{R}. \tag{22}$$

Since the resulting pair $(\widehat{\boldsymbol{A}}_C(k), \widehat{\boldsymbol{B}}_{u,C}(k))$ is *stabilizable*, the gain matrix $\widehat{\boldsymbol{K}}(k)$ can be obtained for the controllable subsystem using the transformed cost matrices ((21) and (22)) and the same method (16). Finally, the control input for the controllable valves can be computed as

$$u(k) = -\hat{\mathbf{K}}(k)\hat{x}(k). \tag{23}$$

3.6. Enforcing Physical Constraints

Before applying the controller inputs, a clipping component 24 is used to enforce real-world constraints on the outflow and opening of each valve. This ensures that the outflow of a controlled SN cannot be negative or exceed physically achievable values (e.g., pressure head) given the current water level height of the SN. At each time step, once all control outflows are calculated, each control outflow is constrained to a nonnegative range limited by the maximum allowable outflow $Q^*_{i,\max}$ also by using a clipping function. The clipping function is given by the piecewise function:

$$u_i(k) = \begin{cases} Q^*_{i,\max}, & u_i(k) > Q^*_{i,\max} \\ u_i(k), & 0 \le u_i(k) \le Q^*_{i,\max} \\ 0 & u_i(k) < 0 \end{cases}. \tag{24}$$

Once the control outflows are clipped, they are transformed using 11 to determine the opening of each orifice. A second clipping component is used to constrain each orifice to a nonnegative area limited by the maximum orifice area. Once these values are determined, they are used to set the state of each valve.

4. Implementation

To promote transparency and accelerate future research of these topics, all of the models and source code from this study are available as part of an open source effort. Those interested in applying or contributing to these efforts are encouraged to join this web portal (open-storm.org).

4.1. Physical Simulations

Many studies often evaluate the performance of control algorithms on the linear models they are based on. However, if this simplified linear model does not adequately capture the physical hydraulic-hydrologic dynamics, it may give the impression that the controller performs better than it actually would in the real world. To address this concern, our approach applied the linear controller to a physically based model. In this fashion, the linearized model was used to make control decisions, while the physically based model reflected what could be expected in the real world.

Control performance was evaluated using the U.S. Environmental Protection Agency SWMM, a popular hydrologic-hydraulic computational model that has been successfully used in the planning, analysis and design of urban drainage systems (Avellaneda et al., 2017; Cantone & Schmidt, 2011; Gironás et al., 2010; Wang & Guo, 2018). SWMM numerically solves the one-dimensional Saint-Venant equations to accurately model transient surface runoff and open-channel flow. Stormwater *hydrology* is modeled based on a collection of homogeneous subcatchment areas that receive precipitation and generate runoff (Rossman & Huber, 2016), while stormwater *hydraulics* are modeled by routing runoff through a network of channels, storage units, and orifices. Although SWMM is computationally more complex than our integrator-delay model, accurately modeling the water levels in the SNs and channel flows is vital to understanding how the proposed algorithms may actually perform when subjected to real-world physics.

While SWMM provides a powerful simulation engine and rudimentary control rules (e.g., site-scale water level control), it was not designed for system-level control algorithms such as the one in this study. To that end, we implemented a customized modeling framework that uses the SWMM engine but executes the model in a stepwise fashion (Mullapudi et al., 2017; Riaño-Briceño et al., 2016). Rather than running the model for the duration of an entire storm, the model is halted every time step, after which the states can be extracted and an external logic module (in our case, an LQR controller) can be used to set the states of valves and gates across the entire system. While the routing step in our SWMM model was set to 5 s, the control input was updated every 60 simulation steps, or 5 min, to more realistically match the sampling frequency of our sensor nodes.

Since the physics engine, which is written in the *C* programming language, is implemented as a stand-alone library, the framework provides a wrapper to interface SWMM with modern and popular languages, including *Python* and *Matlab*. This allows for the seamless interaction of modern computational and control libraries with the physically based modeling of SWMM without necessarily having to implement the controller in the original SWMM model itself. More importantly, implementing this methodology does not necessarily depend on SWMM. Instead, SWMM is used as an evaluation engine for the controller. This, in fact, is much more comprehensive than most control approaches, which evaluate a control algorithm using the simplified linear model that the controller is based on. A summary of how the algorithms were implemented is provided in supporting information S1 of this paper, with pseudocode to guide replication.

Figure 3. The study catchment in Ann Arbor, MI, represented as (a) a Stormwater Management Model and (b) a network for the linearized model reflecting the relative sizes of the storage nodes, their interconnection, and subcatchments.

4.2. Case Study Area

The control approach was evaluated on a 4-km^2 catchment in Ann Arbor, Michigan (Figure 3), which is presently being retrofitted for real-time control by the authors (Figure 1, sites A and B). This particular catchment has been of interest to local officials due to stream erosion, calling for improved means to reduce flows at the outlet of the watershed (Lawson et al., 2017; Pratt, 2016). The catchment is composed of 11 storage basins, ranging in volume from 370 to 32,000 m^3. A calibrated SWMM model of the catchment was made available to the authors by city managers, thus reflecting the most up-to-date knowledge of the real system. To represent valves, each SN in the model was retrofitted with an adjustable 0.1-m^2 orifice, located at the bottom of the SN. Each orifice had a higher invert elevation than the overflow height of all downstream SNs, and all conduits between SNs were circular in geometry with variation in length from 40 to 400 m and Manning roughness coefficient of 0.01. The choice to use this modeled catchment as a case study was also motivated by its low baseflow and limited influences of groundwater effects (HRWC, 2013), which reduces one element of uncertainty when conducting the initial evaluation of the proposed control algorithm for managing stormwater runoff. The soil types are largely type C and some type D according to U.S. Geological Survey and U.S. Department of Agriculture soils data, and a Green-Ampt model was used to model soil infiltration in the subcatchments (Rawls et al., 1983). The model was last calibrated in 2015 (CDM Smith, 2015). Additional model details for the SNs and links can be found in Tables S2 and S3.

A linear control model was formulated from the SWMM model using the methods described in section 3. For the catchment in this study, $\mathbf{A}(k)$ is a 55 × 55 matrix; $\mathbf{B}_{\mathrm{u}}(k)$ is a 55 × 11 matrix; and $\mathbf{B}_{\mathrm{d}}$ is a 55 × 19 matrix. The state vector $x(k)$ is a 55 × 1 vector, and the control vector u (k) is an 11 × 1 vector. After applying the transformation matrix $\mathbf{T}$ to decompose the integrator-delay model and isolate the controllable components, $\widehat{\boldsymbol{A}}_{\boldsymbol{C}}(k)$ and $\widehat{\boldsymbol{B}}_{\boldsymbol{u},\boldsymbol{C}}(k)$, the time-varying gain matrix $\widehat{\boldsymbol{K}}(k)$ was approximated using (16). While the model was simulated at a 5-s resolution, control actions were constrained to 5 min windows to be consistent with the sensor and control networks currently being deployed by the authors (Wong & Kerkez, 2016a). All simulations were carried out using up to 100 cores in parallel on a high-performance Linux cluster at the University of Michigan.

The cluster consists of approximately 27,000 compute cores each with at least 4 GB RAM (http://arc-ts.umich.edu/flux/).

5. Performance Evaluation

Following common practice in stormwater engineering, the modeled catchment was subjected to a variety of synthetic *design storms* (Guo, 2001). To account for storms of various intensity and duration, the physically based model was simulated with rainfall from *Soil Conservation Survey* Type-II design curves (U.S. SCS, 1986), which are commonly used in the United States infrastructure design (Akitsu et al., 2011). Storms ranged from 15 min to 24 hr in duration and 5- to 200-year return period. Statistical storm data provided by NOAA Atlas 14 defined the intensity of the storms for a given storm duration and return period (Unruh & Yekta, 2013). For example, a design storm of 24-hr duration with a 10-year return period, henceforth referred to as a 10-year, 24-hr storm, has an average cumulative rainfall of 83 mm in Southeast Michigan.

Beyond the idealized design storms, the modeled catchment was also evaluated under real rainfall data collected from 1 April to 1 December 2013, which corresponded with the time period during which the SWMM model was calibrated (CDM Smith, 2015). As opposed to using individual events, a long-term simulation offers additional insight into performance under rainfall variability while allowing model initial conditions to settle toward realistic values throughout the duration of the simulation (Gironás et al., 2009). Both synthetic storms and long-term rainfall data were sampled at a 5-min resolution and serve as the rainfall time series for each the simulation. It should be noted again that the control algorithm is feedback-based, rather than predictive, which means that control decisions are based on current conditions and not subject to weather uncertainty. Nonetheless, to evaluate performance under uncertainty, an additional analysis was carried out during which virtual sensor noise was injected into the water levels retrieved from the physically based model. Noise levels were parameterized based on a water level sensors currently being used (MaxBotix MB7384, 2012), sampled at each time step from a Gaussian distribution of mean zero and standard deviation σ = 2.5 mm. The impact of two larger noise levels was also investigated, wherein the realistic noise levels were amplified by 5 and 10 times to evaluate how robust the algorithm is under high levels of measurement uncertainty.

To evaluate the performance of the LQR-based feedback controller, a baseline performance objective was first established by evaluating how the uncontrolled system responds to a relatively small event (2-year, 24-hr storm). During this storm, there were no overflows at any of the SNs and the peak flows in the catchment reached 0.3 m^3/s at the outlet. It was then evaluated if the controlled system could reach the same baseline performance during larger events. While infrastructure designed for 100-year, 24-hr events may also perform very well for a 2-year, 24-hr storm, it however would be over designed. Instead, rather than continue to dig up pipes or build bigger SNs, this was intended to reflect the benefits of retrofitting existing infrastructure with gates and valves and operating them in coordination to improve the use of existing storage elements and limit flows throughout the network. In effect, this reduces flooding as well as stream erosion, which is often triggered through the exceedance of geomorphically significant flow magnitude (Bledsoe, 2002; Pomeroy et al., 2008; Tillinghast et al., 2011). This reasoning also aligns with many current infrastructure design philosophies, which seek to capture larger storms and release them as smaller storms (MacRae, 1996; Nehrke & Roesner, 2004; Rohrer & Roesner, 2006; e.g., capture a 10-year event and release with outflows comparable to a 2-year event prior to the addition of control measures). To maintain a clear relationship between tuning parameters and LQR control performance (15), the tuning parameter ρ was set to 3,500 following manual tuning during the 10-year storm, while the **Q** and **R** matrices were set to be identity matrices. In this configuration, **Q** was a 55 × 55 identity matrix, and **R** was an 11 × 11 identity matrix.

The performance was evaluated across the entire system by combining the volume of flooding along with the flow exceedance ($Q^*_{i,\max}$ =0.29 m^3/s) across the duration of an entire storm. Specifically,

$$P = \sum_{\text{nodes}} \sum_{\text{timesteps}} Q_{i,\text{flood}}(k) + \alpha \cdot \max\left(Q_{i,\text{outflow}}(k) - Q^*_{i,\max}, 0\right). \tag{25}$$

The weighing parameter α can be tuned to reflect the relative importance of each objective (localized flooding vs. downstream erosion, e.g.). In this analysis, α was chosen to be 0.1 to scale the outflows to have the same magnitude as the overflows.

First, the performance P for both the controlled and uncontrolled systems was compared for a 10-year, 24-hr event, assuming that all 11 SNs were controlled. In our study area, this event is designated by regulatory guidelines as the design storm that all new developments must meet (Pratt, 2016). Next, using the same storm, the remaining configurations for controlled systems were evaluated, ranging from only one site being controlled to all 11 sites being controlled in coordination. Altogether, these 2,048 configurations were then ranked to determine which specific configuration provided the best performance, seeking to identify which sites and features may be indicative of good performance. This search was then repeated for 5-, 25-, 50-, 100-, and 200-year, 24-hr storms to confirm if the same configuration was consistent for larger storm events of similar duration. Once the top configuration was identified, its performance was compared to the uncontrolled case over a comprehensive array of storms, ranging across 5-, 10-, 25-, 50-, 100-, and 200-year return periods, across durations from 15 min to 24 hr. Finally, to investigate how much smaller storage volumes could be constructed when control is used, the volume of the controlled SNs was reduced until the system-level performance matched the performance of the uncontrolled system.

6. Results

6.1. Performance of a Fully-Controlled Network

On average, when all SNs were controlled, the LQR-based control approach outperformed the uncontrolled system, both at the scale of individual sites and at the watershed outlet. Specifically, when evaluated on the 10-year, 24-hr storm, only one SN had an overflow event and outflows at each SN did not exceed the critical flow level of 0.3 m^3/s. Plotted in Figure 4 is a dynamic comparison for the uncontrolled (gray line) and coordinated control (blue line) of SNs across the system for a 10-year, 24-hr storm. The outflows of the uncontrolled SNs exhibited the familiar hydrograph shape, with a distinct peak and recession period, whereas the outflow hydrographs from the controlled sites exhibited dynamics with lower flows over a longer period of time. This also resulted in longer retention times and higher water levels in the controlled system since water was held within the SNs so as to not exceed the outflow threshold.

6.2. Top-Performing Individual Control Sites

For the 10-year, 24-hr storm, when only one SN was controlled at a time, 6 of the 11 possible sites showed a notable improvement compared to the uncontrolled system, as measured by the performance across the entire catchment (Figure 5). The baseline performance for the uncontrolled case was $P_0 = 2{,}100$. The control of one site, in particular (SN4), exhibited an improvement in reducing system-wide overflows and outflows compared to the uncontrolled system, where $P_{SN4} = 1{,}100$. Control of 3 of the 11 SNs did not result in improvements compared to the uncontrolled system because the uncontrolled system already met the control objectives. Only one site (SN9) performed worse when controlled, exhibiting local flooding compared to the uncontrolled case, where $P_{SN9} = 2{,}200$. This SN had the smallest storage capacity in the entire system but a relatively large contributing catchment area. Adding a controller led to closure of the valve at the onset of a storm, after which the SN filled up and could not be drained fast enough once the peak of the storm arrived.

6.3. Benefits of Increasing the Number of Controlled Sites

The impact of adding controllers was evaluated based on design performance (25). The addition of the first control valve added the biggest benefit. The addition of more control valves improved the performance, but each successive valve led to marginal returns (Figure 5). The major exception to this trend occurred when the control network was expanded to include 10 and 11 controlled SNs, which resulted to slight degradation in performance due to local flooding at smaller SNs. In all, over 15,000 simulations (carried out across all possible 2,048 possible control configurations and over 24-hr 1-, 2-, 5-, 10-, 25-, 50-, 100-, and 200-year storms) showed that adding more control valves improved the relative performance of the overall system, as measured by the performance function (25). The time required to simulate a single 24-hr storm event was approximately 5 min, requiring over 50 days in total computation time for 15,000 simulations.

When analyzing the performance of all possible 2,048 control configurations, SNs 4, 6, 10, and 11 consistently appeared in configurations that ranked in the top 10% (Figure 6). Interestingly, these same SNs were those that showed the relatively best performance when controlled individually. Overall, out of all 2,048 possible

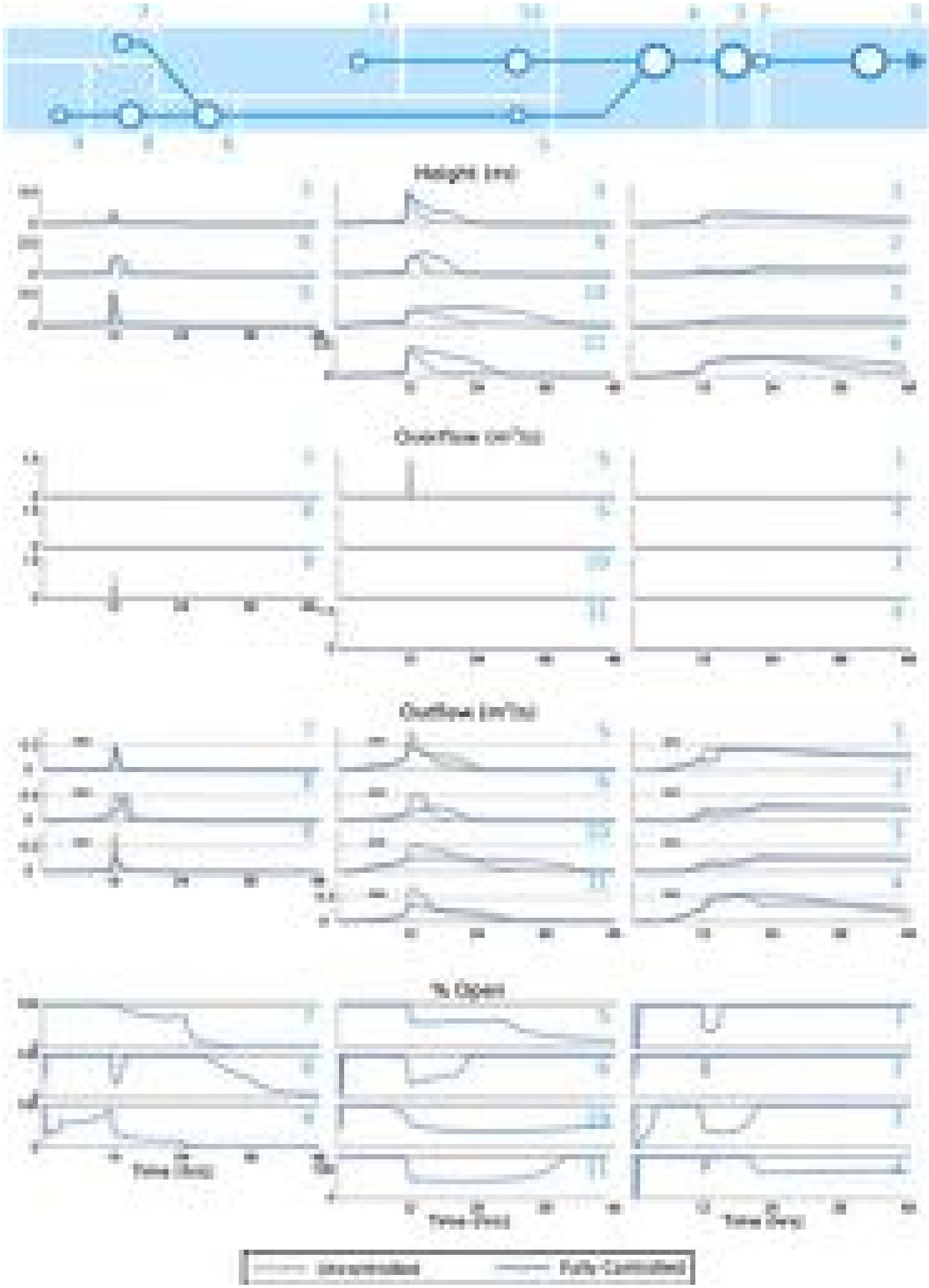


Figure 4. System diagram of storage nodes, where the connected boxes are scaled to represent the relative area of the total subcatchment that contributes runoff to each respective storage node. The response in height, flow, overflow, and valve opening at each node is plotted for cases with zero controllers (no control, gray line) and 11 controllers operated in coordination by the system-level controller.

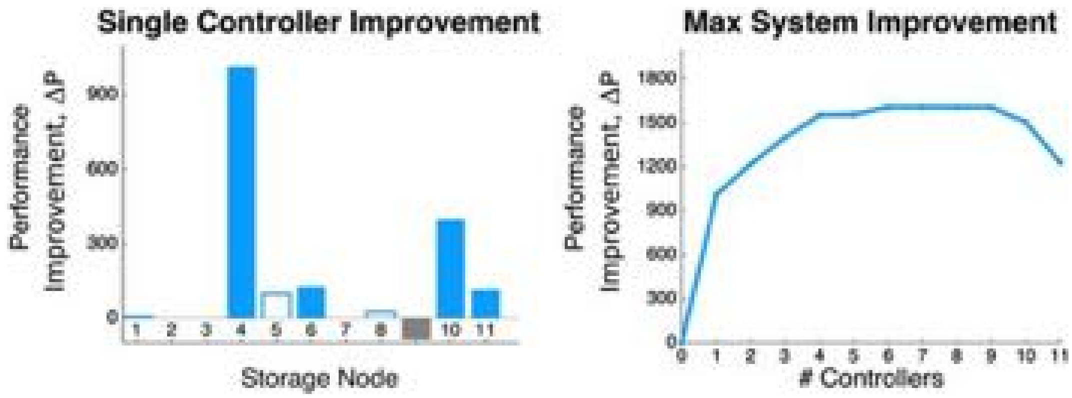


Figure 5. System-wide performance improvement as evaluated by equation (25) compared to the case where there is no control for a 10-year, 24-hr storm event (P_0 = 2,100) when only one valve is controlled at a time (left) and when increasing the number of controlled valves (right).

combinations of controlled sites, the control of SNs 4, 6, 10, and 11 resulted in the best performance for the 10-year, 24-hr storm using the least number of control points. No physical features (volume, location in watershed, etc.) were consistent across SNs 4, 6, 10, and 11 as an explanation as to why these sites performed the best.

In addition to the 10-year, 24-hr storm, the same minimal control configuration (SNs 4, 6, 10, and 11) was compared to the uncontrolled system across a spectrum of design storms (5 to 200-year storm events spanning 15 min to 24 hr in duration). Overall, this configuration improved performance as seen by the notable expansion in the number of cases, or zones, associated with relatively high performance (no overflows and low outflows), as indicated by the larger dark blue region in Figure 7. By reducing overflows and lowering outflows, the minimal controlled system, in turn, improved the metric P from 2,100 to 500 for a 10-year, 24-hr event (Figure 7). In fact, for an approximately 30-year, 24-hr storm event, the minimal controlled system had the same performance (P = 2,100) as the uncontrolled system during a 10-year, 24-hour event (solid lines, Figure 7). In other words, the controlled system was able to handle much bigger storms, without compromising performance. Furthermore, it was determined that the SNs that were controlled could be reduced in volume by over 50% and still achieve the same performance as the original uncontrolled system.

6.4. Weather Variability and Measurement Noise

The result of running the control algorithm for eight months of rainfall data is shown in Figure 8, comparing the uncontrolled system to the controlled system under zero noise and 10 times the realistic noise level. For illustrative purposes, the figure plots the depth of the outlet as well as its outflows. All remaining nodes are

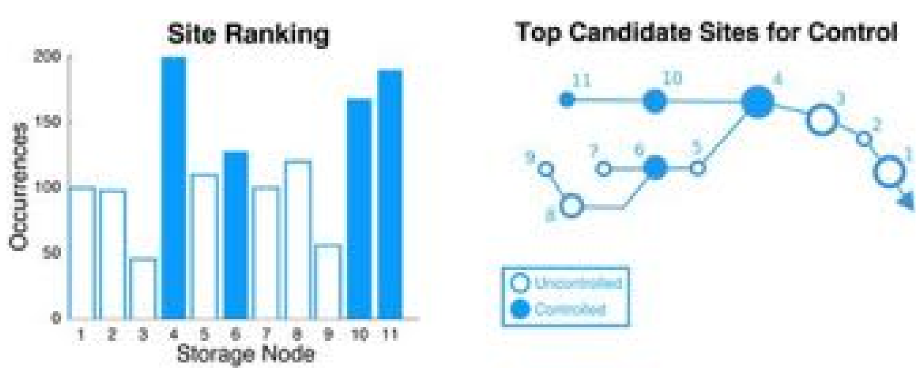


Figure 6. The ranking and location of the storage nodes that appeared most frequently in the top 10% of performances for all controller combinations.

10.1029/2018WR022657

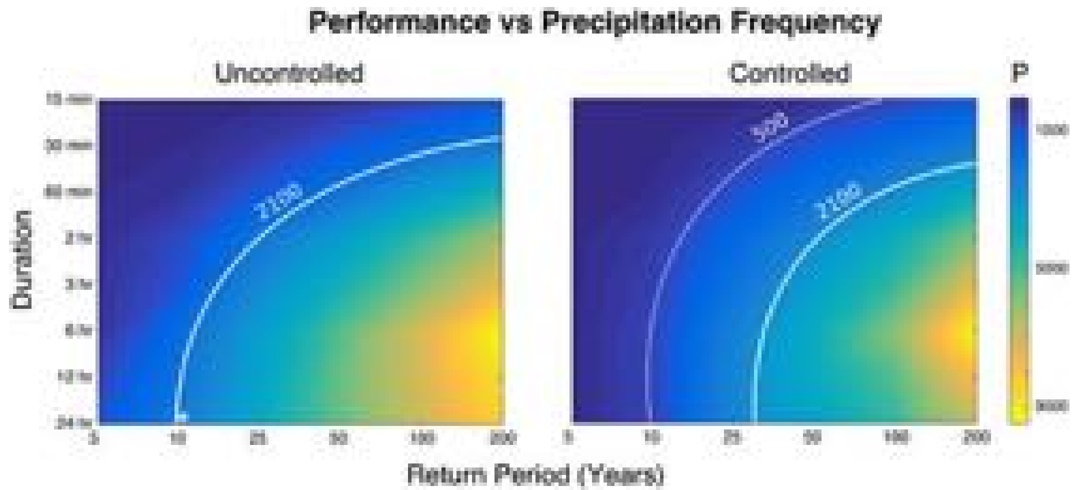


Figure 7. System performance for a range of design storms with various durations and intensities. Nodes 4, 6, 10, and 11 are controlled in the controlled case. The controlled system "shifts" the performance zone notably, achieving the same performance for a 30-year storm as the uncontrolled system for a 10-year storm. The star in the uncontrolled case marks the performance of the same four controlled nodes when they have half the original storage volume.

plotted in supporting information S1 of this paper. For the simulated time period, except for nodes 8 and 9, the controlled outflow did not exceed the critical flow threshold of $Q^{*}_{i,\max} = 0.029$ m^3/s (Figures S1–S3). This threshold corresponded to the response that would have been seen with the uncontrolled system given a 1-year 5-min storm event.

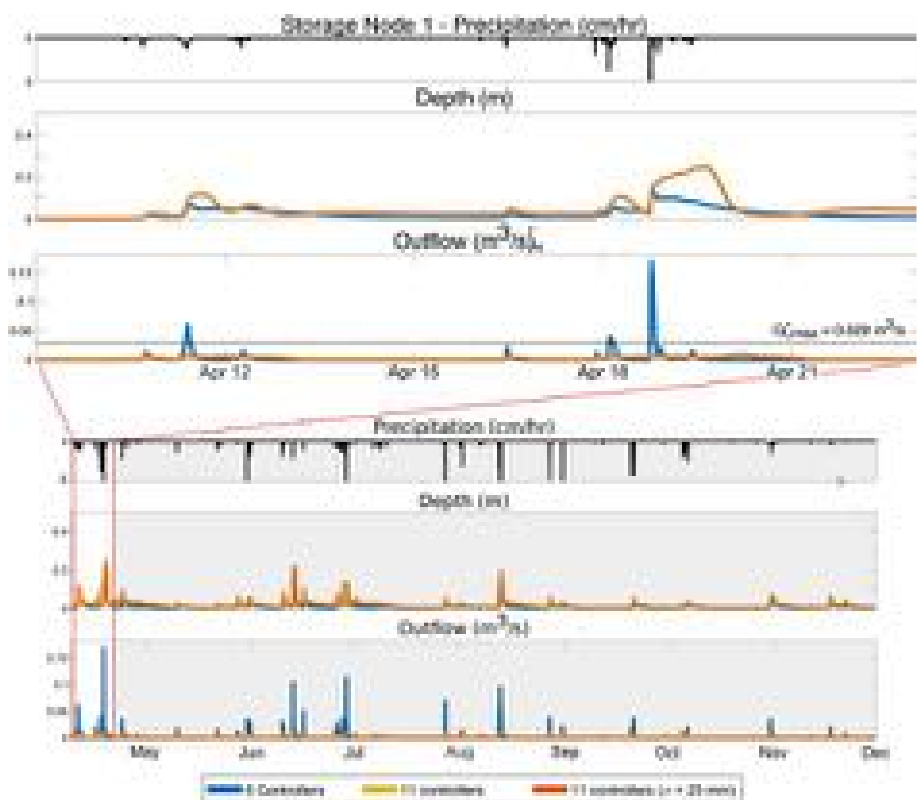


Figure 8. The depth and outflow from node 1 simulated from 1 April to 1 December 2013 and a detail of the rainfall leading up to 19 April—the first storm used in the initial calibration. Conditions are plotted for no control and control with both perfect measurements and measurements with white noise (σ = 25 mm).

For the realistic noise scenario (σ = 2.5 mm), the difference in control performance was not visually evident, which is why the figure plots only the performance for the highest, 10-fold noise level. Even with artificial measurement noise as much as 10 times larger than a realistic noise levels, the controlled system outperformed the uncontrolled system. For example, P_{mod} = 1,237 for the uncontrolled case, whereas with 11 controllers, P_{mod} = 77 with σ = 25 mm. Even when controlling SNs 4, 6, 10, and 11 at 50% volume and σ = 25 mm, P_{mod} = 360. The overall performance is summarized in Table S1, which compares the performance of an uncontrolled system, a fully controlled system, a system with only four controlled basis, as well a scenario wherein the four basins are modeled at 50% of their actual volume.

7. Discussion

7.1. Control Performance

Despite the nonlinear hydraulic and hydrologic dynamics inherent to water systems like urban watersheds, an LQ feedback controller was able to notably improve flow management throughout the modeled catchment. This was not only true for the idealized design storm scenarios but also when controlling the system using historical rainfall data and subjecting it to measurement noise. It is important to note that there was, at the onset, no guarantee that this was going to be the case, since the linear representation of the control dynamics may have seemed oversimplified compared to more traditional and physically based urban watershed models. This validates the use of feedback-based methods, even for the control of complex and nonlinear stormwater catchments.

It is likely that nonlinearities and travel times may become more important as the scale of the controlled systems grows. While this may be worth exploring in future studies, it may also result in the need for more complex control approaches. To that end, the authors contend that the control of larger watersheds could be achieved by controlling individual subwatersheds, such as the one in our case study, and setting their outflows to meet a cumulative goal at the larger scale. As alluded to earlier, this may also be more realistic, since many management and design decisions occur at these smaller headwater scales.

In this case, the framework presented in this study shows great promise for the initial analysis of real-time control at the urban headwater scale (1–5 km^2). Should the control of larger areas become important, control techniques from other domains (e.g., power grid operations) could be adopted. These include centralized controllers like structured and reciprocal LQR (Bareiss & van den Berg, 2015; Yang & Cimen, 1996), as well as decentralized approaches like droop control (Riccobono et al., 2016; Zhu et al., 2013) and agent-based and game-theoretic approaches (Lemos & Pinto, 2012; Maestre et al., 2014; Zhang & Li, 2007) which require a control model to approximate the behavior of neighboring systems. While the deployment of real-time controllable stormwater valves is still in its infancy, the results of our simulations suggest that physical watershed properties can indeed be used to formulate a valid control model for real-time control. This presents exciting new research opportunities to investigate the generalizability and scalability of the framework beyond our study catchment by coupling it with these broader control techniques toward the control neighboring catchments and ultimately entire urban watersheds.

In our study, the control objectives were tuned to reduce outflows and flooding, but the controller and cost function could readily be extended to meet other goals. For example, systems could be tuned to maximally retain water by keeping storage levels near capacity. This would increase hydraulic residence times after storms and thus help with the treatment of sediment-bound and dissolved pollutants. Our study also showed how outflows from the catchment could be "shaped" beyond a traditional hydrograph. In this study, the controlled hydrograph was flat for the majority of its duration, rather than exhibiting a clear peak. By dynamically changing the setpoint of the controller, other outflows patterns could be achieved.

The opportunity to set desired outflows and water levels based on management objectives will open up entirely new management possibilities, which should be evaluated through future research. For example, control valves could be used to mimic "predevelopment" conditions, which is often the goal of many stormwater infrastructure projects (Guan et al., 2015). Furthermore, rather than operating the system in a one-size-fits-all configuration, valves could be controlled based on multiobjective management goals. For example, the system could be operated for water quality benefits during smaller, more frequent storms and operated for flood control during larger storms. This further highlights the flexibility of real-time controlled systems, as their operation can adapt not only to unaccounted conditions such as noise but also with changing watershed-level management goals, something that is difficult to accomplish using passive infrastructure.

7.2. Number of Control Points

As expected, the addition of each subsequent valve improves the performance of the control system. This is intuitive, since each control point provides additional dynamic storage and flexibility to buffer flows. More importantly, however, the benefit of successively adding valves is marginal, whereby adding more valves may not improve performance. Economically, this is important, since it suggests that the entire catchment may not need to be controlled, but rather that simulation and engineering judgment can be used to determine the number of required control valves for a catchment. The number of required valves could then be chosen based on a specific management goal or by finding the point at which investment into more valves will not provide major returns.

Ultimately, each stormwater system has performance limits, which are a function of the hydrology, infrastructure, control objectives, and costs, as well as the specific control algorithms. Real-time control will only be able to push water system to a certain point, beyond which new infrastructure construction may need to be considered. As illustrated in our own study, if construction is needed, new sites can be smaller when real-time control is used. This is particularly important in many urban areas, where cost of construction is high and land availability may be limited.

In addition to having a sufficient number of control points, it is also important to determine where to place valves to maximize catchment-wide benefits. Given the lack of prior studies on this topic, our approach exhaustively simulated every possible configuration of valves across the entire catchment for multiple events, which required over 15,000 model runs. For the catchment studied in our case study, the locations selected for control could be prioritized based on their individual performance. For our study catchment, this means that the simulations of only one valve at a time (all other sites uncontrolled) could be used to rank sites, after which multivalve configurations could be made by combining valves that had the best individual performance. As such, the number of simulations required for valve selection may only need to be as high as the number of candidate sites in a catchment, which may speed up future analyses for site selection. In practice, rather than requiring a specific configuration for a given number of valves, valves could be added without needing to change the location of the valve placed before it. This is very important, since valves could be added one-by-one to benefit the overall system, rather than requiring a preset configuration. Beyond exhaustive simulation, theoretical placement approaches (e.g., Liu et al., 2011) should also be evaluated, but they will need to be adapted to the unique temporal and nonlinear dynamics of stormwater systems.

The physical characteristics of what makes one site more suitable for control than another are still not very clear. In our study, most of the SNs that ranked the highest in their ability to improve catchment-wide performance had a relatively low catchment area to volume ratio. In other words, they received very limited local runoff but had large storage volume. This made them relatively suitable for buffering flows from upstream sites. As such, in-line storage may be a big factor in the ability of a site to contribute to catchment-wide benefits. However, this was not necessarily true of all "good" control sites and instead may be dependent on the actual catchment being studied. As such, more studies will be required in the future to determine if this is a reliable feature when selecting new control sites.

It did become clear in our study, however, that there are types of sites that may lead to worse performance compared to the uncontrolled case. This was particularly evident for SN5 and SN9, which overtopped when controlled. This occurred because the site had a small storage volume but large contributing runoff area, which did not permit it to react to rapid changes in runoff. For such smaller SNs, which are subjected to very flashy inputs, feedback control should likely not be applied unless the cost function is adjusted for more conservative outcomes. This may also be overcome by predictive control, which will not only respond to real-time states but also to forecasts for weather and runoff. Given the performance of the feedback-based LQR controller during both event and long-term simulations, the application of model predictive control (Hashemy et al., 2013; van Overloop, 2006; van Overloop et al., 2008) in urban hydrologic catchments appears very promising and will be evaluated in future studies. In fact, the role of weather uncertainty remains unstudied in the emerging field of real-time stormwater control and poses a promising research frontier.

This study serves as a baseline for assessing the integration of active control measures into urban catchments using dynamical control. While the approach shows great promise, several limiting assumptions were made that will need to be addressed in future studies. In all simulations, the SNs were initially empty, and some remained filled after a storm. Once the storm had passed, nodes could be slowly drained to meet the outflow constraint. While this is not possible in passive systems, it may also become a problem if another storm begins before the SNs have been entirely drained. Alternatively, some catchments, such as those in the dry regions of the world, may be configured for stormwater capture and reuse. In those cases, keeping SNs full becomes an objective, which poses risks to the control of the system if not enough storage is available to buffer incoming storms. This, again, stresses the importance of weather forecasts and their inherent uncertainty, which will need to be studied to determine how a system can be prepared ahead of incoming storms. Since our algorithm relies only on feedback, this may require the future investigation of model predictive control formulation. Without constraints, model predictive control mathematically becomes an LQR controller (Garcia et al., 1989; Morari & Lee, 1999), which suggests that model predictive control may be applied successfully using the integrator delay formulation presented in this study.

Given the nascent nature of real-time control in urbanized hydrologic catchments, there is an urgent need to develop a framework to compare controlled and uncontrolled catchments on an equal footing. While it may be tempting to showcase plots of controlled hydrographs, the number of plots can quickly balloon, even for small systems. The cost functions that are used to parameterize control algorithms do not underpin the language used by decision makers and may ineffectively communicate the benefits to be gained by real-time control. To that end, an equivalence analysis will be necessary to contextualize and synthesize these benefits in terms of traditional systems. As opposed to the tables used in previous studies, we believe that visualizations, such as those in Figure 7, will provide a baseline intuition that can be used to more effectively promote adoption.

Further, while varying degrees of noise were considered, our centralized LQ controller assumed full knowledge of all states and the impacts of estimation to compensate for incomplete knowledge remain to be investigated. While modern sensors are becoming much more reliable and accurate, the role of sensor placement, measurement uncertainty, and sensor reliability must be studied to ensure robust control performance. Given the dynamical formulation of our framework, this could be accomplished through formal estimation approaches (e.g., Kalman filters; Faragher, 2012).

8. Conclusions

Active control poses an exciting frontier with the potential to transform stormwater management of urban watersheds. We introduce a novel framework for analyzing the impact of real-time control across urban headwater catchments. By confirming the ability of feedback control to achieve desired flows and reduce flooding, the approach offers an alternative to new construction, which is currently the only solution to cope with changes in land use and weather. The approach would, of course, need to be evaluated further not only in simulation but also in the real world, which should now be very feasible given that the necessary sensing and control technologies have been developed. The retrofitting of catchments should also be aided by the discovery that only a few key locations may need to be controlled, but this should still be validated on catchment-by-catchment basis.

Much research remains to be conducted to determine the generalizability and scalability of the methods proposed in this paper. In particular, the control of larger urban watersheds should be evaluated. The authors contend that control at this larger scale may be most effectively achieved through the control of many smaller catchment "building blocks." The need to segment control into smaller clusters may also be motivated by the practicality of working across ownership boundaries, insurance requirements, and social constraints. Social factors may ultimately become the most important barrier to the adoption of real-time control, since the best control algorithms may only be as good as the willingness of the public to adopt them. As such, there will be many opportunities to engage other disciplines in the emerging area of research. To encourage outreach and engagement, detailed technical information regarding software and implementation of this study are provided online at http://open-storm.org/.

Acknowledgments
This work was financially supported by the Great Lakes Protection Fund under project 1035. Simulation tools can be found at https://github.com/open-storm/, and results can be found at http://open-storm.org/data/.

References

Akitsu, K., Evan, C. H., & David, S. (2011). Comparison of synthetic design storms with observed storms in southern Arizona. *Journal of Hydrologic Engineering*, *16*(11), 935–941.

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, *111*, D05109. https://doi.org/10.1029/2005JD006290

Askarizadeh, A., Rippy, M. A., Fletcher, T. D., Feldman, D. L., Peng, J., Bowler, P., et al. (2015). From rain tanks to catchments: Use of low-impact development to address hydrologic symptoms of the urban stream syndrome. *Environmental Science & Technology*, *49*(19), 11,264–11,280. https://doi.org/10.1021/acs.est.5b01635

Aulinas, M., Nieves, J. C., Cortés, U., & Poch, M. (2011). Supporting decision making in urban wastewater systems using a knowledge-based approach. *Environmental Modelling & Software*, *26*(5), 562–572. https://doi.org/10.1016/j.envsoft.2010.11.009

Avellaneda, P., Jefferson, A., Grieser, J., & Bush, S. (2017). Simulation of the cumulative hydrological response to green infrastructure. *Water Resources Research*, *53*, 3087–3101. https://doi.org/10.1002/2016WR019836

Bareiss, D., & van den Berg, J. (2015). Generalized reciprocal collision avoidance. *The International Journal of Robotics Research*, *34*(12), 1501–1514. https://doi.org/10.1177/0278364915576234

Barreiro-Gomez, J., Obando, G., Riano-Briceno, G., Quijano, N., & Ocampo-Martinez, C. (2015). Decentralized control for urban drainage systems via population dynamics: Bogotá case study. Paper presented at the Control Conference (ECC), 2015 European, 2426–2431.

Barreiro-Gomez, J., Riaño-Briceño, G., Ocampo-Martínez, C., & Quijano, N. (2017). Data-driven evolutionary-game-based control for drinking-water networks. In *Real-time monitoring and operational control of drinking-water systems* (pp. 363–383). Cham: Springer.

Bartos, M., Wong, B., & Kerkez, B. (2017). Open storm: A complete framework for sensing and control of urban watersheds. arXiv Preprint arXiv:1708.05172.

Benedict, M. A., McMahon, E. T., & The Conservation (2012). *Green infrastructure: Linking landscapes and communities*. Washington, DC: Island Press.

Bledsoe, B. P. (2002). Stream erosion potential and stormwater management strategies. *Journal of Water Resources Planning and Management*, *128*(6), 451–455. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:6(451)

Boyer, K. B., & Kieser, M. S. (2012). Urban stormwater mangement—An MS4 success story for western Michigan university. *Journal of Green Building*, *7*(1), 28–39. https://doi.org/10.3992/jgb.7.1.28

Bryant, S., & Wadzuk, B. M. (2017). Modeling of a real-time controlled green roof. Paper presented at the World Environmental and Water Resources Congress 2017, 357–364.

Cantone, J., & Schmidt, A. (2011). Improved understanding and prediction of the hydrologic response of highly urbanized catchments through development of the Illinois Urban Hydrologic Model. *Water Resources Research*, *47*, W08538. https://doi.org/10.1029/2010WR009330

Carpenter, J. F., Vallet, B., Pelletier, G., Lessard, P., & Vanrolleghem, P. A. (2014). Pollutant removal efficiency of a retrofitted stormwater detention pond. *Water Quality Research Journal*, *49*(2), 124–134. https://doi.org/10.2166/wqrjc.2013.020

CDM Smith (2015). City of Ann Arbor stormwater model calibration and analysis project—Final report. Retrieved from https://www.a2gov.org/departments/systems-planning/planning-areas/water-resources/Documents/A2_SWM_Report_20150601.pdf

Cheng, L., & AghaKouchak, A. (2014). Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Science Reports*, *4*, 7093.

Dahleh, M., Dahleh, M. A., & Verghese, G. (2004). Lectures on dynamic systems and control (Vol. 4, pp. 1–100). Department of Electrical Engineering and Computer Science, Massachuasetts Institute of Technology, *4*(100), 1–100.

Dorato, P., Abdallah, C. T., Cerone, V., & Jacobson, D. H. (1995). *Linear-quadratic control: An introduction*. Englewood Cliffs, NJ: Prentice Hall.

Eggimann, S., Mutzner, L., Wani, O., Schneider, M. Y., Spuhler, D., Moy de Vitry, M., et al. (2017). The potential of knowing more: A review of data-driven urban water management. *Environmental Science and Technology*, *51*(5), 2538–2553. https://doi.org/10.1021/acs.est.6b04267

Emerson, C. H., Welty, C., & Traver, R. G. (2005). Watershed-scale evaluation of a system of storm water detention basins. *Journal of Hydrologic Engineering*, *10*(3), 237–242. https://doi.org/10.1061/(ASCE)1084-0699(2005)10:3(237)

Faragher, R. (2012). Understanding the basis of the Kalman filter via a simple and intuitive derivation [lecture notes]. *IEEE Signal Processing Magazine*, *29*(5), 128–132. https://doi.org/10.1109/MSP.2012.2203621

Fletcher, T. D., Vietz, G., & Walsh, C. J. (2014). Protection of stream ecosystems from urban stormwater runoff: The multiple benefits of an ecohydrological approach. *Progress in Physical Geography*, *38*(5), 543–555. https://doi.org/10.1177/0309133314537671

Francis, B. A., & Wonham, W. M. (1976). The internal model principle of control theory. *Automatica*, *12*(5), 457–465.

Gaborit, E., Anctil, F., Pelletier, G., & Vanrolleghem, P. (2016). Exploring forecast-based management strategies for stormwater detention ponds. *Urban Water Journal*, *13*(8), 841–851. https://doi.org/10.1080/1573062X.2015.1057172

Gaborit, E., Muschalla, D., Vallet, B., Vanrolleghem, P. A., & Anctil, F. (2013). Improving the performance of stormwater detention basins by real-time control using rainfall forecasts. *Urban Water Journal*, *10*(4), 230–246. https://doi.org/10.1080/1573062X.2012.726229

Garcia, C. E., Prett, D. M., & Morari, M. (1989). Model predictive control: Theory and practice—A survey. *Automatica*, *25*(3), 335–348. https://doi.org/10.1016/0005-1098(89)90002-2

García, L., Barreiro-Gomez, J., Escobar, E., Téllez, D., Quijano, N., & Ocampo-Martínez, C. (2015). Modeling and real-time control of urban drainage systems: A review. *Advances in Water Resources*, *85*, 120–132. https://doi.org/10.1016/j.advwatres.2015.08.007

Gill, M. A. (1978). Flood routing by the Muskingum method. *Journal of Hydrology*, *36*(3–4), 353–363. https://doi.org/10.1016/0022-1694(78)90153-1

Gilpin, A., & Barrett, M. (2014). Interim report on the retrofit of an existing flood control facility to improve pollutant removal in an urban watershed. Paper presented at the World Environmental and Water Resources Congress 2014, 65–74.

Gironás, J., Roesner, L. A., Davis, J., Rossman, L. A., & Supply, W. (2009). Storm water management model applications manual. National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency Cincinnati, OH.

Gironás, J., Roesner, L. A., Rossman, L. A., & Davis, J. (2010). A new applications manual for the storm water management model (SWMM). *Environmental Modelling and Software*, *25*(6), 813–814.

Goff, K. M., & Gentry, R. W. (2006). The influence of watershed and development characteristics on the cumulative impacts of stormwater detention ponds. *Water Resources Management*, *20*(6), 829–860. https://doi.org/10.1007/s11269-005-9010-2

Grigg, N. S. (2012). *Water, wastewater, and stormwater infrastructure management*. Boca Raton, FL: CRC Press.

Grigg, N. S. (2015). Infrastructure report card: Purpose and results. *Journal of Infrastructure Systems*, *21*(4), 02514001.

Guan, M., Sillanpää, N., & Koivusalo, H. (2015). Assessment of LID practices for restoring pre-development runoff regime in an urbanized catchment in southern Finland. *Water Science and Technology*, *71*(10), 1485–1491. https://doi.org/10.2166/wst.2015.129

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645–1660. https://doi.org/10.1016/j.future.2013.01.010

Guo, Y. (2001). Hydrologic design of urban flood control detention ponds. *Journal of Hydraulic Engineering*, *6*(6), 472–479. https://doi.org/10.1061/(ASCE)1084-0699(2001)6:6(472)

Hart, J. K., & Martinez, K. (2006). Environmental sensor networks: A revolution in the earth system science? *Earth-Science Reviews*, *78*(3–4), 177–191. https://doi.org/10.1016/j.earscirev.2006.05.001

Hashemy, S. M., Monem, M. J., Maestre, J. M., & Overloop, P. J. V. (2013). Application of an in-line storage strategy to improve the operational performance of main irrigation canals using model predictive control. *Journal of Irrigation and Drainage Engineering*, *139*(8), 635–644. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000603

HRWC (2013). Malletts creekshed report. Retrieved from https://www.hrwc.org/wp-content/uploads/Malletts_8x11.5.pdf

Keeley, M., Koburger, A., Dolowitz, D. P., Medearis, D., Nickel, D., & Shuster, W. (2013). Perspectives on the use of green infrastructure for stormwater management in Cleveland and Milwaukee. *Environmental Management*, *51*(6), 1093–1108. https://doi.org/10.1007/s00267-013-0032-x

Kerkez, B., Gruden, C., Lewis, M. J., Montestruque, L., Quigley, M., Wong, B. P., et al. (2016). Smarter stormwater systems. *Environmental Science and Technology*, *50*(14), 7267–7273. https://doi.org/10.1021/acs.est.5b05870

Kessler, R. (2011). Stormwater strategies: Cities prepare aging infrastructure for climate change. *Environmental Health Perspectives*, *119*(12), A514–A519. https://doi.org/10.1289/ehp.119-a514

Klepiszewski, K., & Schmitt, T. (2002). Comparison of conventional rule based flow control with control processes based on fuzzy logic in a combined sewer system. *Water Science and Technology*, *46*(6–7), 77–84. https://doi.org/10.2166/wst.2002.0665

Lancaster, P., & Rodman, L. (1995). *Algebraic Riccati equations*. Oxford.

Lawson, R., Riggs, E., Weiker, D. & Doubek, J. (2017). Total suspended solids reduction—Implementation plan for Malletts Creek: October 2011–September 2016. Retrieved from http://www.hrwc.org/wp-content/uploads/2011/11/MallettsCreek_BiotaTMDL_FINAL.pdf

Lee, J. G., Selvakumar, A., Alvi, K., Riverson, J., Zhen, J. X., Shoemaker, L., & Lai, F. (2012). A watershed-scale design optimization model for stormwater best management practices. *Environmental Modelling & Software*, *37*, 6–18. https://doi.org/10.1016/j.envsoft.2012.04.011

Lee, J. H., & Bang, K. W. (2000). Characterization of urban stormwater runoff. *Water Research*, *34*(6), 1773–1780. https://doi.org/10.1016/S0043-1354(99)00325-5

Lemos, J. M., & Pinto, L. F. (2012). Distributed linear-quadratic control of serially chained systems: Application to a water delivery canal [applications of control]. *IEEE Control Systems*, *32*(6), 26–38.

Lin, F. (2007). *Robust control design: An optimal control approach*. Chichester, UK: John Wiley. https://doi.org/10.1002/9780470059579

Litrico, X., & Fromion, V. (2009). *Modeling and control of hydrosystems*. Dordrecht, Heidelberg: Springer Science & Business Media. https://doi.org/10.1007/978-1-84882-624-3

Liu, Y., Slotine, J., & Barabási, A. (2011). Controllability of complex networks. *Nature*, *473*(7346), 167–173. https://doi.org/10.1038/nature10011

MacRae, C. (1996). Experience from morphological research on Canadian streams: Is control of the two-year frequency runoff event the best basis for stream channel protection? Paper presented at the Effects of Watershed Development and Management on Aquatic Ecosystems, 144-162.

Maestre, J., de la Peña, D. M., & Camacho, E. (2014). Distributed MPC based on agent negotiation. In *Distributed model predictive control made easy* (pp. 465–477). Switzerland: Springer.

Malaterre, P. O., & Baume, J. P. (1998). Modeling and regulation of irrigation canals: Existing applications and ongoing researches. Paper presented at the Systems, Man, and Cybernetics, 1998.1998 {IEEE} International Conference On*, 4* 3850–3855 o.4.

Marinaki, M., & Papageorgiou, M. (2003). Linear-quadratic regulators applied to sewer network flow control. Paper presented at the European Control Conference (ECC)*,* 2003, 2407–2412.

Marsalek, J. (2005). Evolution of urban drainage: From cloaca maxima to environmental sustainability. Acqua E Citta, I Convegno Nazionale Di Idraulica Urbana, Cent.Stud.Idraul.Urbana, Santâ€™Agnello Di Sorrento, Italy, 28–30.

MaxBotix MB7384 (2012). Retrieved from http://www.maxbotix.com/Ultrasonic_Sensors/MB7384-Ultrasonic_Snow_Depth_Sensor.htm

Middleton, J. R., & Barrett, M. E. (2008). Water quality performance of a batch-type stormwater detention basin. *Water Environment Research*, *80*(2), 172–178. https://doi.org/10.2175/106143007X220842

Montestruque, L., & Lemmon, M. (2015). Globally coordinated distributed storm water management system. Paper presented at the Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks, 10.

Morari, M., & Lee, H. J. (1999). Model predictive control: Past, present and future. *Computers and Chemical Engineering*, *23*(4–5), 667–682. https://doi.org/10.1016/S0098-1354(98)00301-9

Mullapudi, A., Wong, B. P., & Kerkez, B. (2017). Emerging investigators series: Building a theory for smart stormwater systems. *Environmental Science: Water Research & Technology*, *3*(1), 66–77.

Muschalla, D., Vallet, B., Anctil, F., Lessard, P., Pelletier, G., & Vanrolleghem, P. A. (2014). Ecohydraulic-driven real-time control of stormwater basins. *Journal of Hydrology*, *511*, 82–91. https://doi.org/10.1016/j.jhydrol.2014.01.002

Nehrke, S. M., & Roesner, L. A. (2004). Effects of design practice for flood control and best management practices on the flow-frequency curve. *Journal of Water Resources Planning and Management*, *130*(2), 131–139. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:2(131)

Ocampo-Martinez, C., Puig, V., Cembrano, G., & Quevedo, J. (2013). Application of predictive control strategies to the management of complex networks in the urban water cycle [applications of control]. *IEEE Control Systems*, *33*(1), 15–41.

Pagano, M. A., & Perry, D. (2008). Financing infrastructure in the 21st century city. *Public Works Management and Policy*, *13*(1), 22–38. https://doi.org/10.1177/1087724X08321015

Petrucci, G., Rioust, E., Deroubaix, J., & Tassin, B. (2013). Do stormwater source control policies deliver the right hydrologic outcomes? *Journal of Hydrology*, *485*, 188–200. https://doi.org/10.1016/j.jhydrol.2012.06.018

Pomeroy, C., Postel, N., O'Neill, P., & Roesner, L. (2008). Development of storm-water management design criteria to maintain geomorphic stability in Kansas City metropolitan area streams. *Journal of Irrigation and Drainage Engineering*, *134*(5), 562–566. https://doi.org/10.1061/(ASCE)0733-9437(2008)134:5(562)

Pratt, E. N. (2016). Rules and guidelines—Procedures and design criteria for stormwater.

Ramírez-Llanos, E., & Quijano, N. (2010). A population dynamics approach for the water distribution problem. *International Journal of Control*, *83*(9), 1947–1964. https://doi.org/10.1080/00207179.2010.501389

Rawls, W. J., Brakensiek, D. L., & Miller, N. (1983). Green-ampt infiltration parameters from soils data. *Journal of Hydraulic Engineering*, *109*(1), 62–70. https://doi.org/10.1061/(ASCE)0733-9429(1983)109:1(62)

Riaño-Briceño, G., Barreiro-Gomez, J., Ramirez-Jaime, A., Quijano, N., & Ocampo-Martinez, C. (2016). MatSWMM—An open-source toolbox for designing real-time control of urban drainage systems. *Environmental Modelling & Software*, *83*, 143–154. https://doi.org/10.1016/j.envsoft.2016.05.009

Riccobono, A., Ferdowsi, M., Hu, J., Wolisz, H., Jahangiri, P., Müller, D., et al. (2016). Next generation automation architecture for DC smart homes. Paper presented at the Energy Conference (ENERGYCON), 2016 IEEE International, 1–6.

Rohrer, C., & Roesner, L. (2006). Matching the critical portion of the flow duration curve to minimise changes in modelled excess shear. *Water Science and Technology*, *54*(6–7), 347–354. https://doi.org/10.2166/wst.2006.590

Rosenberg, E. A., Keys, P. W., Booth, D. B., Hartley, D., Burkey, J., Steinemann, A. C., & Lettenmaier, D. P. (2010). Precipitation extremes and the impacts of climate change on stormwater infrastructure in Washington state. *Climatic Change*, *102*(1–2), 319–349. https://doi.org/10.1007/s10584-010-9847-0

Rossman, L., & Huber, W. (2016). *Storm water management model reference manual volume I—Hydrology (revised)*. Cincinnati, OH: US Environmental Protection Agency.

Rossman, L. A. (2010). Storm water management model user's manual, version 5.0 National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency Cincinnati, OH.

Schütze, M., Campisano, A., Colas, H., Schilling, W., & Vanrolleghem, P. A. (2002). Real-time control of urban wastewater systems-where do we stand today? *Journal of Hydrology*, *299*(3–4), 335–348. https://doi.org/10.1016/j.jhydrol.2004.08.010

Schuurmans, J. (1997). Control of water levels in open-channels.

Schuurmans, J., Bosgra, O., & Brouwer, R. (1995). Open-channel flow model approximation for controller design. *Applied Mathematical Modelling*, *19*(9), 525–530. https://doi.org/10.1016/0307-904X(95)00053-M

Sinha, A. (2007). *Linear systems: Optimal and robust control*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781420008883

Strecker, E. W., Quigley, M. M., Urbonas, B. R., Jones, J. E., & Clary, J. K. (2001). Determining urban storm water BMP effectiveness. *Journal of Water Resources Planning and Management*, *127*(3), 144–149. https://doi.org/10.1061/(ASCE)0733-9496(2001)127:3(144)

Tillinghast, E., Hunt, W., & Jennings, G. (2011). Stormwater control measure (SCM) design standards to limit stream erosion for piedmont North Carolina. *Journal of Hydrology*, *411*(3–4), 185–196. https://doi.org/10.1016/j.jhydrol.2011.09.027

Unruh, M., & Yekta, G. B. (2013). Precipitation-frequency atlas of the United States, midwestern states. NOAA Atlas 14, 8(2).

U.S. SCS (1986). Urban hydrology for small watersheds, technical release no. 55 (TR-55). US Department of Agriculture, US Government Printing Office, Washington, DC.

van Overloop, P. (2006). *Model predictive control on open water systems*. Delft, Netherlands: IOS Press.

van Overloop, P., Weijs, S., & Dijkstra, S. (2008). Multiple model predictive control on a drainage canal system. *Control Engineering Practice*, *16*(5), 531–540. https://doi.org/10.1016/j.conengprac.2007.06.002

Van Overloop, P. J., Negenborn, R. R., De Schutter, B., & Van De Giesen, N. C. (2010). Predictive control for national water flow optimization in the Netherlands. *Intelligent Infrastructures*, *42*, 439–461.

Walsh, C. J., Roy, A. H., Feminella, J. W., Cottingham, P. D., Groffman, P. M., & Morgan, R. P. (2005). The urban stream syndrome: Current knowledge and the search for a cure. *Journal of the North American Benthological Society*, *24*(3), 706–723. https://doi.org/10.1899/04-028.1

Wang, J., & Guo, Y. (2018). An analytical stochastic approach for evaluating the performance of combined sewer overflow tanks. *Water Resources Research*, *54*, 3357–3375. https://doi.org/10.1029/2017WR022286

Wong, B. P., & Kerkez, B. (2014). Adaptive, decentralized, and real-time sampling strategies for resource constrained hydraulic and hydrologic sensor networks. Paper presented at the HIC 2014 – 11th International Conference on Hydroinformatics, New York, USA. 1 Retrieved from http://academicworks.cuny.edu/cc_conf_hic/235/

Wong, B. P., & Kerkez, B. (2016a). Wong Brandon Preclaro and Kerkez Branko. GitHub Repository. Retrieved from https://github.com/klabum/iot

Wong, B. P., & Kerkez, B. (2016b). Adaptive measurements of urban runoff quality. *Water Resources Research*, *52*, 8986–9000. https://doi.org/10.1002/2015WR018013

Xu, M., Van Overloop, P., & Van De Giesen, N. (2011). On the study of control effectiveness and computational efficiency of reduced Saint-Venant model in model predictive control of open channel flow. *Advances in Water Resources*, *34*(2), 282–290. https://doi.org/10.1016/j.advwatres.2010.11.009

Yang, T., & Cimen, H. (1996). Applying structured singular values and a new LQR design to robust decentralized power system load frequency control. Paper presented at the Industrial Technology, 1996. (ICIT'96), Proceedings of the IEEE International Conference On, 880–884.

Zhang, Y., & Li, S. (2007). Networked model predictive control based on neighbourhood optimization for serially connected large-scale processes. *Journal of Process Control*, *17*(1), 37–50. https://doi.org/10.1016/j.jprocont.2006.08.009

Zhen, X., Yu, S. L., & Lin, J. (2004). Optimal location and sizing of stormwater basins at watershed scale. *Journal of Water Resources Planning and Management*, *130*(4), 339–347. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:4(339)

Zhu, L., Liu, J., Cupelli, M., & Monti, A. (2013). Decentralized linear quadratic Gaussian control of multi-generator MVDC shipboard power system with constant power loads. Paper presented at the Electric Ship Technologies Symposium (ESTS), 2013 IEEE, 308–313.

Environmental Science & Technology

Feature

pubs.acs.org/est

Smarter Stormwater Systems

Branko Kerkez,[*,†] Cyndee Gruden,[‡] Matthew Lewis,[§] Luis Montestruque,[∥] Marcus Quigley,[⊥] Brandon Wong,[†] Alex Bedig,[⊥] Ruben Kertesz,[∥] Tim Braun,[∥] Owen Cadwalader,[⊥] Aaron Poresky,[#] and Carrie Pak[∇]

[†]University of Michigan, Department of Civil and Environmental Engineering, Ann Arbor, Michigan 48109, United States

[‡]University of Toledo, Department of Civil Engineering, Toledo, Ohio 43606, United States

[§]Michigan Aerospace Corporation, Ann Arbor, Michigan 48108, United States

[∥]Emnet LLC, South Bend, Indiana 46617, United States

[⊥]OptiRTC, Inc., Boston, Massachusetts 02116, United States

[#]Geosyntec Consultants, Atlanta, Georgia, United States

[∇]Clean Water Services, Hillsboro, Oregon 97123, United States

ABSTRACT: Existing stormwater systems require significant investments to meet challenges imposed by climate change, rapid urbanization, and evolving regulations. There is an unprecedented opportunity to improve urban water quality by equipping stormwater systems with low-cost sensors and controllers. This will transform their operation from static to adaptive, permitting them to be instantly "redesigned" to respond to individual storms and evolving land uses.

■ INTRODUCTION

The design of stormwater and sewer systems is based on historical observations of precipitation and land use. These systems require significant investments to meet challenges imposed by rapid urbanization, evolving regulations and an uncertain climate. As a result, runoff from urban environments is threatening environmental health by lowering the quality of receiving waters, including fisheries, recreational sites and sources of drinking water. There is an unprecedented opportunity, however, to improve urban water flow and quality by equipping existing stormwater systems with low-cost sensors and controllers. This will enable a new generation of *intelligent* green and gray stormwater networks, which will adapt their operation to maximize water quality benefits in response to individual storm events and changing landscapes.

■ STATIC SOLUTIONS TO A DYNAMIC PROBLEM

The vast majority of the world's population resides in or near urban centers, underscoring the need to sustainably manage anthropogenic environmental impacts. Urbanization and land development are disruptive to the hydrologic cycle since they result in an altered, more impervious landscape, which promotes increased runoff at the expense of infiltration and evapotranspiration.[1,2] While most cities maintain a dedicated stormwater infrastructure, ecosystems near many postindustrial cities in the U.S. are adversely impacted by exfiltration and overflows from combined sewers.[3−5] These overflows have increased due to leaks in aging infrastructure and shrinking municipal budgets.

The increase in the volume, velocity and contaminants in stormwater runoff has caused a crisis in receiving water bodies.[6−9] Harmful algal blooms, associated with anthropogenic inputs of nutrients, have resulted in unsafe drinking water, impaired fisheries and damage to recreational waters.[10−14] As such, managing pollutant loadings from urban stormwater has become one of our most pressing environmental challenges.[15,16]

Expansion and upsizing of *gray* infrastructure are perhaps the most common solutions to coping with increased runoff resulting from changing weather and land use.[17] Aggressive climate adaptation via traditional tools may lead to over-designed gray infrastructure, which conveys water too quickly to streams, leading to floodplain encroachment, increases in runoff volumes, and stream erosion. To preserve stream stability and ecological function, advances in stormwater science are calling for traditional peak attenuation designs to be replaced with those that reduce stream erosion during smaller, more frequent storms.[18] As communities seek more

Published: May 26, 2016

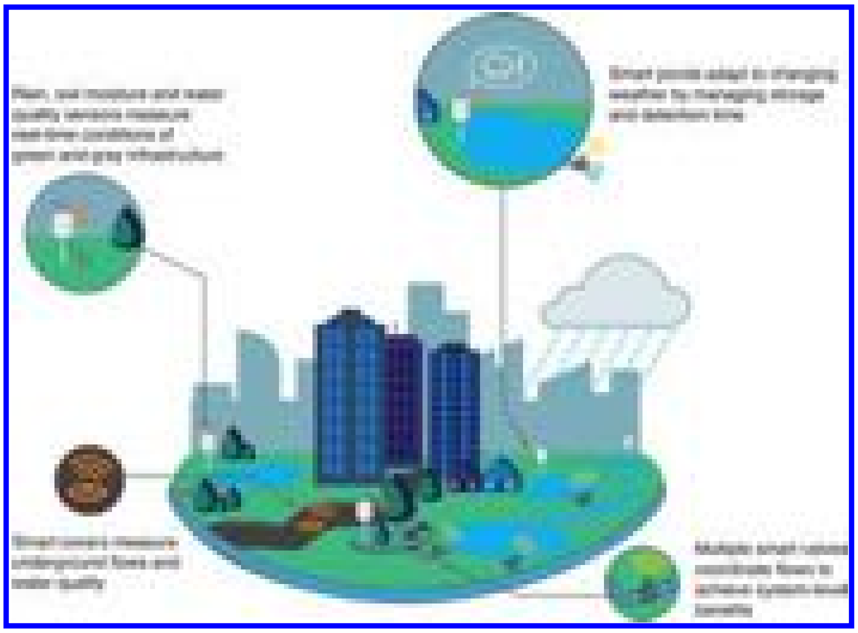


Figure 1. System-level stormwater measurement and control.

resilient and adaptive stormwater solutions, novel and non-traditional alternatives to new construction must be considered.

One such alternative is provided by *green infrastructure (GI)*, which augments impervious urban areas with pervious solutions such as bioswales, green roofs and rain gardens.[19−21] GI is designed to restore some ecosystem functions to preurbanization levels by capturing runoff and contaminants before they enter the stormwater system. These solutions have experienced a significant rise in popularity due to their promise to offer a low impact alternative toward buffering flows and improving runoff water quality.[22] Much research remains to be conducted, however, to test the efficacy and scalability of GI as an alternative to gray infrastructure. To that end, more cost-effective sensing solutions are required to assess the in situ performance and improve the maintenance of GI.[23,24]

While stormwater systems do change (albeit slowly), their design performance is often regarded as static due to limited ability to adapt to changing climate and land uses. More importantly, stormwater solutions are engineered on a site-by-site basis, with little consideration given to ensuring that local benefits are actually adding up to achieve a collective outcome.[25] Rather than offering an alternative, a new solution promises to augment, rather than replace, green and gray infrastructure. This approach relies heavily on sensor and information technology to make existing stormwater systems more adaptive by embedding them with connectivity and *intelligence*.

■ REAL-TIME ADAPTIVE MANAGEMENT

The past decade has witnessed significant advances and reduction in the cost of novel sensors, wireless communications and data platforms. In large, much of this development has accompanied the recent boom on the *Internet of Things (IoT)*, a technological movement that promises to build the next generation of interconnected and *smart* buildings and cities.[26] The stormwater sector has been slow in its adoption of these technologies, especially in the context of high-resolution and real-time decision-making. Present uses of sensors range from regulatory compliance[27,28] to performance studies of individual stormwater facilities.[29] These technological advances have the potential to become highly transformative, however, by enabling stormwater infrastructure to evolve from static to highly adaptive (Figure 1). By coupling the flow of water with the flow of information, modern stormwater infrastructure will adapt itself in real-time to changing storms and land uses, while simultaneously providing a highly cost-effective solution for cities that are otherwise forced to spend billions on stormwater reconstruction.[30]

Given advances of modern sensors and data acquisition systems, it is now feasible to monitor green and gray infrastructure projects pre- and postconstruction to provide in situ performance metrics. This is afforded by a significant reduction in the cost of sensors and cloud-hosted real-time data systems. Many commercial and open-source platforms, specifically geared toward demands imposed by storm and sewer applications, are now available and promising to lower the cost of wireless sensor deployments. Water flow, stage, precipitation and soil-moisture can now be measured seamlessly and continuously. The development of robust and affordable in situ water quality sensors for nutrients, metals or bacteria is still evolving.

While new measurements will provide significant insight into the study and management of stormwater systems, it is the ability to directly and proactively control these systems that presents the biggest potential impact to water quality. Low-cost, reliable and secure actuators (e.g., valves, gates, pumps) can now be attached to existing stormwater systems to control the flow of water in pipes, ponds and green infrastructure. Examples include inflatable pillows that can be used to take advantage of underused inline storage,[31] or *smart* outlet structures that control water levels in response to real-time data and weather forecasts (Figure 2).

While real-time process control in water and wastewater treatment has been studied extensively and continues to be a fruitful area of research,[33] there is now the opportunity to distribute these treatment ideas to the watershed scale. This presents an exciting new paradigm: *retrofitting existing stormwater infrastructure through cost-effective sensors and actuators will*

DOI: 10.1021/acs.est.5b05870
Environ. Sci. Technol. 2016, 50, 7267−7273

Figure 2. Example sensing and control devices (a) Remote valve for basin control, (b) smart sensing manhole cover, and (c) an open-source sensor node for distributed measurement and control.[32]

transform its operation from static to adaptive, permitting it to be instantly "redesigned" to respond to changing conditions. There is an inherent complexity associated with control of city-scale systems, however, as they are comprised of a variety of gray and green solutions and driven by complex storm patterns, hydrologic phenomena, and water quality dynamics. The number of studies addressing real-time water quality control is limited but promising, ranging from local- to city-scale control.

REAL-TIME CONTROL OF INDIVIDUAL STORMWATER FACILITIES

Many existing studies focus on the real-time control of stormwater basins and ponds, which are some of the most common elements in a stormwater system.[34−36] Pollutant removal in basins comprises a complex interaction between a number of mechanisms, including sedimentation, flotation, infiltration, biological conversion, and degradation.[37] Traditionally, these facilities are designed as compromises between flood control (detention) and water quality control (retention), with limited ability to adapt functionality to individual storm events. Retrofitting an existing site with a real-time control valve permits it to serve both as a detention and retention basin, as well as a spectrum of in-between configurations. One control rule, for example, opens a valve to drain a pond if a storm is forecasted, which creates additional storage for incoming runoff. Similarly, runoff can be strategically retained after a storm to improve settling and biological uptake. It has been shown, for example, that by temporarily converting a detention basin to a retention basin, the removal efficiency of total suspended solids (TSS) increased from 39% (189 120 g inflow vs 98 269 g outflow) to 90% (e.g., 59 807 g inflow vs 8055 g outflow) and ammonia-nitrogen increased from 10% (101.1 g inflow vs 79.2 g outflow) to nearly 90% (e.g., 163.5 g inflow vs 7.8 g outflow).[37,38] Using data from these studied, Mushalla et al.[39] simulated that retaining water using real-time controls may result in up to a 60% improvement in small particle removal compared to a traditional design.

Some studies are also beginning to show that real-time control can play a significant role in removing biological, metal and dissoved contaminants. A controlled basin in Pflugerville, Texas, achieved 6-fold reduction in nitrate plus nitrite-nitrogen compared to the same preretrofit dry basin (0.66 mg/L to 0.11 mg/L) by extending detention time and releasing water before a storm to create additional storage.[40] While biological uptake likely contributed to nitrogen removal, reliable and affordable in situ sensors for many dissolved pollutants are still needed to fully understand the impacts of control to dissoved pollutant removal in natural treatment systems.

Real-time control of a retrofitted detention pond showed that the removal of *Escherichia coli* was improved by strategically retaining water for 24 h after a storm rather than allowing the water to flow though the pond as originally designed.[41] For the controlled basin the outlet concentrations were an order of magnitude lower than inlet concentrations (1940 MPN/100 mL in vs 187 MPN/100 mL out; and 3410 MPN/100 mL in vs 768 MPN/100 mL out), whereas the uncontrolled basin showed limited removal and even increased *E. coli* at the outlet (4350 MPN/100 mL in vs 8860 MPN/100 mL out; 10800 MPN/100 mL in vs 11000 MPN/100 mL out). Since streambed concentrations of *E. coli* were three times higher than in the streamwater, the primary mechanisms for removal were attributed to sedimentation and increased exposure to sunlight. This example also speaks to the need to be cognizant of flow releases from controlled basins, as high outflows can resuspend pollutants. As such, real-time control can be used to modulate the flow rate from storage facilities to reduce downstream erosion and pollutant loads. Such strategies begin to place real-time control into a much broader systems context, whereby each individual stormwater facility not only generates local benefits, but can also be used to improve flow and water quality at the city-scale.

Flow modulation for stream protection was demonstrated at two pilot sites owned by Clean Water Services (CWS) in Washington County, Oregon. In one system (sized to retain 0.2 in. of rainfall), the addition of real-time control to an existing wet pond reduced the volume and duration of channel forming discharges by approximately 25%. In a second facility (a dry detention pond), the use of real-time control was used to minimize release rates in smaller, more frequent storm events while maintaining the ability to match predevelopment peak flows during larger storms. This enhancement was modeled to reduce the volume of erosive flows by nearly 60% and the volume of wet weather discharges by nearly 70% compared to a passive basin (Figure 3). Additionally, the use of real-time control increased the average residence time of this facility from 1 to 19 h. In a simulation case study real-time control reduced

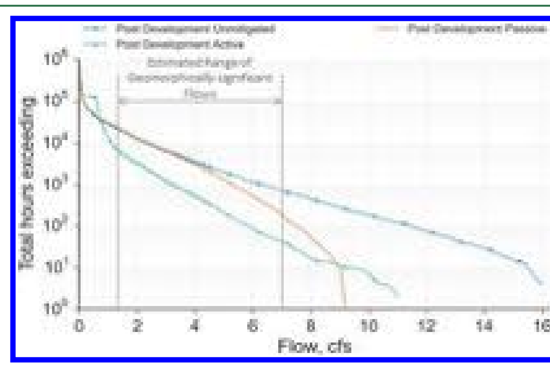


Figure 3. Improvement achieved by retrofitting an existing basin to reduce erosive flows.

DOI: 10.1021/acs.est.5b05870
Environ. Sci. Technol. 2016, 50, 7267−7273

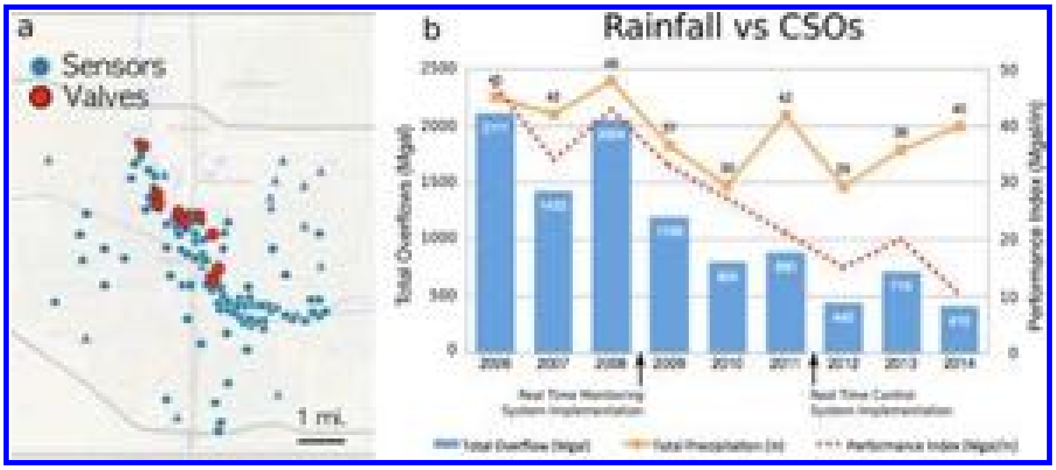


Figure 4. Comparison of combined sewer overflows (CSOs) before and after commissioning of real-time sensing and control system in South Bend, IN.

the required pond volume by 30–50%, compared to a passive facility, while achieving the same level of flow-duration control performance. Finally, based on whole lifecycle cost estimates, it was determined that a real-time control retrofit of an existing stormwater detention facility would be approximately three times lower in lifecycle cost than the equivalent passive alternative.[42]

■ SCALING UP

An insight into the scalability of real-time control is provided by a large-scale control network that is presently deployed in South Bend, Indiana.[43] The network encompasses 100 km^2 and is comprised of 120 real-time flow and water depth sensors (Figure 4), which share information every 5 min. The system has been retrofitted with control valves located at nine CSO regulators to modulate flow into the city's interceptor line. The control valves allow more water to enter the interceptor line when conveyance capacity is available, while avoiding surcharging the interceptor, which may cause surface flooding or structural damage. The system operates by taking advantage of excess conveyance capacity within the interceptor line, which is driven dynamically by spatial or temporal features of specific storms.

The distributed control strategy uses an agent-based control scheme to optimize the water collection system, whereby each infrastructure component trades its own storage or conveyance capacity to other upstream assets, similar to traders in a stock market.[44] Even before the system was controlled, benefits were achieved by means of monitoring alone. By isolating maintenance issues in its first year of operation (2008), the system helped the utility eliminate critical dry weather sewer overflows, which were occurring an average of 27 times per year. Overall, the control system reduced total sewer overflow volumes from 2100 MGal to 400 MGal from 2006–2014 (Figure 4). Even after adjusting for total annual rainfall, a near 5-fold performance improvement (ratio of overflows to precipitation) was achieved. While a reduction in *E. coli* concentrations (443 cfu/100 mL to 234 cfu/100 mL) in the downstream sewer locations was also observed, a more comprehensive ecological study is warranted to study the impacts of real-time control to *E. coli* removal mechanisms. It is estimated that over one billion gallons of untreated sewer flows were blocked from flowing into the river, suggesting that real-time control played a role in improving water quality.

■ KNOWLEDGE GAPS

Systems Thinking. While nascent, research on real-time stormwater control is not limited by technology, but rather by a much more fundamental need to understand the complex spatiotemporal dynamics that govern water flow and quality across large urban areas. One of the largest challenges with existing stormwater solutions relates to their design as single entities. This means that benefits achieved at a local scale may often be masked or eliminated at the city scale if the performance of an individual element is not designed in a broader systems context.[25,45] Perhaps the biggest benefit of control relates to the ability to leverage real-time interconnection to guide the behavior of individual elements to achieve city-scale benefits.

There is a need to build upon prior and ongoing research efforts on best management practices (BMPs)[20,29,46,47] to understand how individual green and gray stormwater solutions perform when stressed by varying climate, storms, and runoff dynamics. Many studies focus on hydrologic control and removal of solids and bacteria, but much work still remains to be done to determine the impacts of these solutions to the treatment of metals, nutrients and emerging contaminants. This will require the expanded development of cheap and reliable sensors for these pollutants. Furthermore, there is an urgent need to fill a knowledge and measurement gap on the interconnectedness of BMPs across various scales and runoff dynamics (e.g., first flush vs peak flow). By improving the understanding of stormwater networks as a function of scale, it will then be possible to posit how very large systems (ten to a hundred ponds, for example) should be controlled or tuned in real-time to achieve a collective outcome.

Uncertainty. The role of uncertainty is rarely acknowledged in the design of traditional stormwater systems, since it is assumed that many transient system behaviors will average out into a cumulative performance over time. The benefits of real-time control, however, are highly underpinned by uncertainties related to weather forecasts, models, control algorithms, and sensor measurements. Some elements of the system will always remain unmeasured or not understood. Furthermore, many control decisions will continue to be based on hydraulic parameters, such as flow or residence time. Until reliable and low-cost water quality sensors become available, water quality control decisions will rely on statistical correlations or physical

DOI: 10.1021/acs.est.5b05870
Environ. Sci. Technol. 2016, 50, 7267–7273

models. It will be important to quantify the role of the resulting "error bars" on the performance of real-time control.

As with many controlled systems, there may be an inherent risk to infrastructure, private property, or even human life due to poorly designed control algorithms. Since risk relates directly to uncertainty, reliable and consistent real-time operations can only be achieved by exhaustively quantifying the role of uncertainty in control operations. Furthermore, even the best controllers and sensors may only achieve marginal benefits if storms cannot be predicted adequately, thus calling for the need to begin investigating the value of weather forecasts in control operations. Many other examples can be given, but studies exploring the role of uncertainty have yet to be conducted.

■ OUTLOOK AND BROADER ADOPTION

Real-time control promises to revolutionize the management of urban water quality by providing the ability to significantly improve the operation of existing stormwater assets. As the community of researchers grows, there will be a need to develop baseline performance metrics, study sites, measurement platforms, and data sets. Research on stormwater capture and direct use (reuse) has recently increased[48] due to the potential of reclaimed stormwater to serve arid regions. In drought-prone regions of the U.S., where stormwater direct use is becoming one of the few viable water recovery options, sensing and real-time control will improve stormwater extraction compared to static or natural treatment options. Controlling the timing and magnitude of flows and improving removal of contaminants before they reach the plant will also result in a reduction in resources required for treatment in combined sewer systems.

Outfitting stormwater infrastructure with sensors and digital control systems introduces new opportunities for efficiency and new risks of failure. Responsible use of these systems extends beyond deployment, requiring ongoing effort to maintain trust in the data produced and the integrity with which control actions are followed. As with all Best Management Practices,[20] standards will be required to facilitate broader adoption of real-time control and to assess the risks introduced by the use of information sourced from these embedded systems. Future standards may focus around data formats, sensor requirements or actuator specifications, and will need to ensure interoperability between various sites. Failure to recognize, plan for, and manage the ongoing cyber security risks introduced by the distributed installation of sensors and actuators in stormwater infrastructure will result in new risks to public health and safety, which may undermine trust in broader efforts to deliver the potential benefits of these technologies.

There will be a need to address regulatory compliance, ownership, governance, and operational jurisdictions relating to real-time controlled systems. Unlike existing deployments of sensor and control systems in wastewater treatment, digital stormwater infrastructure is deployed across a watershed, outside of buildings staffed by an operations team. A key tension relates to jurisdiction, both in terms of who owns the infrastructure being controlled and which software system provides this dynamic capacity. Many cities may only wish to try retrofitting some sites, with the plan to augment their systems over time as they see benefits. This raises the possibility that many software systems may operate simultaneously and interfere with a global goal. If control systems are deployed by a spectrum of public and private stakeholders, they should nonetheless interoperate to provide capacity for watershed-scale control and maintenance. Governance models must be explored to facilitate cooperation and liability concerns. While solutions to these concerns can build on successful models used for ownership and operation of passive controls, they may require further thought in their translation to real-time controlled systems.

Beyond technical challenges, the ecosystem of municipalities and engineering firms must adapt to accommodate real-time control within a large umbrella of green and gray infrastructure solutions. Broader community engagement is necessary to facilitate dissemination and adoption of real-time stormwater control. Compliance regulators, such as state and federal environmental protection agencies, must be highlighted as members of this community, since many cities are wary of innovation because of perceptions that regulators will reject nontraditional solutions. Environmental consulting firms, municipalities, and researchers will need to acquire nontraditional skillsets, which span electrical engineering and computer science. To help with this effort, a major initiative is presently underway to organize an open-source consortium and share reference implementations on real-time stormwater control (http://open-storm.org). While open-source options for sensing and control are alluring due to their perceived cost, examples of holistic open-source approaches, which integrate environmental science, technology and engineering design, have yet to be developed. To that end, this consortium will serve as a hub for reference applications, standards, architectures, sensors, hardware and algorithms, to show that it is well within the abilities of most academic groups, municipalities and engineering firms to begin instrumenting and controlling stormwater infrastructure.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bkerkez@umich.edu.

Notes

The authors declare no competing financial interest.

Biography

Branko Kerkez (Assistant Professor) and Brandon Wong (Graduate Student) are part of the *Real-time Water System Lab* in the Department of Civil and Environmental Engineering at the University of Michigan. Cyndee Gruden is an Associate Professor at the University of Toledo, Department of Civil Engineering. Dr. Matt Lewis is the CTO of *Michigan Aerospace*, in Ann Arbor, Michigan. Dr. Luis Montestruque (CTO), Ruben Kertesz and Tim Braun are employed at *EmNet* in South Bend Indiana. Marcus Quigley (CEO), Alex Bedig and Owen Cadwalader are employed at *OptiRTC*, while Aaron Poresky works with *GeoSyntec*. Carrie Pak works at *Clean Water Services* in Hillsboro, Oregon.

■ ACKNOWLEDGMENTS

Branko Kerkez, Cyndee Gruden, Matthew Lewis and Brandon Wong are supported by the Great Lakes Protection Fund. We acknowledge the CWS staff, including Richard Boyle, Jadene Stensland and Andy Braun, Doug Schuh, and Jeff Van Note, as well as the city of South Bend, including Gary Gilot, Eric Horvath, Patrick Henthorn, Al Greek and Jack Dillon. All data used to generate the figures in this paper are available at http://open-storm.org/data

DOI: 10.1021/acs.est.5b05870
Environ. Sci. Technol. 2016, 50, 7267−7273

REFERENCES

(1) Zhu, T.; Lund, J. R.; Jenkins, M. W.; Marques, G. F.; Ritzema, R. S. Climate Change, Urbanization, and Optimal Long-Term Floodplain Protection. *Water Resour. Res.* **2007**, *43* (6).n/a10.1029/2004WR003516

(2) Mallin, M. A.; Johnson, V. L.; Ensign, S. H. Comparative Impacts of Stormwater Runoff on Water Quality of an Urban, a Suburban, and a Rural Stream. *Environ. Monit. Assess.* **2009**, *159* (1−4), 475−491.

(3) Semadeni-Davies, A.; Hernebring, C.; Svensson, G.; Gustafsson, L.-G. The Impacts of Climate Change and Urbanisation on Drainage in Helsingborg, Sweden: Combined Sewer System. *J. Hydrol.* **2008**, *350* (1−2), 100−113.

(4) Divers, M. T.; Elliott, E. M.; Bain, D. J. Constraining Nitrogen Inputs to Urban Streams from Leaking Sewers Using Inverse Modeling: Implications for Dissolved Inorganic Nitrogen (DIN) Retention in Urban Environments. *Environ. Sci. Technol.* **2013**, *47* (4), 1816−1823.

(5) Sercu, B.; Van De Werfhorst, L. C.; Murray, J. L. S.; Holden, P. A. Sewage Exfiltration as a Source of Storm Drain Contamination during Dry Weather in Urban Watersheds. *Environ. Sci. Technol.* **2011**, *45* (17), 7151−7157.

(6) Booth, D. B.; Jackson, C. R. Urbanization of Aquatic Systems: Degradation Thresholds, Stormwater Detection, and the Limits of Mitigation. *J. Am. Water Resour. Assoc.* **1997**, *33* (5), 1077−1090.

(7) Finkebine, J. K.; Atwater, J. W.; Mavnic, D. S. Stream Health after Urbanization. *J. Am. Water Resour. Assoc.* **2000**, *36* (5), 1149−1160.

(8) Wang, L.; Lyons, J.; Kanehl, P.; Bannerman, R. Impacts of Urbanization on Stream Habitat and Fish Across Multiple Spatial Scales. *Environ. Manage.* **2001**, *28* (2), 255−266.

(9) Barco, J.; Hogue, T. S.; Curto, V.; Rademacher, L. Linking Hydrology and Stream Geochemistry in Urban Fringe Watersheds. *J. Hydrol.* **2008**, *360* (1−4), 31−47.

(10) Sahagun, L. *High Cost of Fighting Urban Runoff Examined in Report*; LA Times. September 2, 2013.

(11) Wines, M. *Behind Toledo's Water Crisis, a Long-Troubled Lake Erie*; New York Times. August 5, **2014**.

(12) Doughton, S. Toxic Runoff Kills Adult Coho Salmon Study Finds*Seattle Times*2015

(13) Ahn, J. H.; Grant, S. B.; Surbeck, C. Q.; DiGiacomo, P. M.; Nezlin, N. P.; Jiang, S. Coastal Water Quality Impact of Stormwater Runoff from an Urban Watershed in Southern California. *Environ. Sci. Technol.* **2005**, *39* (16), 5940−5953.

(14) Carey, R. O.; Wollheim, W. M.; Mulukutla, G. K.; Mineau, M. M. Characterizing Storm-Event Nitrate Fluxes in a Fifth Order Suburbanizing Watershed Using in Situ Sensors. *Environ. Sci. Technol.* **2014**, *48* (14), 7756−7765.

(15) Vörösmarty, C. J.; McIntyre, P. B.; Gessner, M. O.; Dudgeon, D.; Prusevich, A.; Green, P.; Glidden, S.; Bunn, S. E.; Sullivan, C. A.; Liermann, C. R.; et al. Global Threats to Human Water Security and River Biodiversity. *Nature* **2010**, *467* (7315), 555−561.

(16) United States Environmental Protection Agency. *Nonpoint Source Pollution: The Nation's Largest Water Quality Problem*; 1997.

(17) Rosenberg, E. A.; Keys, P. W.; Booth, D. B.; Hartley, D.; Burkey, J.; Steinemann, A. C.; Lettenmaier, D. P. Precipitation Extremes and the Impacts of Climate Change on Stormwater Infrastructure in Washington State. *Clim. Change* **2010**, *102* (1−2), 319−349.

(18) Hawley, R. J.; Vietz, J. G. Addressing the Urban Stream Disturbance Regime. *Freshwater Sci.* **2016**, *35* (1).27810.1086/684647

(19) Coffman, L. S.; Goo, R.; Frederick, R. Low-Impact Development: An Innovative Alternative Approach to Stormwater Management. *Bridges* **1999**, *10*, 118.

(20) Strecker, E.; Quigley, M. M.; Urbonas, B. R.; Jones, J.; Clary, J. Determining Urban Storm Water BMP Effectiveness. *J. Water Resour. Plan. Manag* **2001**, *127* (3), 144−149.

(21) Askarizadeh, A.; Rippy, M. A.; Fletcher, T. D.; Feldman, D. L.; Peng, J.; Bowler, P.; Mehring, A. S.; Winfrey, B. K.; Vrugt, J. A.; AghaKouchak, A.; et al. From Rain Tanks to Catchments: Use of Low-Impact Development To Address Hydrologic Symptoms of the Urban Stream Syndrome. *Environ. Sci. Technol.* **2015**, *49* (19), 11264−11280.

(22) Hunt, W. F.; Davis, A. P.; Traver, R. G. Meeting Hydrologic and Water Quality Goals through Targeted Bioretention Design. *J. Environ. Eng.* **2012**, *138* (6), 698−707.

(23) Keeley, M.; Koburger, A.; Dolowitz, D. P.; Medearis, D.; Nickel, D.; Shuster, W. Perspectives on the Use of Green Infrastructure for Stormwater Management in Cleveland and Milwaukee. *Environ. Manage.* **2013**, *51* (6), 1093−1108.

(24) Barbosa, A. E.; Fernandes, J. N.; David, L. M. Key Issues for Sustainable Urban Stormwater Management. *Water Res.* **2012**, *46* (20), 6787−6798.

(25) Petrucci, G.; Rioust, E.; Deroubaix, J.-F.; Tassin, B. Do Stormwater Source Control Policies Deliver the Right Hydrologic Outcomes? *J. Hydrol.* **2013**, *485*, 188−200.

(26) Atzori, L.; Iera, A.; Morabito, G. The Internet of Things: A Survey. *Comput. Networks* **2010**, *54* (15), 2787−2805.

(27) Pellerin, B. A.; Stauffer, B. A.; Young, D. A.; Sullivan, D. J.; Bricker, S. B.; Walbridge, M. R.; Clyde, G. A.; Shaw, D. M. Emerging Tools for Continuous Nutrient Monitoring Networks: Sensors Advancing Science and Water Resources Protection. *J. Am. Water Resour. Assoc.* **2016**, n/a−n/a.

(28) Crawford, J. T.; Loken, L. C.; Casson, N. J.; Smith, C.; Stone, A. G.; Winslow, L. A. High-Speed Limnology: Using Advanced Sensors to Investigate Spatial Variability in Biogeochemistry and Hydrology. *Environ. Sci. Technol.* **2015**, *49* (1), 442−450.

(29) Dienhart, A.; Erickson, K.; Wennen, C.; Henjum, M.; Hozalski, R.; Novak, P.; Arnold, W.; Potter, K. W.; Frevert, D. K. In Situ Sensors for Measuring Pollutant Loads in Urban Streams and Evaluating Stormwater BMP Performance. In *Innovations in Watershed Management under Land Use and Climate Change. Proceedings of the 2010 Watershed Management Conference, Madison, Wisconsin, USA, 23−27 August 2010.*; American Society of Civil Engineers (ASCE), 2010; pp 870−879.

(30) Regional sewer district chooses costly tunnels over "green" infrastructure, though vacant land abounds in Cleveland http://www.cleveland.com/drain/index.ssf/2014/03/regional_sewer_district_plans.html (accessed January 1, 2015).

(31) Ridgeway, K.; Rabbaig, M. A Dam Good Idea. *Water Environment Federation Magazine*. July **2007**.

(32) Open Storm Consortium http://open-storm.org (accessed May 10, 2016).

(33) Katebi, R.; Johnson, M. A.; Wilkie, J. *Control and Instrumentation for Wastewater Treatment Plants*; Springer Science & Business Media, 2012.

(34) Gaborit, E.; Muschalla, D.; Vallet, B.; Vanrolleghem, P. A.; Anctil, F. Improving the Performance of Stormwater Detention Basins by Real-Time Control Using Rainfall Forecasts. *Urban Water J.* **2013**, *10* (4), 230−246.

(35) Middleton, J. R.; Barrett, M. E. Water Quality Performance of a Batch-Type Stormwater Detention Basin. *Water Environ. Res.* **2008**, *80* (2), 172−178.

(36) Jacopin, C.; Lucas, E.; Desbordes, M.; Bourgogne, P. Optimisation of Operational Management Practices for the Detention Basins. *Water Sci. Technol.* **2001**, 277−285.

(37) Carpenter, J. F.; Vallet, B.; Pelletier, G.; Lessard, P.; Vanrolleghem, P. A. Pollutant Removal Efficiency of a Retrofitted Stormwater Detention Pond. *Water Qual. Res. J. Can.* **2014**, *49* (2), 124.

(38) Gaborit, E.; Muschalla, D.; Vallet, B.; Vanrolleghem, P. A.; Anctil, F. Improving the Performance of Stormwater Detention Basins by Real-Time Control Using Rainfall Forecasts. *Urban Water J.* **2013**, *10* (4), 230−246.

(39) Muschalla, D.; Vallet, B.; Anctil, F.; Lessard, P.; Pelletier, G.; Vanrolleghem, P. A. Ecohydraulic-Driven Real-Time Control of Stormwater Basins. *J. Hydrol.* **2014**, *511*, 82−91.

(40) Poresky, A.; Boyle, R.; Cadwalader, O. Taking Stormwater Real Time Controls to the Watershed Scale: Evaluating the Business Case and Developing an Implementation Roadmap for an Oregon MS4. In *Proceedings of the California Stormwater Quality Asssociation*; 2015.

DOI: 10.1021/acs.est.5b05870
Environ. Sci. Technol. 2016, 50, 7267−7273

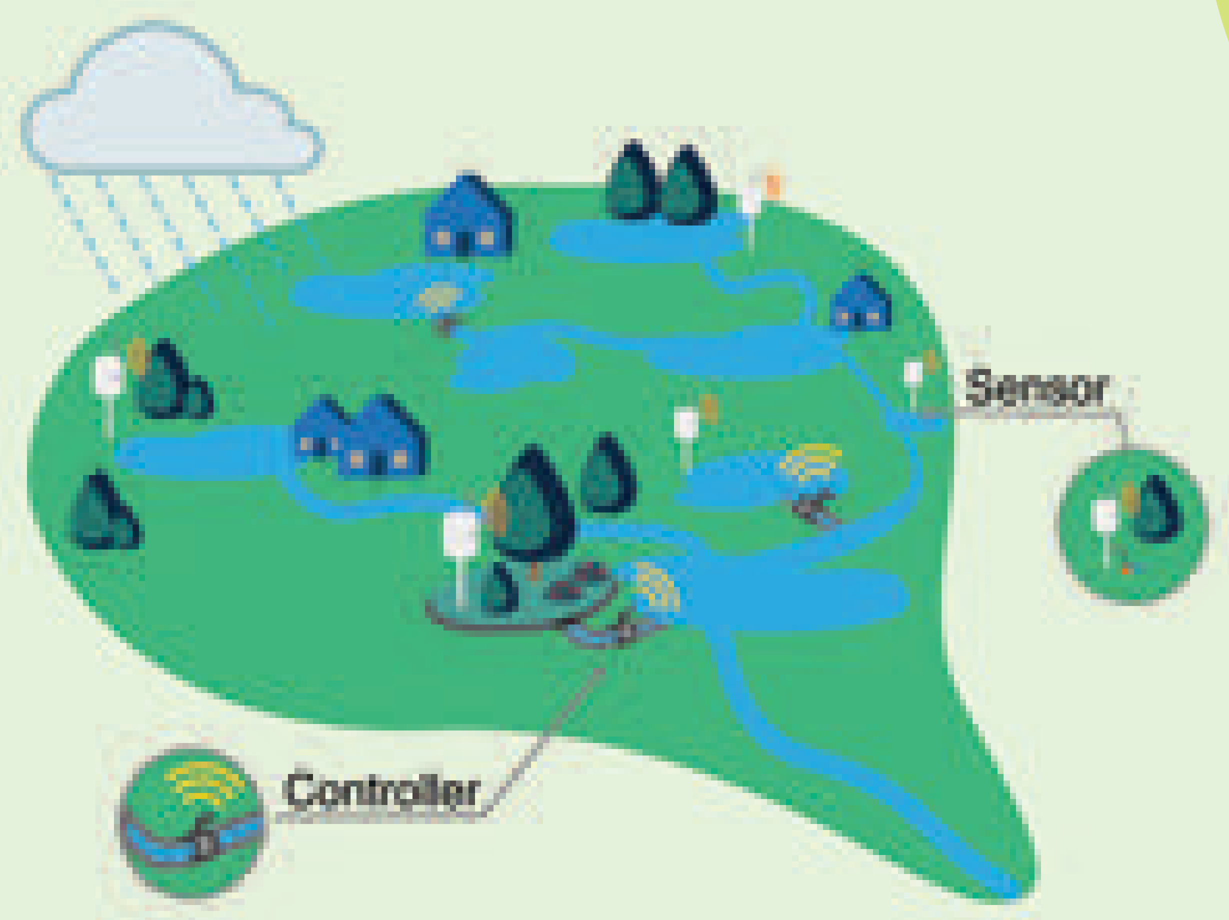
(41) Gilpin, A.; Barrett, M. Interim Report on the Retrofit of an Existing Flood Control Facility to Improve Pollutant Removal in an Urban Watershed. In *World Environmental and Water Resources Congress 2014*; ASCE, 2014; pp 65−74.
(42) Poresky, A.; Boyle, R.; Cadwalader, O. Piloting Real Time Control Retrofits of Stormwater Facilities: Two Oregon Case Studies and Beyond. In *Proceedings of the Pacific Northwest Clean Water Association*, 2015.
(43) Montestruque, L.; Lemmon, M. D. Globally Coordinated Distributed Storm Water Management System. In *Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks - CySWater'15*; ACM Press: New York, 2015; pp 1−6.
(44) *Encyclopedia of Machine Learning*; Sammut, C., Webb, G. I., Eds.; Springer US: Boston, MA, 2010.
(45) Emerson, C. H.; Welty, C.; Traver, R. G. Watershed-Scale Evaluation of a System of Storm Water Detention Basins. *J. Hydrol. Eng.* **2005**, *10* (3), 237−242.
(46) Davis, A. P.; Hunt, W. F.; Traver, R. G.; Clar, M. Bioretention Technology: Overview of Current Practice and Future Needs. *J. Environ. Eng.* **2009**, *135* (3), 109−117.
(47) Hathaway, J. M.; Hunt, W. F.; Jadlocki, S. Indicator Bacteria Removal in Storm-Water Best Management Practices in Charlotte, North Carolina. *J. Environ. Eng.* **2009**, *135*, 1275.
(48) Makropoulos, C. K.; Butler, D. Distributed Water Infrastructure for Sustainable Communities. *Water Resour. Manag.* **2010**, *24* (11), 2795−2816.

DOI: 10.1021/acs.est.5b05870
Environ. Sci. Technol. 2016, 50, 7267−7273

Volume 3 Number 1 January 2017 Pages 1–178

Environmental Science Water Research & Technology

rsc.li/es-water



ISSN 2053-1400

ROYAL SOCIETY OF CHEMISTRY

PAPER
Branko Kerkez *et al.*
Emerging investigators series: building a theory for smart stormwater systems

Environmental Science
Water Research & Technology

PAPER

View Article Online
View Journal | View Issue

Cite this: *Environ. Sci.: Water Res. Technol.*, 2017, **3**, 66

Emerging investigators series: building a theory for smart stormwater systems†

Abhiram Mullapudi, Brandon P. Wong and Branko Kerkez*

Received 15th August 2016,
Accepted 7th October 2016

DOI: 10.1039/c6ew00211k

rsc.li/es-water

Retrofitting stormwater systems with sensors and controllers will allow cities to be operated as real-time, distributed treatment plants. Unlike static infrastructure, which cannot adapt its operation to individual storms or changing land uses, "smart" stormwater systems will use system-level coordination to maximize watershed pollutant removal and treatment. We illustrate that this vision is not limited by technology, which has matured to the point at which it can be ubiquitously deployed. Rather, the challenge is much more fundamental and rooted in a system-level understanding of environmental science. Once distributed stormwater systems become highly instrumented and controlled, how should they be operated to achieve desired watershed outcomes? The answer to this question demands the development of a theoretical framework for smart stormwater systems. In this paper, we lay out the requirements for such a theory. Acknowledging that the adoption of these systems may still be years away, we also present a modeling framework to allow for the simulation of controlled stormwater systems before they become commonplace. We apply this control framework to two simulated case studies in which stormwater sites are controlled to reduce nitrate loads to downstream water bodies.

Water impact

"Smart" stormwater systems will transform cities into coordinated and real-time controlled treatment plants. Retrofitting existing stormwater elements with sensors and controllers will allow them to change their configuration to maximize watershed-scale pollutant removal. We discuss that fundamental knowledge gaps must be addressed before these systems become a reality and present a simulation framework to model real-time control of urban stormwater.

Rapid advances in sensing, computation, and wireless communication are promising to merge the physical with the virtual. Calls to build the "smart" city of the future are being embraced by decision makers. While the onset of self-driving cars provides a good example that this vision is becoming a reality, the role of information technology in the water sector has yet to be fleshed out. These technologies stand to enable a leap in innovation in the distributed treatment of urban runoff, one of our largest environmental challenges.

Retrofitting stormwater systems with sensors and controllers will allow cities to be controlled in real time as distributed treatment plants. Unlike static infrastructure, which cannot adapt its operation to individual storms or changing land uses, "smart" stormwater systems will use system-level coordination to reduce flooding and maximize watershed pollutant removal. Given the sheer number of stormwater control measures in the United States, even a small improvement in their performance could lead to a substantial reduction in pollutant loads. Intriguingly, such a vision is not limited by technology, which has matured to the point at which it can be ubiquitously deployed. Rather, the challenge is much more fundamental and rooted in a system-level understanding of environmental science. Once stormwater systems become highly instrumented and controlled, how should they actually be operated to achieve desired watershed outcomes? The answer to this question demands the development of a theoretical framework for smart stormwater systems. In this paper, we lay out the requirements for such a theory. Acknowledging that the broad adoption of these systems may still be years away, we also present and evaluate a modeling framework to allow for the simulation of smart stormwater systems before they become commonplace.

Recent urban floods,[1] many of which are driven by flashy events and inadequately sized infrastructure, are an all too common example that aging stormwater infrastructure is struggling to keep pace with a dynamic and changing climate. While flood control often emerges as one of the most promising application areas, to illustrate the flexibility of

Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA. E-mail: bkerkez@umich.edu; Tel: +1 (734) 647 0727

† Electronic supplementary information (ESI) available: Models and data are available at https://github.com/kLabUM/control-sim-es-wrt. See DOI: 10.1039/c6ew00211k

smart stormwater systems, this paper will focus on the impacts on urban water quality.

1 Do best local practices achieve the best global outcomes?

Pollutants in runoff are threatening the health of downstream ecosystems, as evinced by harmful algal blooms, such as those on Lake Erie[2] and Chesapeake Bay.[3,4] Simultaneously, "dry" regions of the country are struggling to find new and clean sources of water. By some estimates, the capture of stormwater in Los Angeles[5] and San Francisco[6] could offset the water used by these cities. This, however, requires at least some level of treatment to ensure that captured stormwater is safe for direct use or aquifer injection. In the face of these challenges, novel solutions for stormwater management are needed.

Reductions in hydraulic or pollutant loads are commonly achieved *via* a set of distributed stormwater solutions,[7,8] such as ponds or treatment wetlands. Our body of knowledge on the treatment potential of these systems is extensive, showing that significant water quality and hydraulic benefits can be achieved at the level of individual sites.[9,10] Most recently, an exciting and growing research area has formed around smaller-scale and more distributed Green Infrastructure (GI) solutions, such as green roofs or bioswales.[11] Most of these solutions are grouped under the broader umbrella of Best Management Practices[12] (BMPs) or Stormwater Control Measures (SCMs).[13]

Given the aggressive adoption of these stormwater practices, rarely is the question asked: does doing the "best" at a local scale translate to doing the best at the watershed scale? Research on this question is limited,[14–16] but paints a cautionary picture. Unless designed as part of a coordinated, city-scale solution, a system of SCMs may actually worsen watershed-scale outcomes. For example, unless coordinated at design-time, hydrographs from individual SCMs may add up to cause larger downstream flows compared to having no SCMs at all.[17] This, in turn, can lead to increased stream erosion and re-suspension of sediment-bound pollutants. More examples can be given, but there is an urgent need to investigate the scalability of SCMs and to ensure that their functionality is tuned in the context of broader stormwater systems.

Even if system-level optimization is used to determine the placement of SCMs,[18,19] it is difficult to guarantee that the overall system will perform as designed. The sheer variability in rainfall,[20] seasonal pollutant loadings,[21] and broader land use changes[22] will always push stormwater systems beyond their intended design for the "average" storm.[23] As such, it becomes imperative to find a way to adapt to these uncertain disturbances. One solution relies on real-time sensing and control. By equipping stormwater elements with control valves, which can be operated in real time based on sensor readings, the overall performance of the entire system can be adapted to achieve watershed-scale benefits (Fig. 1a).

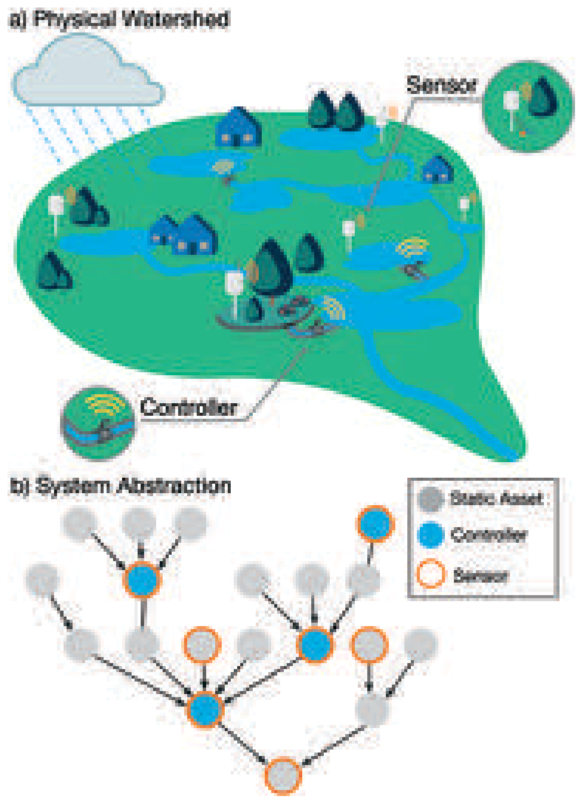


Fig. 1 Application of control and optimization methods to the real-time operation of stormwater systems will be made possible by abstracting physical models (part a) into system-theoretical representations (part b).

1.1 Existing studies on real-time control

The bulk of existing literature on real-time control of stormwater SCMs focuses on water quality and hydraulic impacts at individual sites, particularly ponds and basins. These studies assume that the outlet of a BMP has been retrofitted with a remotely controllable gate or valve. By strategically controlling outflows before or during storm events, internal volumes can be modified and hydraulic retention time (HRT) can be increased. Jacopin *et al.*[24] demonstrated that detention basins, typically designed for flood control, can reduce sediment-based pollutant loading (57% decrease) in downstream water bodies by simply opening and closing a valve. Middleton *et al.*[25] analyzed the water quality response of a controlled detention basin, observing up to a 90% improvement in TSS and ammonia–nitrate removal. Recent studies[26–28] in Quebec, Canada proposed a rule-based control logic for a pond, based on rainfall forecasts, to maximize retention time and reduce hydraulic shocks to the downstream water bodies. These studies reported a 90% improvement in TSS retention. A comprehensive review of these and other studies is summarized by Kerkez *et al.*,[29] along with additional information on how these solutions are deployed in the field. While these studies demonstrate significant potential to improve water quality at the scale of individual sites, the mechanisms behind the removal of pollutants in controlled SCMs remain a research challenge. This is particularly true in the removal of dissolved pollutants, such as ammonia and nitrate. Furthermore, the scalability of real-time control must be evaluated to ensure that local benefits do not overshadow watershed-scale benefits.

Since the 2000 European Union's Water Framework Directive,[30] there has been an increasing emphasis on integrated, system-level control of sewer water distribution systems. The resulting control strategies vary in complexity[31–33] and have since been implemented in a number of urban water networks.[34] Applying these methods to distributed stormwater solutions introduces a new set of challenges. However, unlike in well-maintained sewer networks, the exposed and distributed nature of stormwater systems introduces complexities associated with the urban hydrologic cycle, such as infiltration, evaporation, soil moisture and groundwater dynamics. Furthermore, one major function of stormwater systems relates to the distributed control of a large variety of solid, dissolved and emerging pollutants. Control of sewer networks is often targeted at volume control to mitigate sewer overflows or overloading treatment plants. As such, much work remains to be conducted on investigating how these methods can be applied to the distributed control of stormwater.

2 Toward a framework for smart stormwater systems

Many methods have been developed by the operational research and control theory communities to optimize the operation of networked systems.[35,36] Given their inherent nonlinearity and complexity, existing stormwater models are not compatible with these tools. To that end, our knowledge of treatment processes and the physical nature of stormwater systems must first be embedded in a system-theoretical framework (Fig. 1b). Such a formal and mathematical approach will be crucial toward developing a system-level understanding of stormwater. Not only will this framework help to control future stormwater systems, but it will also create a foundation to answer critical questions, such as how many controllers are needed and where should they be placed to achieve the best system-level benefits? Consequently, how many sensors are needed and where should they be placed to help the control system achieve these objectives?

Until sensors and controllers become ubiquitously deployed across stormwater systems, which may take years to accomplish, there is enough domain knowledge embedded in existing models to begin answering these questions through simulation.

2.1 Limitations of existing simulation approaches

Existing stormwater models can be broadly grouped into two categories: those that focus on hydrology (including hydraulics) and those that focus on water quality. The former range across simple routines, such as Muskingum routing[37] and the Rational method,[38] to more complex hydrodynamic models that solve the St. Venant equation, such as popular packages like SWMM[39] and HEC-RAS.[40] The latter, which include models such as HYDRUS[41,42] and FITOVERT,[43] are used to simulate treatment processes within individual sites, such as wetlands and green infrastructure. While some packages support extended features that model both hydrology and water quality, much work needs to be conducted to improve their accuracy.[44] This often forces a trade-off between comprehensively modeling system-level hydrology and local-level treatment.

Pollutant removal in stormwater is a highly complex and dynamic process. The rate at which pollutants undergo transformation is dependent upon the pollutant type and its interaction with a given stormwater element (oxygen concentration, soil type, biomass, settling time, water temperature, *etc.*). Given the complexity of these interactions, popular stormwater models, such as SWMM, MUSIC[45] and SUSTAIN,[46] often approximate pollutant treatment using first-order decay models:[47]

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -kC \tag{1}$$

where the concentration C of a pollutant is assumed to decrease exponentially following a decay coefficient k. While this may be sufficient for approximating the settling dynamics of sediment-bound pollutants, it does not capture the nuanced and complex transformation of dissolved compounds. This often leads to treating the hydraulic retention time (HRT) as the main proxy for water quality.

To that end, a number of approaches have been developed to extend first-order decay models to account for variations in background concentration,[48] temperature,[47] loading rates[49] or mixing conditions.[50,51] A number of process models have also been developed, applying knowledge from treatment plant operations to stormwater.[52] Langergraber *et al.*[53,54] used finite element analysis to model pollutant transformations in subsurface flow wetlands. While these more comprehensive water quality models are highly promising, their ability to simulate system-level treatment remains to be explored.

Given the need to develop a better understanding of the system-level transport and treatment of stormwater, there is a need to couple existing hydraulic and water quality models.

3 Simulating controlled systems

The real-time operation of gates and valves introduces dynamics that impact hydraulics and water quality. To that end, the biggest limitation of existing models is their ability to simulate the system-wide impacts of real-time control. This includes the ability to dynamically route flows based on a variety of desired control actions, as well as the capacity to simulate a variety of pollutant buildup, washoff, and non-steady state treatment dynamics. While models such as SWMM do have some rudimentary control capabilities, the built-in control logic is limited to site-level control (*e.g.* maintaining levels or flow in a pond).[39] Advanced features, such as system-level control, optimization, or the ability to control around external factors (such as weather forecasts), are not yet implemented.[55]

While it would be possible to extend an existing model to capture all these functionalities, the effort would be significant. To that end, we contend that a coupled modeling approach[56] will be the most flexible way to accomplish this. By coupling models, rather than translating their features into one large model, it becomes possible to construct a modeling chain whose complexity varies based on the scientific or management question that needs to be answered. More importantly, if individual models undergo updates by their respective domain experts, these new features would become available to the coupled model as well without much implementation overhead.

In our coupled modeling approach (Fig. 2), each element in the broader stormwater system can be represented as a storage node, which receives inflows q_1, q_2,…, q_n from upstream nodes, each of which has a corresponding concentration c_1, c_2,…, c_n for a pollutant of interest. The node has an outflow q_{out} which, unlike in static hydraulic infrastructure, is governed by a real-time control action u. A treatment potential k governs the removal or transformation of the pollutant based on a number of hydraulic and water quality states.

Given that control actions change the hydraulic behavior, which in turn affects the treatment of the pollutants, it becomes necessary to implement a modeling cycle that couples these processes in an interconnected, step-wise fashion. In our implementation, the hydraulic simulation can be carried out by any number of hydraulic models, ranging from simple hydraulic routing schemes to more complex models such as SWMM or MUSIC. Outputs from the hydraulic model are fed to the water quality model, which, depending on the pollutant of interest, can range from simple first-order process-based methods to more complex finite-element models. Finally, the control module processes the outputs from the hydraulic and water quality models. Based on the objective, which can depend on the states of multiple elements in the overall system, it sets the discharge rate q_{out} by closing or opening the outlet. The benefits of the coupled approach relate to its flexibility since individual elements can be connected together to represent highly complex stormwater networks.

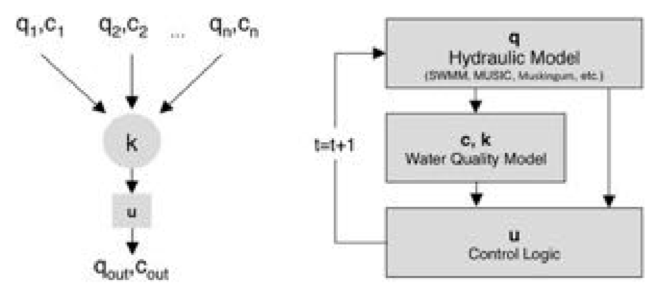


Fig. 2 Each element in the broader stormwater system can be modeled in a step-wise fashion that simulates hydraulic, water quality and control dynamics.

4 Simulated studies

To illustrate the potential benefits that can be achieved through real-time stormwater control, we applied the proposed simulation framework to two simulated case studies, which were inspired by our current research efforts in the midwestern United States. Multiple sites are currently being retrofitted for control and will be compared to these simulations in the coming years. The analysis was targeted on nitrate removal since most of the existing literature focuses on hydraulic control or sediment-bound pollutants.

1. Local scale: the first study investigated the impacts of real-time control on nitrate removal in a single stormwater pond.

2. System scale: the second study evaluated how nitrate removal can be coordinated between a system of controlled stormwater elements.

4.1 Model implementation

Given the scope of the use cases, a simple flow balance module was sufficient to simulate the hydraulic behavior of each element. The change in water volume was modeled as the difference between inflows and outflows, which could be used to calculate the water height h in each element based on its area A. Outflows from each element were proportional to the instantaneous pressure head, unless the element was controlled. Such controlled elements were assumed to be equipped with an outlet structure (*i.e.* butterfly or gate valve) that can be used to regulate outflows such that:

$$0 \le q_{\text{out}} \le \sqrt{2gh} \tag{2}$$

Inflow into upstream elements was based on a hydrograph that was directly measured at one of our study sites in Ann Arbor, Michigan (Fig. 3). For the purpose of this study, this hydrograph can be considered as a synthetic, but realistic, input into the simulation. Overflows were simulated in the case that the storage volume was exceeded. For simplicity, infiltration was assumed to be negligible in the study sites.

A water quality model was developed to simulate nitrogen removal in each stormwater element. While nitrogen removal processes are complex, we can simplify their function for this example by assuming that the removal of nitrogen in stormwater ponds and wetlands occurs through two primary pathways: nitrification (conversion of ammonia to nitrate) and de-nitrification (conversion of nitrate to nitrogen gas).[47,57] Nitrification is an aerobic process (oxygen acts an electron acceptor), while denitrification is anoxic (nitrate acts as an electron acceptor). While denitrification requires sufficient biomass, it is often not limited by this requirement since plants, grass and other sources of carbon are readily present in stormwater ponds and wetlands.[58] As such, oxygen availability becomes a critical factor in nitrogen removal. This can readily be tuned through hydraulic control since retention can be used to create anaerobic conditions.

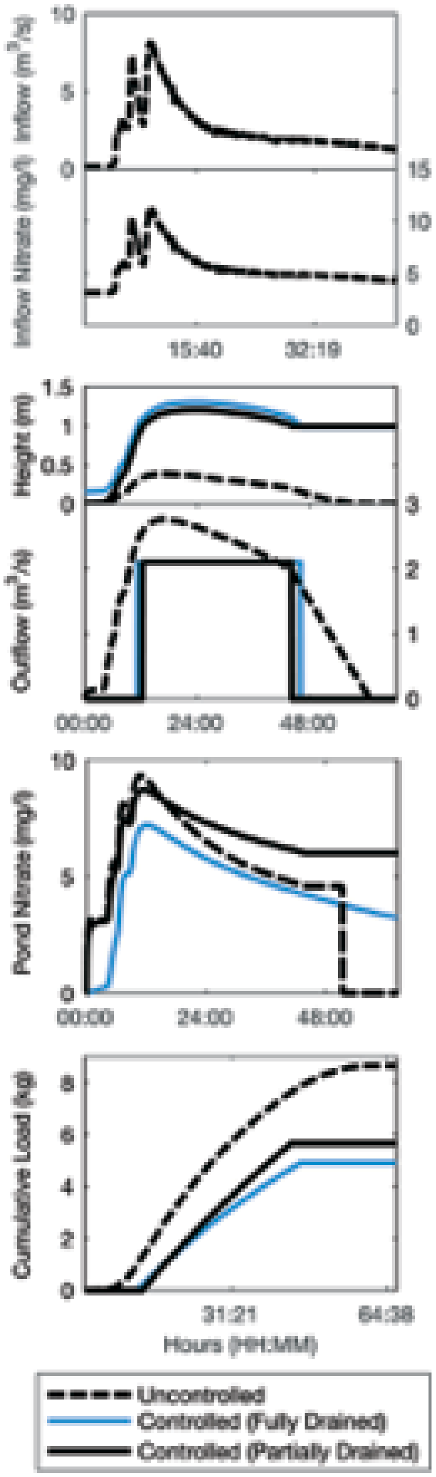


Fig. 3 Impact of real-time control on hydraulic behavior and nitrate treatment, showing inflow concentrations (top panel), pond water height and outflows (second panel), nitrate concentrations inside the pond (third panel), and cumulative nitrate loads exiting the pond (bottom panel).

We constrained our case studies by focusing only on denitrification, assuming that the majority of nitrogen entering our system was in the form of nitrate. While ammonia is present in some stormwater systems, prior measurement of our study sites, as well as other literature studies,[47] indicated a nitrate-dominated runoff. Future studies will investigate the more complex dual-pathway conversion. A synthetic time series for nitrate inflow concentrations was generated to simulate loads to upstream elements. This was achieved by assuming a rough correlation between flow and nitrate (2 mg L^{-1} per m^3 s^{-1}), which was based on prior measurements.[29]

The water treatment for each element was simulated using a continuously stirred tank reactor (CSTR) representation, which is commonly used to simulate similar processes in wastewater treatment plants.[59] Given the dynamic flow conditions that result from real-time control, a closed-form solution that is based on hydraulic residence time does not adequately capture the change in concentration of the pollutant. As such, it becomes necessary to expand it into a complete CSTR mass-balance relation[60–62] to model the concentration C of the dissolved pollutant:

$$\frac{\mathrm{d}c}{\mathrm{d}t}V + \frac{\mathrm{d}v}{\mathrm{d}t}C = q_{\mathrm{in}}C_{\mathrm{in}} - q_{\mathrm{out}}C - kCV \qquad (3)$$

At each time step, the CSTR module communicates with the hydraulic module to update the hydraulic states $\left(\frac{\mathrm{d}V}{\mathrm{d}t}, V, q_{\mathrm{in}} \text{ and } q_{\mathrm{out}}\right)$. The transformation rate k is computed at each time step based on the hydraulic conditions of the stormwater element. Specifically, denitrification can begin once the oxygen concentration at the soil–water interface drops below a minimum threshold (following a first-order decay assumption). Once this occurs, a constant removal rate k is activated. After the element drains, soil is exposed to the air and must be submerged before denitrification can begin again. As such, the model assumes that cumulative denitrification is maximized when the water is in contact with the most anaerobic soil area.

All simulations were implemented in MATLAB Simulink[63] using a fixed time step solver (ode8 Dormand-Prince[64]) at 5 minute intervals. The step-wise coupled modeling approach was implemented by representing each module (hydraulic, water quality, and control) as an individual Simulink object (Fig. 4). All of the source code, inputs and implementation details are attached to this paper as ESI.†

4.2 Case study 1: local control

The first case study is motivated by the objective of controlling a single stormwater basin, which was originally designed for flood remediation as a detention pond (flow-through). The model parameters and physical attributes are provided in the appendix of this paper. In its original configuration, the pond merely serves to attenuate peak flows, with little emphasis on water quality. By equipping this pond with a

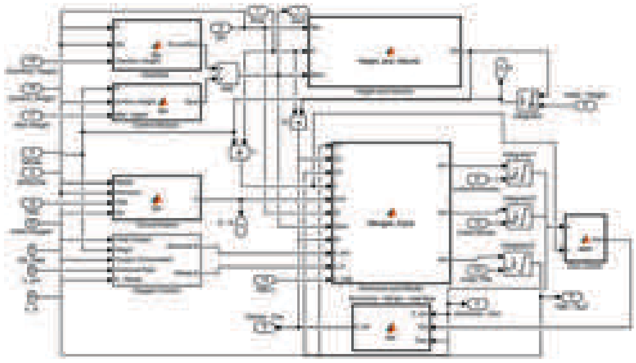


Fig. 4 MATLAB Simulink implementation of the first case study. The overall model functions in a step-wise fashion and couples stand-alone hydraulic, water quality, and control models.

control valve, its original functionality can remain unaffected during large storms by simply keeping the valve open. Major water quality benefits can arise, however, by controlling this pond during smaller and more frequent events.

When enabled, the control algorithm keeps the valve closed and only opens it if the water height exceeds 1.0 m to prevent the pond from overflowing. As a further constraint, when the height exceeds 1.0 m, the valve is modulated to ensure that outflows do not exceed 2 $m^3 s^{-1}$, which is the threshold at which downstream sediments are assumed to be re-suspended. This additional condition ensures that the controller behaves realistically, as the real-world operation would likely have a bound on the outflows to prevent downstream erosion. Two variations of the control algorithm are also evaluated. The first strategy completely drains the pond before a rain event, thus maximizing captured volumes. Based on the magnitude of the rain event (assumed to be known through a weather forecast), the second strategy only partially drains the pond, maximizing the anaerobic conditions at the soil–water interface and thus speeding up denitrification of the inflows. In this case study, the height of the partially drained configuration was set to 0.15 m, assuming that this height would be sufficient to maintain the saturated conditions and prevent the diffusion of oxygen into the soil.[57]

Compared to the uncontrolled scenario, which only attenuated the peak flow, both controlled scenarios retained a water height of 1.0 m after the storm (Fig. 3). Since the pond can be drained at a later time, this volume of water was effectively removed from downstream infrastructure during the storm event. In static stormwater systems, volume reduction strategies are typically only assumed to be possible through upstream infiltration and capture. As such, control may effectively serve as a volume reduction strategy by shifting flows outside of the storm window. Furthermore, outflows for the controlled scenarios resembled a "step", which kept flows below a predetermined erosion threshold. This reduces downstream sediment loads, compared to the uncontrolled scenario, whose outflows spent over 50% of the time exceeding the 2 $m^3 s^{-1}$ erosion threshold.

Nitrate inside the pond and the effluent revealed distinct dynamics between each control configuration. In the uncontrolled scenario, very limited treatment occurred due to short hydraulic retention time. The effluent concentrations peaked before dropping to zero since the pond was drained completely following the storm. The controlled scenarios did not see this drop-off in internal nitrate because the flows were retained for treatment. The partially drained scenario showed lower nitrate concentrations at the beginning of the storm due to higher anaerobic soil area and denitrification potential.

While internal concentrations are an indicator of treatment dynamics inside the pond, perhaps the best measure of treatment capacity is given by the cumulative nitrate load exiting the pond (bottom panel, Fig. 3). The uncontrolled scenario exhibited the largest cumulative nitrate loads since the runoff effectively just flowed through the pond with limited treatment. The controlled pond showed a nearly 45% mass reduction (from 8.6 kg to 4.7 kg) in nitrate due to increased volume capture, HRT and denitrification. The partially drained control strategy did indicate an improved load reduction compared to the fully drained control strategy (14% improvement). This suggests that, rather than simply draining the pond before a storm event, improved load reductions may be achieved through more complex control approaches. More complex control comes at the cost of uncertainty, however. The partially drained controller assumed prior knowledge about inflows to decide how much water to drain before a storm. If these decisions are made around weather forecasts, the uncertainty embedded in the inputs may cause adverse impacts, such as overflows. The anticipated benefits of any control strategy should thus always be weighted against the uncertainty of any inputs.

4.3 Case study 2: system-level control

The second case study evaluated how control strategies may change when a system of multiple stormwater assets is controlled. A system of four elements was considered, consisting of three parallel ponds draining into a constructed wetland (Fig. 5). Two of the upstream ponds were controlled while the treatment wetland and the other pond remained uncontrolled. The objective was to control the upstream ponds to boost the nitrate treatment and reduce the effluent concentrations at the outlet of the wetland. The configuration was based on a real-world site currently being retrofitted for control in southeastern Michigan.

Due to their large biomass area, wetlands have a higher nitrate treatment capacity than ponds.[65] As such, the control objective was to keep the downstream wetland "active", by maximizing its water height and thus the biomass treatment area. While a prolonged inundation may damage the emergent vegetation in the wetland, the proposed control algorithm maximizes the treatment area of the wetland only during the duration of the storm event, which should improve the treatment while only briefly inundating the

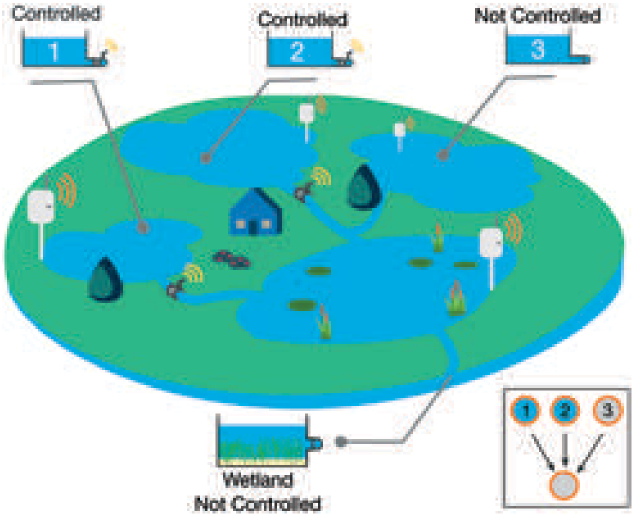


Fig. 5 System-level control case study: three ponds, two of which are controlled, draining into a treatment wetland.

wetland. In the uncontrolled scenario, the flows from the upstream ponds actually added up causing the wetland to overflow (Fig. 6, fourth column), which also impaired treatment.

The controlled scenario (see Appendix for implementation details) balanced the outflows from the two ponds to ensure that the wetland remained filled (2 m – just below its overflow height) as long as the uncontrolled third pond was discharging. Once the third pond was entirely drained, the upstream ponds retained any additional inflows, as long as it would not cause them to overflow. This strategy eliminated downstream overflows while simultaneously increasing the wetland's anaerobic treatment area. As such, flows from the third pond were exposed to a higher degree of denitrification compared to those in the uncontrolled case. Overall, the controlled system achieved a 46.48% (from 17.9 kg to 9.6 kg) reduction in cumulative nitrate loads. While some of this overall reduction was driven by the fact that the two controlled ponds remained filled after the storm, thus retaining some nitrate mass upstream, two major benefits arose compared to the uncontrolled scenario. Firstly, the wetland effluent concentrations were reduced over time, showing a 15.25% reduction in concentration. Secondly, the case study showed that a subset of upstream elements may be controlled to reduce downstream hydraulic loads, which, similar to the first case study, has the potential to reduce erosion.

A natural extension of this control strategy would be the direct control of the wetland. In many real-world situations, however, not all elements of the system will be controllable. In these instances, system-level benefits may still be achieved *via* control of other elements. The purpose of this case study was to illustrate one possible example focused on system-level nutrient control. While simple, this control strategy was nonetheless effective at improving the hydraulic and water quality behavior of the overall system. More complex control strategies will be evaluated in the future, especially in the context of larger and more heterogeneous stormwater systems.

5 Discussion

Current state-of-the-art stormwater solutions are still primarily focused on static (non-controlled) solutions. As such, our analysis compared real-time control to static solutions, which were designed in accordance to modern engineering practices. Sensor-driven, real-time control of stormwater presents an exciting new paradigm and research area. It is presently unclear, however, how results generated by existing research, as well as the case studies presented in this paper, can be scaled to large watersheds. Many existing studies focus solely on the control of individual elements and, specifically, on sediment reduction or flood remediation. While the case studies in this paper took a step toward simulating the removal of more complex dissolved pollutants in a multi-element system, it is important to note that the control logic was uniquely tailored to one specific storm and study area. The efficacy of the controls in our case studies was reliant on the ability to hold water after a storm to allow for extended treatment. This strategy may be impacted by limits on hydraulic retention time. Modifying the water levels and residence times may introduce issues related to aesthetics, plant survival and mosquito breeding.[66] Thus, the potential benefits to water flow and quality must be studied as part of a multi-objective optimization problem. Much of the real world is underpinned by significant uncertainty, especially related to weather forecasts. Since these forecasts determine when and how much water needs to be released, the stochastic nature of weather must be taken into consideration when controlling such systems.

Control strategies may also change entirely if the removal of different pollutants is required. A simple example can be given by watersheds in which runoff is dominated by ammonia rather than nitrate, thus requiring stages of both nitrification and denitrification. The intricacy of control strategies will likely increase with the number of objectives[67] and the complexity of runoff dynamics. This introduces the exciting paradigm of controlling the overall system to create treatment chains in which individual elements are tuned to achieve specific objectives. By tuning the hydraulic behavior of each element, there will be an unprecedented opportunity to begin applying process-based knowledge from wastewater treatment to distributed stormwater modeling. The modeling of such complete control approaches will be made easier by the simulation approach proposed in Fig. 2, which will allow for coupling of knowledge spanning hydrology, hydraulics, and water quality.

6 Knowledge gaps

While research is needed to improve our fundamental understanding and modeling of system-level stormwater, two major knowledge gaps become evident when we view stormwater

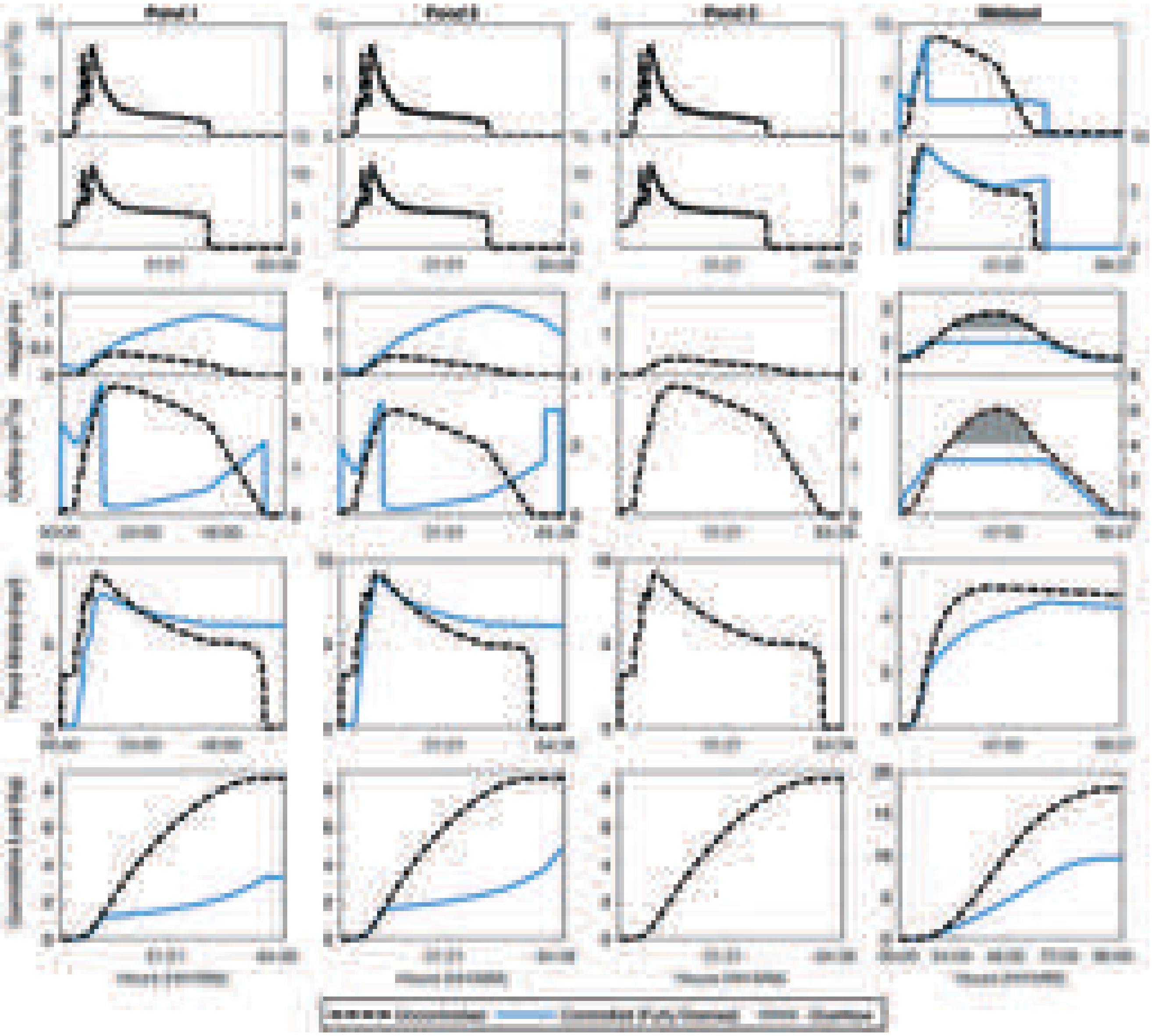


Fig. 6 Impact of real-time control on hydraulics and nitrate treatment across a system of stormwater elements: inflow concentrations (top row), pond water height and outflows (second row), nitrate concentrations inside each element (third row), and cumulative nitrate loads exiting the pond (bottom row).

control in a system-theoretical framework. This can be accomplished by visualizing it as a *feedback loop* (Fig. 7), a technique common in the control communities and dynamical systems theory.[68] This loop estimates the difference between a desired watershed outcome (downstream nitrate concentrations, for example) and the actual watershed outcome and *feeds* it into the control logic to drive the system toward the desired outcome. The physical requirements of this feedback loop, which include sensors, controllers and the physical infrastructure, already exist or have matured to the point at which they do not present major research challenges. Rather, our biggest knowledge gaps span the *virtual* components of the feedback loop and include the (1) assimilation of noisy, sparse, and heterogeneous sensor data into real-time models (state estimation), and (2) the automated synthesis of control logic in response to these estimates.

6.1 Toward a new generation of real-time models

Unlike in static infrastructure systems, where adaptation strategies take place on monthly or yearly time scales, real-time control reduces adaptation to minutes or seconds. Existing stormwater models have not been designed to interface with real-time data. Rather, sensor data is often used merely as a convenience to parameterize the model. It is not uncommon for these predictions to drift away from real-world conditions over the modeling horizon. Given the need to base control actions on the best sources of information, a

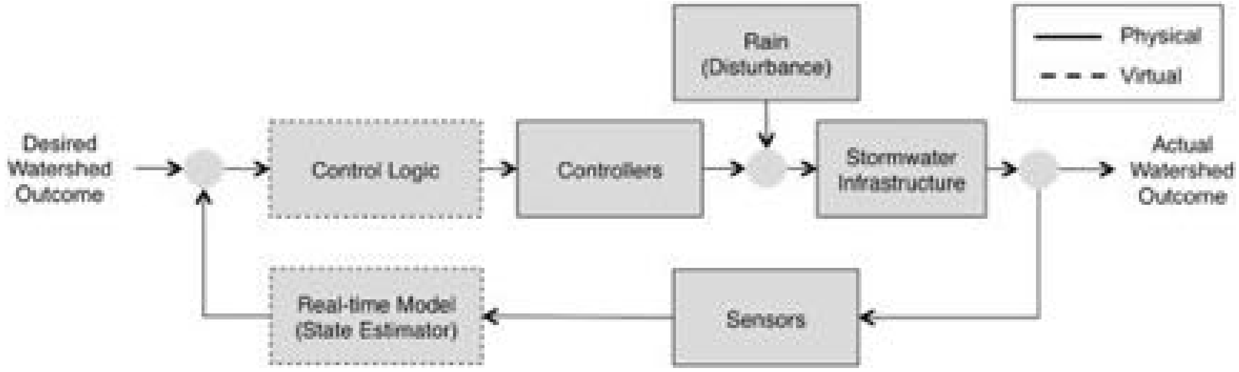


Fig. 7 The stormwater *feedback control loop*. A desired watershed outcome is compared, in real time, to an actual watershed state based on sensor measurements. The control logic then adjusts the states of valves, gates and pumps to drive the system toward the desired state. Disturbances, such as precipitation, may drive the system away from the desired outcome and must be controlled against when the feedback loop repeats.

new generation of data-driven and real-time models must be developed. Rather than executing unchecked into the future, they will "learn" from the data and update their states to reflect changing field conditions. Such models will need to be self-calibrating, robust to uncertainty, and computationally efficient to execute in the amount of time required to make control decisions. This raises the question: how complex does a system model need to be to enable an effective and robust control loop? While the answer to this question remains to be investigated, many other control applications (aircraft autopilots, for example) suggest that it is very likely that stormwater control models will not need to be as complex as the models currently used for simulation. This does not mean that existing physically driven models or our proposed simulation framework (Fig. 2) will not be needed. In fact, existing simulation approaches will be critical in the planning and design of control systems, while real-time models will be used for the actual control.

In our case studies, an assumption was made that control actions were informed by known *in situ* conditions, such as water flows, pond levels, and nitrate concentrations. This will be far from true in many real-world control systems, where sensors will be sparsely placed and noisy. New models will thus have to be developed to make predictions at locations that are uninstrumented and for parameters that are unmeasured. By quantifying the uncertainty inherent in such models, it will also be possible to develop sensor placement algorithms to determine how many sensors are required and where they should be placed to improve real-time model performance. Many of the methods required for these tasks already exist in other communities (system identification, data assimilation, machine learning, *etc.*), but their application to stormwater systems remains to be investigated.

6.2 Control algorithms

While there is potential to apply more advanced control algorithms to our case studies, the application of complex control logic to stormwater presents an open area of research. It is unclear which real-time control and optimization techniques will be the most robust and suitable for distributed stormwater systems. Most current studies, as well as the case studies presented in this paper, have been built around simple rule-based control (*e.g.* drain a pond before a storm). While such control approaches preserve intuition and incorporate operator expertise, they do not scale for systems of arbitrary sizes. This impedes the ability to transfer lessons from one watershed to another. The complexity of operational rules will increase drastically with the size of watersheds or extended control objectives. The logic associated with operating a network of distributed stormwater assets, consisting of hundreds or thousands of controllers, will become overwhelming unless formal mathematical methods are developed to abstract the physical stormwater dynamics into a system-theoretical framework. These mathematical underpinnings will finally allow for performance or safety guarantees to be provided. This, in turn, will enable new methods to determine how many controllers are needed and where they should be placed to ensure that desired watershed outcomes are met.

7 Conclusions

The goal of this paper was to illustrate the need for a "smart" stormwater systems theoretical framework. Before such systems become adopted, much work remains to be conducted on simulating their performance, which can be accomplished by coupling existing hydrologic, hydraulic and water quality models. As demonstrated by our case studies, real-time control of stormwater has the potential to significantly improve the performance of existing infrastructure, introducing new alternatives to tightly manage nutrients, metals and other pollutants in urban watersheds. Considering the current funding mechanisms for stormwater, especially in the United States, the cost of retrofitting will provide a more budget-conscious alternative to new construction while achieving similar or better water quality outcomes. Aside from technical or research gaps, which must be addressed before these systems become a reality, it will be imperative to encourage a broad community of researchers, engineers, and cities to adopt these technologies as part of their existing toolboxes. To that end, our team has been spearheading the open-storm.org portal, a collaborative and open-source initiative aimed at sharing end-to-end blueprints and tutorials on software, hardware and sensors required to instrument and control urban watersheds. Updates (photos, videos, results, *etc.* open-storm.org will track and disseminate its future work.

8 Appendix

Parameterization of models used for the case studies. The MATLAB Simulink models and data used for generating the plots are available at https://github.com/kLabUM/control-sim-es-wrt.

8.1 Case study 1: local control

- Area: 500 m^2
  - Max height: 1.5 m
  - K_{Nitrate} = 42.048 per year
  - K_{Oxygen} = 31.536 per year

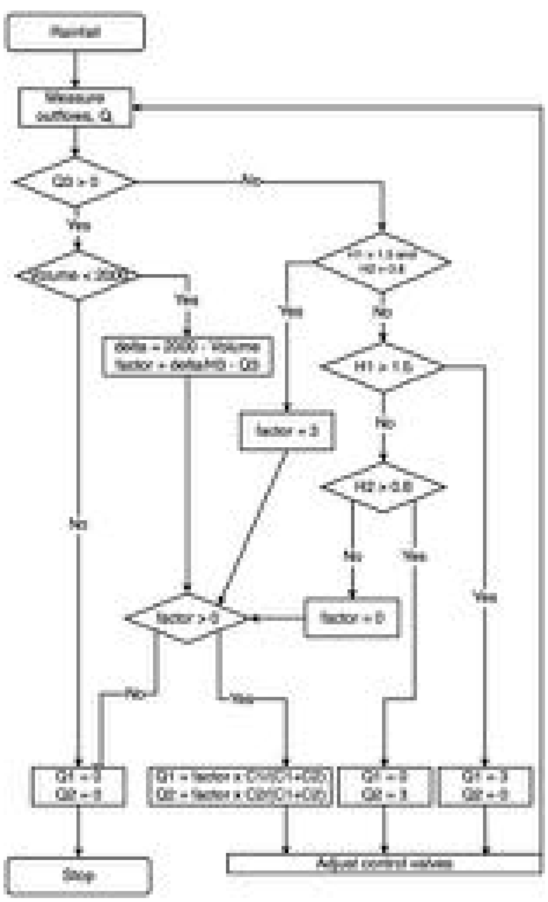


Fig. 8 Rule-based algorithm used for controlling the system in the second case study.

8.2 Case study 2: system-level control (Fig. 8)

- Pond 1
  - Area: 1000 m^2
  - Max height: 2.5 m
  - K_{Nitrate} = 21.024 per year
  - K_{Oxygen} = 525.60 per year
- Pond 2
  - Area: 600 m^2
  - Max height: 2.5 m
  - K_{Nitrate} = 21.024 per year
  - K_{Oxygen} = 1051.2 per year
- Pond 3
  - Area: 1000 m^2
  - Max height: 2.5 m
  - K_{Nitrate} = 15.768 per year
  - K_{Oxygen} = 1051.2 per year
- Wetland
  - Area: 1000 m^2
  - Max height: 2.4 m
  - Weir height: 1.5 m
  - K_{Nitrate} = 25.228 per year
  - K_{Oxygen} = 1051.2 per year

Acknowledgements

The authors acknowledge the financial support of the Great Lakes Protection Fund and the University of Michigan.

References

1 D. Frosch and C. McWhirter, *Houstons Rapid Growth, Heavy Rains, Heightend Flood Risk*, 2016.
2 A. M. Michalak, E. J. Anderson, D. Beletsky, S. Boland, N. S. Bosch, T. B. Bridgeman, J. D. Chaffin, K. Cho, R. Confesor, I. Daloglu, J. V. DePinto, M. A. Evans, G. L. Fahnenstiel, L. He, J. C. Ho, L. Jenkins, T. H. Johengen, K. C. Kuo, E. LaPorte, X. Liu, M. R. McWilliams, M. R. Moore, D. J. Posselt, R. P. Richards, D. Scavia, A. L. Steiner, E. Verhamme, D. M. Wright and M. A. Zagorski, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 6448–6452.
3 F. Powledge, *BioScience*, 2005, **55**, 1032–1038.
4 D. F. Boesch, R. B. Brinsfield and R. E. Magnien, *J. Environ. Qual.*, 2001, **30**, 303–320.
5 Geosyntec, Stormwater Capture Master Plan Interim Report, Geosyntec Technical Report December, 2014.
6 N. Garrison, *Stormwater Capture Potential in Urban and Suburban California*, Natural resources defense council technical report, 2014.
7 P. Hamel, E. Daly and T. D. Fletcher, *J. Hydrol.*, 2013, **485**, 201–211.
8 M. J. Burns, T. D. Fletcher, C. J. Walsh, A. R. Ladson and B. E. Hatt, *Landsc. Urban Plan.*, 2012, **105**, 230–240.
9 D. W. Stanley, *Water Environ. Res.*, 1996, **68**, 1076–1083.
10 J. N. Carleton, T. J. Grizzard, A. N. Godrej, H. E. Post, L. Lampe and P. P. Kenel, *Water Environ. Res.*, 2000, 72, 295–304.

11 M. A. Benedict, T. Edward and J. D. Mcmahon, in *Green Infrastructure: A Strategic Approach to Land Conservation*, The Conservation Fund, 2006, ch. 1, p. 5.
12 B. Urbonas, *Water Sci. Technol.*, 1994, **29**, 347–353.
13 A. R. Cizek and W. F. Hunt, *Ecol. Eng.*, 2013, **57**, 40–45.
14 G. Petrucci, E. Rioust, J.-F. Deroubaix and B. Tassin, *J. Hydrol.*, 2013, **485**, 188–200.
15 J. Sage, E. Berthier and M. C. Gromaire, *Environ. Manage.*, 2015, **56**, 66–80.
16 G. Petrucci, J.-F. Deroubaix and B. Tassin, *CWRS 2014: Evolving Water Resources Systems: Understanding, Predicting and Managing Water-Society Interactions*, 2014, vol. 364, pp. 1–7.
17 C. H. Emerson, M. Asce, C. Welty and R. G. Traver, *J. Hydrol. Eng.*, 2005, **10**, 237–242.
18 S.-K. Ciou, J.-T. Kuo, P.-H. Hsieh and G.-H. Yu, Optimization Model for BMP Placement in a Reservoir Watershed, *J. Irrig. Drain. Eng.*, 2012, **138**(8), 736–747.
19 X.-Y. J. Zhen, S. L. Yu and J.-Y. Lin, *J. Water Resour. Plan. Manage.*, 2004, **130**, 339–347.
20 I. Chaubey, C. Haan, S. Grunwald and J. Salisbury, *J. Hydrol.*, 1999, **220**, 48–61.
21 Y. Ouyang, P. Nkedi-Kizza, Q. Wu, D. Shinde and C. Huang, *Water Res.*, 2006, **40**, 3800–3810.
22 A. Goonetilleke, E. Thomas, S. Ginn and D. Gilbert, *J. Environ. Manage.*, 2005, **74**, 31–42.
23 Department of Environmental Protection, Pennsylvania Stormwater Best Management Practices Manual, Bureau of Watershed Management, 2006.
24 C. Jacopin, E. Lucas, M. Desbordes and P. Bourgogne, *Water Sci. Technol.*, 2001, **44**, 277–285.
25 J. R. Middleton and M. E. Barrett, *Water Environ. Res.*, 2008, **80**, 172–178.
26 D. Muschalla, B. Vallet, F. Anctil, P. Lessard, G. Pelletier and P. A. Vanrolleghem, *J. Hydrol.*, 2014, **511**, 82–91.
27 E. Gaborit, D. Muschalla, B. Vallet, P. A. Vanrolleghem and F. Anctil, *Urban Water J.*, 2013, **10**, 230–246.
28 E. Gaborit, F. Anctil, G. Pelletier and P. A. Vanrolleghem, *Urban Water J.*, 2015, 1–11.
29 B. Kerkez, C. Gruden, M. J. Lewis, L. Montestruque, M. Quigley, B. P. Wong, A. Bedig, R. Kertesz, T. Braun, O. Cadwalader, A. Poresky and C. Pak, *Environ. Sci. Technol.*, 2016, **50**, 7267–7273.
30 The European Parliament and the Council of European Union, Official Journal of the European Communities, 2000, 01–73.
31 L. Benedetti, P. Prat, I. Nopens, M. Poch, C. Turon, B. De Baets and J. Comas, *Water Sci. Technol.*, 2009, **60**, 2035.
32 K. Seggelke, R. Löwe, T. Beeneken and L. Fuchs, Implementation of an integrated real-time control system of sewer system and waste water treatment plant in the city of Wilhelmshaven, *Urban Water J.*, 2013, **10**, 330–341.
33 D. Fiorelli, G. Schutz, K. Klepiszewski, M. Regneri and S. Seiffert, Optimised real time operation of a sewer network using a multi-goal objective function, *Urban Water J.*, 2013, **10**, 342–353.
34 A. L. Mollerup, P. S. Mikkelsen, D. Thornberg and G. Sin, Controlling sewer systems – a critical review based on systems in three EU cities, *Urban Water J.*, 2016, DOI: 10.1080/1573062X.2016.1148183.
35 Y. Sheffi, *Urban transportation networks*, Prentice Hall, 1984.
36 K. J. Aström, R. M. Murray and R. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*, Princeton University Press, 2006.
37 G. W. Brunner and J. Gorbrecht, A Muskingum-Cunge Channel Flow Routing Method for Drainage Networks, *ASCE Journal of Hydraulics*, 1991, **117**(5), 629–642.
38 D. A. Chin, A. Mazumdar and K. Roy, *Water-resources engineering*, Prentice Hall Englewood Cliffs, 12th edn, 2000.
39 L. A. Rossman, *Storm Water Management Model User's Manual Version 5.1*, US Environmental Protection Agency, 2010.
40 G. Brunner and CEIWR-HEC, *HEC -RAS River Analysis System User Manual*, US Army Corps of Engineers, Institute for Water Resources, 5th edn, 2016.
41 A. Rizzo, G. Langergraber, A. Galvão, F. Boano, R. Revelli and L. Ridolfi, *Ecol. Eng.*, 2014, **68**, 209–213.
42 T. Pálfy and G. Langergraber, *Ecol. Eng.*, 2014, **68**, 105–115.
43 D. Giraldi, M. de Michieli Vitturi and R. Iannelli, *Environ. Model. Softw.*, 2010, **25**, 633–640.
44 C. B. S. Dotto, M. Kleidorfer, A. Deletic, T. D. Fletcher, D. T. McCarthy and W. Rauch, *Water Sci. Technol.*, 2010, **62**, 837–843.
45 T. H. F. Wong, T. D. Fletcher, H. P. Duncan, J. R. Coleman and G. A. Jenkins, *Global Solutions for Urban Drainage*, 2002, pp. 1–14.
46 F.-H. Lai, T. Dai, J. Zhen, J. Riverson, K. Alvi and L. Shoemaker, *Proceedings of the Water Environment Federation*, 2007, vol. 5, pp. 946–968.
47 R. H. Kadlec and S. Wallace, *Treatment wetlands*, CRC press, 2008.
48 A. H. L. Shepherd, G. Tchobanoglous and M. E. Grismer, *Water Environ. Res.*, 2001, **73**, 597–606.
49 C. Mitchell and D. McNevin, *Water Res.*, 2001, **35**, 1295–1303.
50 J. Persson and H. B. Wittgren, *Ecol. Eng.*, 2003, **21**, 259–269.
51 T. H. F. Wong, T. D. Fletcher, H. P. Duncan and G. A. Jenkins, *Ecol. Eng.*, 2006, **27**, 58–70.
52 G. Langergraber, *Vadose Zone J.*, 2008, **7**, 830–842.
53 G. Langergraber, D. P. L. Rousseau, J. García and J. Mena, *Water Sci. Technol.*, 2009, **59**, 1687–1697.
54 G. Langergraber and J. Šimůnek, *Vadose Zone J.*, 2005, **4**, 924–938.
55 G. Ria, J. Barreiro-Gomez, A. Ramirez-Jaime, N. Quijano and C. Ocampo-Martinez, *Environ. Model. Softw.*, 2016, **83**, 143–154.
56 J. L. Goodall, B. F. Robinson and A. M. Castronova, *Environ. Model. Softw.*, 2011, **26**, 573–582.
57 K. R. Reddy, W. H. Patrick and C. W. Lindau, *Limnol. Oceanogr.*, 1989, **34**, 1004–1013.
58 J. R. White and K. R. Reddy, *The Wetlands Handbook*, 2009, pp. 213–227.
59 M. Henze, *Activated sludge models ASM1, ASM2, ASM2d and ASM3*, IWA Publishing, 2000.
60 H. Alvord and R. Kadlec, *Ecol. Eng.*, 1996, **90**, 97–107.

61 R. H. Kadlec, *Transformations of Nutrients in Natural and Constructed Wetlands*, 2001, pp. 365–391.
62 R. K. Munson, S. B. Roy, S. A. Gherini, A. L. MacNeill, R. J. M. Hudson and V. L. Blette, *Water, Air, Soil Pollut.*, 2002, **134**, 255–272.
63 The MathWorks Inc., MATLAB.
64 J. Dormand and P. Prince, *J. Comput. Appl. Math.*, 1980, **6**, 19–26.
65 L. Scholes, D. M. Revitt and J. B. Ellis, *J. Environ. Manage.*, 2008, **88**, 467–478.
66 R. L. Knight, W. E. Walton, G. F. O'Meara, W. K. Reisen and R. Wass, *Ecol. Eng.*, 2003, **21**, 211–232.
67 E. D. Tillinghast, W. F. Hunt, G. D. Jennings and P. D'Arconte, *J. Hydrol. Eng.*, 2012, **17**, 1381–1388.
68 K. Ogata, *Modern Control Engineering*, Prentice Hall, 5th edn, 2009, p. 201.

Advances in Water Resources 127 (2019) 167–179

Contents lists available at ScienceDirect

Advances in Water Resources

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

Hydrograph peak-shaving using a graph-theoretic algorithm for placement of hydraulic control structures

Matthew Bartos*, Branko Kerkez

Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA

A R T I C L E I N F O

Keywords:
Control systems
Controller placement
Graph theory
Stormwater control

A B S T R A C T

The need to attenuate hydrograph peaks is central to the design of stormwater and flood control systems. However, few guidelines exist for siting hydraulic control structures such that system-scale benefits are maximized. This study presents a new graph-theoretic algorithm for stabilizing the hydrologic response of watersheds by placing controllers at strategic locations in the drainage network. This algorithm identifies subcatchments that dominate the peak of the hydrograph, and then finds the “cuts” in the drainage network that maximally isolate these subcatchments, thereby flattening the hydrologic response. Evaluating the performance of the algorithm through an ensemble of hydrodynamic simulations, we find that our controller placement algorithm produces consistently flatter discharges than randomized controller configurations—both in terms of the peak discharge and the overall variance of the hydrograph. By attenuating flashy flows, our algorithm provides a powerful methodology for mitigating flash floods, reducing erosion, and protecting aquatic ecosystems. More broadly, we show that controller placement exerts an important influence on the hydrologic response and demonstrate that analysis of drainage network structure can inform more effective stormwater control policies.

1. Introduction

In the wake of rapid urbanization, aging infrastructure and a changing climate, effective stormwater management poses a major challenge for cities worldwide (Kerkez et al., 2016). Flash floods are one of the largest causes of natural disaster deaths in the developed world (Doocy et al., 2013), and often occur when stormwater systems fail to convey runoff from urban areas (Wright and Marchese, 2017). At the same time, many cities suffer from impaired water quality due to inadequate stormwater control (Walsh et al., 2005). Flashy flows erode streambeds, release sediment-bound pollutants, and damage aquatic habitats (Booth and Jackson, 1997; Finkenbine et al., 2000; Walsh et al., 2005; Wang et al., 2001), while untreated runoff may trigger fish kills and toxic algal blooms (Sahagun, 2013; Wines, 2014). Engineers have historically responded to these problems by expanding and upsizing stormwater control infrastructure (Rosenberg et al., 2010). However, larger infrastructure frequently brings adverse side-effects, such as dam-induced disruption of riparian ecosystems (Dam, 2000), and erosive discharges due to overdesigned conveyance infrastructure (Kerkez et al., 2016). As a result, recent work has called for the replacement of traditional peak attenuation infrastructure with targeted solutions that better reduce environmental impacts (Arora et al., 2015; Hawley and Vietz, 2016).

As the drawbacks of oversized stormwater infrastructure become more apparent, many cities are turning towards decentralized stormwater solutions to regulate and treat urban runoff while reducing adverse impacts. Green infrastructure, for instance, uses low-impact rain gardens, bioswales, and green roofs to condition flashy flows and remove contaminants (Askarizadeh et al., 2015; Coffman et al., 1999; Strecker et al., 2000). *Smart* stormwater systems take this idea further by retrofitting static infrastructure with dynamically controlled valves, gates and pumps (Bartos et al., 2018; Kerkez et al., 2016; Mullapudi et al., 2018; 2017). By actuating small, distributed storage basins and conveyance structures in real-time, *smart* stormwater systems can halt combined sewer overflows (Montestruque and Lemmon, 2015), mitigate flooding (Kerkez et al., 2016), and improve water quality at a fraction of the cost of new construction (Bartos et al., 2018; Kerkez et al., 2016). While decentralized stormwater management tools show promise towards mitigating urban water problems, it is currently unclear how these systems can be designed to achieve maximal benefits at the watershed scale. Indeed, some research suggests when stormwater control facilities are not designed in a global context, local best management practices can lead to adverse system-scale outcomes—in some cases inducing downstream flows that are more intense than those produced under unregulated conditions (Emerson et al., 2005; Petrucci et al., 2013).

* Corresponding author.
E-mail addresses: mdbartos@umich.edu (M. Bartos), bkerkez@umich.edu (B. Kerkez).

https://doi.org/10.1016/j.advwatres.2019.03.016
Received 13 September 2018; Received in revised form 26 March 2019; Accepted 27 March 2019
Available online 28 March 2019

Thus, as cities begin to experiment with decentralized stormwater control, the question of *where* to place control structures becomes crucial. While many studies have investigated the ways in which active control can realize system-scale benefits (using techniques like feedback control (Wong and Kerkez, 2018), market-based control (Montestruque and Lemmon, 2015), or model-predictive control, (Gelormino and Ricker, 1994; Mollerup et al., 2016)), the location of control structures within the drainage network may serve an equally important function. Hydrologists have long recognized the role that drainage network topology plays in shaping hydrologic response (Gupta and Mesa, 1988; Gupta et al., 1986; Kirkby, 1976; Mantilla et al., 2011; Marani et al., 1991; Mesa and Mifflin, 1986; Tejedor et al., 2015a,b; Troutman and Karlinger, 1985). It follows that strategic placement of hydraulic control structures can shape the hydrograph to fulfill operational objectives, such as maximally flattening flood waves and regulating erosion downstream. To date, however, little research has been done to assess the problem of optimal placement of hydraulic control structures in drainage networks:

- Recent studies have investigated optimal placement of green infrastructure upgrades like green roofs, rain tanks and bioswales (Meerow and Newell, 2017; Norton et al., 2015; Schilling and Logan, 2008; Schubert et al., 2017; Yao et al., 2015; Zellner et al., 2016; Zhang et al., 2015). However, these studies generally focus on quantifying the potential benefits of green infrastructure projects through representative case studies (Schubert et al., 2017; Yao et al., 2015; Zellner et al., 2016; Zhang et al., 2015), and do not intend to present a generalized framework for placement of stormwater control structures. As a result, many of these studies focus on optimizing multiple objectives (such as urban heat island mitigation (Norton et al., 2015), air quality (Meerow and Newell, 2017), or quality of life considerations (Schilling and Logan, 2008)), or use complex socio-physical models and optimization frameworks (Zellner et al., 2016), making it difficult to draw general conclusions about controller placement in drainage networks.
- Studies of pressurized water distribution networks have investigated the related problems of valve placement (Cattafi et al., 2011; Creaco et al., 2010), sensor placement (Perelman and Ostfeld, 2013), subnetwork vulnerability assessment (Yazdani and Jeffrey, 2011), and network sectorization (Hajebi et al., 2015; Tzatchkov et al., 2008). While these studies provide valuable insights into the ways that complex network theory can inform drinking water infrastructure design, water distribution networks are pressure-driven and cyclic, and are thus governed by different dynamics than natural drainage networks, which are mainly gravity-driven and dendritic.
- Recent studies in distributed reservoir management have revealed that the placement of reservoirs plays an important role in flood control. Ayalew et al. (2015) develop a framework that combines rainfall-runoff modeling, reservoir routing and Monte Carlo simulation to assess reservoir-regulated flood response (Ayalew et al., 2013), and then subsequently use this framework to investigate the effects of reservoir placement on flood frequency (Ayalew et al., 2015). Using a randomly-generated 1,000-year rainfall time series along with a simulated catchment, they find that two retention basins placed in parallel provide better flood control than either (i) two retention basis placed in series along the river main stem, or (ii) a single large retention basin upstream of the watershed outlet. This research demonstrates that placement of hydraulic control structures exerts a powerful influence on the performance of flood control infrastructure, and raises questions about how larger numbers of control structures should best be distributed throughout a watershed to improve flood control.
- Inspiration for the controller placement problem can be drawn from recent theoretical work into the controllability of complex networks. These studies show that the control properties of complex systems ranging from power grids to gene expression pathways are inextricably linked with topological properties of an underlying network representation (Liu and Barabási, 2016). The location of driver nodes needed for complete controllability of a linear system, for instance, can be determined from the maximum matching of a graph associated with that system's state space representation (Liu et al., 2011). For systems in which complete control of the network is infeasible, the relative performance of driver node configurations can be measured by detecting controllable substructures (Ruths and Ruths, 2014), or by leveraging the concept of "control energy" from classical control theory (Shirin et al., 2017; Summers and Lygeros, 2014; Yan et al., 2012; 2015). While these studies bring a theoretical foundation to the problem of controller placement, they generally assume linear system dynamics, and may thus not be well-suited for drainage networks, which are driven by nonlinear runoff formation and channel routing processes.
- Recent studies have drawn on advances in complex network theory to examine the controllability of stream networks (Riasi and Yeghiazarian, 2017) and enhance understanding of geomorphological processes (Czuba and Foufoula-Georgiou, 2015). Riasi and Yeghiazarian apply several theoretical controllability metrics to real-world drainage networks, ultimately finding that control of dendritic river networks requires a relatively large proportion of driver nodes (Riasi and Yeghiazarian, 2017). Czuba and Foufoula-Georgiou investigate spatial and temporal patterns in sediment accumulation in a river network arising from the combined effects of transport dynamics and stream network topology (Czuba and Foufoula-Georgiou, 2015). They find that the emergence of persistent clusters of mass on the network is a major driver of geomorphological change, and conclude that management efforts should seek to "identify the source contributions that synchronize on the network to form clusters", and then break the synchronization by reducing sediment generation in these regions.

Despite the critical need for system-scale stormwater control, there is to our knowledge no robust theoretical framework to guide the placement of hydraulic control structures for the purposes of improving hydrograph peak attentuation. To address this knowledge gap, we formulate a new graph-theoretic algorithm that uses the network structure of watersheds to determine the controller locations that will maximally "de-synchronize" tributary flows. By flattening the discharge hydrograph, our algorithm provides a powerful method to mitigate flash floods and curtail water quality impairments in urban watersheds. Our approach is distinguished by the fact that it is theoretically-motivated, and links the control of stormwater systems with the underlying structure of the drainage network. The result is a fast, generalized algorithm that requires only digital elevation data for the watershed of interest. More broadly, through our graph-theoretic framework we show that network structure plays a dominant role in the control of drainage basins, and demonstrate how the study of watersheds as complex networks can inform more effective stormwater infrastructure design.

2. Algorithm description

Flashy flows occur when large volumes of runoff arrive synchronously at a given location in the drainage network. If hydraulic control structures are placed at strategic locations, flood waves can be mitigated by "de-synchronizing" tributary flows before they arrive at a common junction. With this in mind, we introduce a controller placement algorithm that minimizes flashy flows by removing regions of the drainage network that contribute disproportionately to synchronous flows at the outlet. In our approach, the watershed is first transformed into a directed graph consisting of unit subcatchments (vertices) connected by flow paths (edges). Next, critical regions are identified by computing the catchment's *width function* (an approximation of the distribution of travel times to the outlet), and then weighting each vertex in the network in proportion to the number of vertices that share the same travel time to the outlet. The weights are used to compute a *weighted*

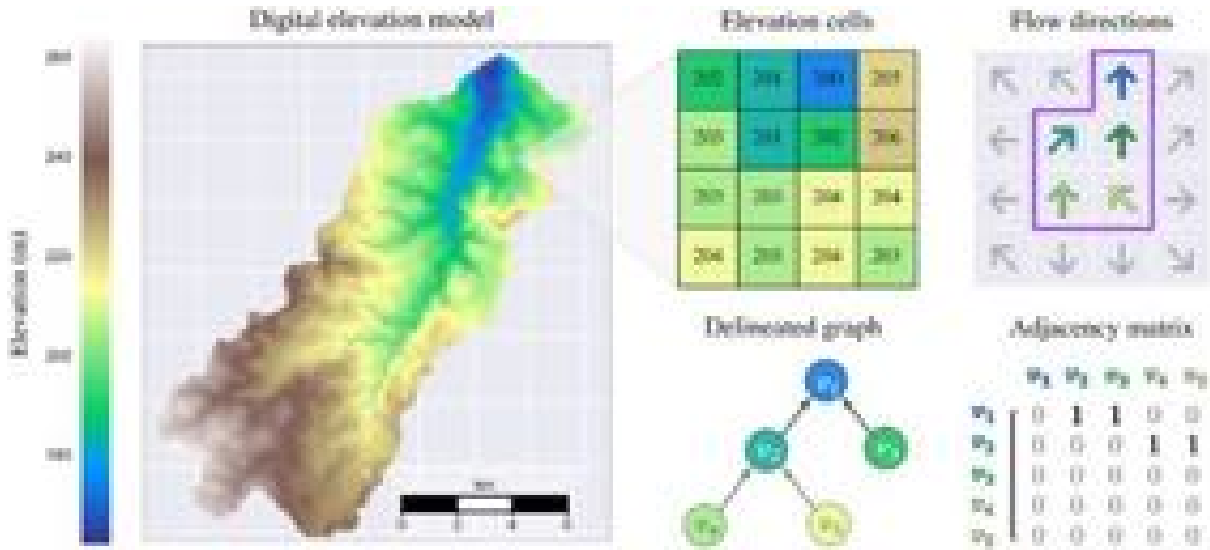


Fig. 1. Left panel: Digital elevation model (DEM) of a watershed with river network highlighted. Right panel (from left to right, top to bottom): (i) DEM detail (colors not to scale); (ii) flow directions; (iii) delineated subcatchment graph; (iv) adjacency matrix representation of graph.

accumulation score for each vertex, which sums the weights of every possible subcatchment in the watershed. The graph is then partitioned recursively based on this weighted accumulation score, with the most downstream vertex of each partition representing a controller location.

2.1. Definitions

Graph representation of a watershed: Watersheds can be represented as directed graphs, in which subcatchments (vertices or cells) are connected by elevation-dependent flow paths (edges). The directed graph can be formulated mathematically as an adjacency matrix, A, where for each element a_{ij}, $a_{ij} \neq 0$ if there exists a directed edge connecting vertex v_j to v_i, and conversely, $a_{ij} = 0$ if there does not exist a directed edge connecting vertex v_j to v_i. Nonzero edge weights can be specified to represent travel times, distances, or probabilities of transition between connected vertices. Flow paths between adjacent cells are established using a routing scheme, typically based on directions of steepest descent (see Fig. 1).

In this study, we determine the connectivity of the drainage network using a *D8 routing* scheme (O'Callaghan and Mark, 1984). In this scheme, elevation cells are treated as vertices in a 2-dimensional lattice (meaning that each vertex v_i is surrounded by eight neighbors $\mathcal{N}_i$). A directed link is established from vertex v_i to a neighboring vertex v_j if the slope between v_i and v_j is steeper than the slope between v_i and all of its other neighbors $\mathcal{N}_i \setminus v_j$ (where v_j has a lower elevation than v_i). The *D8 routing* scheme produces a directed acyclic graph where the indegree of each vertex is between 0 and 8 (with an indegree of 8 indicating that the vertex is a "sink"), and the outdegree of each vertex is 1 (except for the watershed outlet, which has an outdegree of 0). It should be noted that other schemes exist for determining drainage network structure, such as the *D-infinity* routing algorithm, which better resolves drainage directions on hillslopes Tarboton (1997). However, because the routing scheme is not essential to the construction of the algorithm, we focus on the simpler *D8* routing scheme for this study. Similarly, to simplify the construction of the algorithm, we will assume that the vertices of the watershed are defined on a regular grid, such that the area of each unit subcatchment is equal.[1] Fig. 1 shows the result of delineating a river network from a digital elevation model (left), along with an illustration of the underlying graph structure and adjacency matrix representation (right).

Controller: In the context of this study, a controller represents any structure or practice that can regulate flows from an upstream channel segment to a downstream one. Examples include retention basins, dams, weirs, gates and other hydraulic control structures. These structures may be either passively or actively controlled. For the validation assessment presented later in this paper, we will examine the controller placement problem in the context of *volume capture*, meaning that controllers are passive, and that they are large enough to completely remove flows from their upstream contributing areas. However, the algorithm itself does not require the controller to meet these particular conditions.

Mathematically, we can think of a controller as a cut in the graph that removes one of the edges. This cut halts or inhibits flows across the affected edge. Because the watershed has a dendritic structure, any cut in the network will split the network into two sub-trees: (i) the delineated region upstream of the cut, and (ii) all the vertices that are not part of the delineated region. Placing controllers is thus equivalent to removing branches (subcatchments) from a tree (the parent watershed).

Delineation: Delineation returns the set of vertices upstream of a target vertex. In other words, this operation returns the contributing area of vertex v_i. Expressed in terms of the adjacency matrix:

$$V_d(A, v_i) = \{v_j \in V \,|\, (A^n)_{ij} \neq 0 \text{ for some } n \leq D\} \tag{1}$$

where A^n is the adjacency matrix A raised to the n^{th} power, i is the row index, j is the column index, V is the vertex set of A, and D is the graph diameter. Note that $(A^n)_{ij}$ is nonzero only if vertex v_j is located within an n-hop neighborhood of vertex v_i. Note that the delineation operation can also be performed in a single step by analyzing the null space of the graph Laplacian of the watershed's adjacency matrix (Tejedor et al., 2015a).

Pruning: Pruning is the complement of delineation. This operation returns the vertex set consisting of all vertices that are not upstream of the current vertex.

$$V_p(A, v_i) = V \setminus V_d(A, v_i) \tag{2}$$

Subgraphs induced by the delineated and pruned vertex sets are defined as follows:

$$\begin{aligned} A_d(A, v_i) &= A(G[V_d]) \\ A_p(A, v_i) &= A(G[V_p]) \end{aligned} \tag{3}$$

[1] Thus, for watershed models derived from a digital elevation model (DEM), a unit subcatchment is equivalent to a single DEM cell.

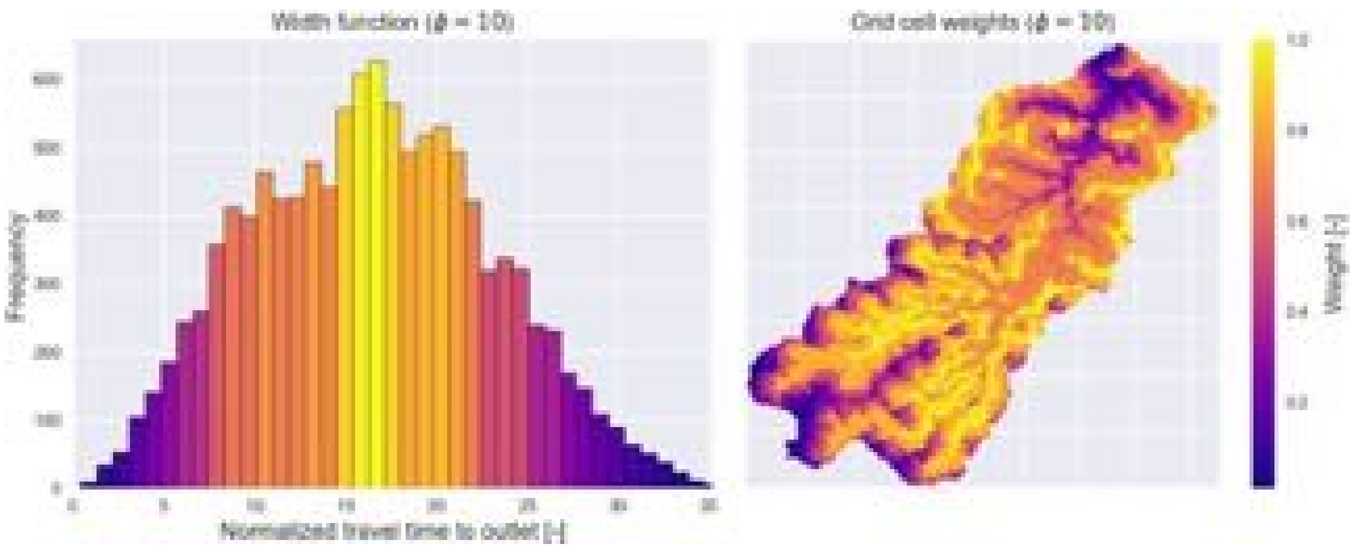


Fig. 2. Left: width function (travel-time histogram) of the watershed, assuming that channelized travel time is ten times faster than on hillslopes ($\phi = 10$). Right: weights associated with each vertex of the graph. Brighter regions correspond to areas that contribute to the peaks of the width function.

where $A(G[V])$ represents the adjacency matrix of the subgraph induced by the vertex set V.

Width function: The width function describes the distribution of travel times from each upstream vertex to some downstream vertex, v_i[2] (Rodriguez-Iturbe and Rinaldo, 2001). In general terms, the width function can be expressed as:

$$H(t, v_i) = \sum_{\gamma \in \Gamma_i} I(\gamma, t) \tag{4}$$

Along with an indicator function, $I(\gamma, t)$:

$$I(\gamma, t) = \begin{cases} 1 & T(\gamma) = t \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where Γ_i is the set of all directed paths to the target vertex v_i, and $T(\gamma)$ is the travel time along path γ. If the travel times between vertices are constant and equal for all vertices, the width function of the graph at vertex v_i can be described as a linear function of the adjacency matrix:[3]

$$H(t, v_i) = (A^t \mathbf{1})(i) \tag{6}$$

where $\mathbf{1}$ signifies the vector of all ones, A^t represents the adjacency matrix A raised to the power t (with t representing the discrete time step), and $(A^t\mathbf{1})(i)$ indicates the ith element of the vector $A^t\mathbf{1}$. In real-world drainage networks, travel times between grid cells are not uniform. Crucially, the travel time for channelized cells will be roughly 1–2 orders of magnitude faster than the travel time in hillslope cells (Rodriguez-Iturbe and Rinaldo, 2001; Tak and Bras, 1990). Thus, to account for this discrepancy, we define ϕ to represent the ratio of hillslope to channel travel times:

$$\phi = \frac{t_h}{t_c} \tag{7}$$

where t_h is the travel time for hillslopes and t_c is the travel time for channels. Fig. 2 (left) shows the width function for an example watershed, under the assumption that channel velocity is ten times faster than hillslope velocity ($\phi = 10$). The width functions for various values of ϕ are shown in Figs. S3 and S4 in the Supplementary Information.

Note that when the effects of hydraulic dispersion are ignored, the width function is equivalent to the geomorphological impulse unit hydrograph (GIUH) of the basin (Rodriguez-Iturbe and Rinaldo, 2001). The GIUH represents the response of the basin to an instantaneous impulse of rainfall distributed uniformly over the catchment; or equivalently, the probability that a particle injected randomly within the watershed at time $t = 0$ exits the watershed through the outlet at time $t = t'$.

Accumulation: The accumulation at vertex v_i describes the number of vertices located upstream of v_i (or alternatively, the upstream area (Moore et al., 1991)). It is equivalent to the cumulative sum of the width function with respect to time[4]:

$$C(v_i) = \left(\sum_{t=0}^{\infty} A^t \mathbf{1} \right)(i) \tag{8}$$

Fig. 3 (left) shows the accumulation at each vertex for an example catchment. Because upstream area is correlated with mean discharge (Rodriguez-Iturbe and Rinaldo, 2001), accumulation is frequently used to determine locations of channels within a drainage network (Moore et al., 1991).

Weighting function: To identify the vertices that contribute most to synchronous flows at the outlet, we propose a weighting function that weights each vertex by its rank in the travel time distribution. Let τ_{ij} represent the known travel time from a starting vertex v_j to the outlet vertex v_i. Then the weight associated with vertex v_j can be expressed in terms of a weighting function $W(v_i, v_j)$:

$$w_j = W(v_i, v_j) = \frac{H(\tau_{ij}, v_i)}{\max_t(H(v_i))} \tag{9}$$

where τ_{ij} represents the travel time from vertex v_j to vertex v_i, $H(\tau_{ij}, v_i)$ represents the width function for an outlet vertex v_i evaluated at time τ_{ij}, and the normalizing factor $\max_t(H(v_i))$ represents the maximum value of the width function over all time steps t. In this formulation, vertices are weighted by the rank of the associated travel time in the width function. Vertices that contribute to the maximum value of the width function (the mode of the travel time distribution) will receive the highest possible weight (unity), while vertices that contribute to the smallest

[2] The width function $H(x)$ was originally defined by Shreve to yield the number of links in the network at a topological distance x from the outlet (Shreve, 1969). Because travel times may vary between hillslope and channel links, we present a generalized formulation of the width function here.

[3] While mathematically concise, this equation is computationally inefficient. See Section S1 in the Supplementary Information for the efficient implementation used in our analysis.

[4] See Section S1 in the Supplementary Information for the efficient implementation of the accumulation algorithm.

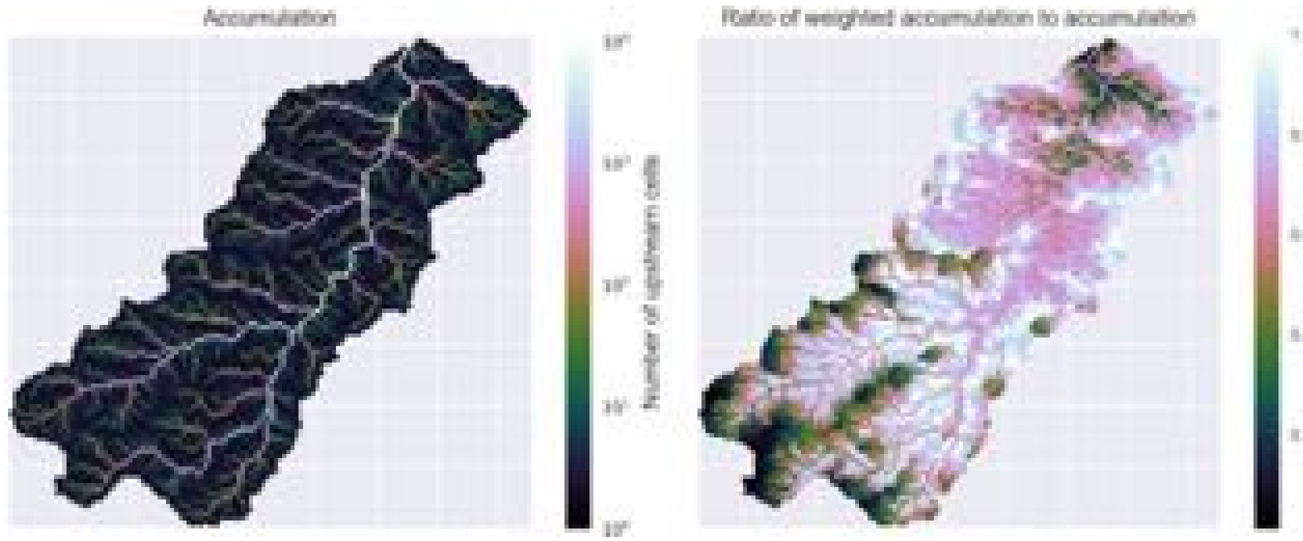


Fig. 3. Left: accumulation (number of cells upstream of every cell). Right: ratio of weighted accumulation to accumulation (C_w/C).

values of the width function will receive small weights. In other words, vertices will be weighted in proportion to the number of vertices that share the same travel time to the outlet. Fig. 2 shows the weights corresponding to each bin of the travel time distribution (left), along with the weights applied to each vertex (right). Weights for varying values of ϕ are shown in Figs. S3 and S4 in the Supplementary Information.

Weighted accumulation: Much like the *accumulation* describes the number of vertices upstream of each vertex v_i, the *weighted accumulation* yields the sum of the weights upstream of v_i. If each vertex v_j is given a weight w_j, the weighted accumulation at vertex v_i can be defined:

$$C_w(v_i, \mathbf{w}) = \left(\sum_{t=0}^{\infty} A^t \mathbf{w} \right)(i) \tag{10}$$

where $\mathbf{w}$ is a vector of weights, with each weight w_j associated with a vertex v_j in the graph. When the previously-defined weighting function is used, the weighted accumulation score measures the extent to which a subcatchment delineated at vertex v_i contributes to synchronous flows at the outlet. In other words, if the ratio of weighted accumulation to accumulation is large for a particular vertex, this means that the subcatchment upstream of that vertex contributes disproportionately to the peak of the hydrograph. Fig. 3 (right) shows the ratio of weighted accumulation to accumulation for the example catchment. The weighted accumulation provides a natural metric for detecting the cuts in the drainage network that will maximally remove synchronous flows, and thus forms the basis of the controller placement algorithm.

2.2. Controller placement algorithm definition

The controller placement algorithm is described as follows. Let A represent the adjacency matrix of a watershed delineated at some vertex v_i. Additionally, let k equal the desired number of controllers, and c equal the maximum upstream accumulation allowed for each controller. The graph is then partitioned according to the following scheme:

1. Compute the width function, $H(t, v_i)$, for the graph described by adjacency matrix A with an outlet at vertex v_i.
2. Compute the accumulation $C(v_j)$ at each vertex v_j.
3. Use $H(t, v_i)$ to compute the weighted accumulation $C_w(v_j)$ at each vertex v_j.
4. Find the vertex v_{opt}, where the accumulation $C(v_{opt})$ is less than the maximum allowable accumulation and the weighted accumulation $C_w(v_{opt})$ is maximized:
$$v_{opt} \leftarrow \underset{v_s \in V_s}{\operatorname{argmax}}(C_w(v_s)) \tag{11}$$
where V_s is the set of vertices such that vertex v_i is reachable from any vertex in V_s and the accumulation C at any vertex in V_s is less than c.
5. Prune the graph at vertex v_{opt}: $A \leftarrow A_p(A, v_{opt})$
6. If the cumulative number of partitions is equal to k, end the algorithm. Otherwise, start at (1) with the catchment described by the new A matrix.

The algorithm is described formally in Algorithm 1. An open-source implementation of the algorithm in the *Python* programming language is also provided (Bartos, 2018a), along with the data needed to reproduce our results. Efficient implementations of the *delineation, accumulation*, and *width function* operations are provided via the `pysheds` toolkit, which is maintained by the authors (Bartos, 2018b).

Fig. 4 shows the controller configuration generated by applying the controller placement algorithm to the example watershed, with $k = 15$ controllers, each with a maximum accumulation of $c = 900$ (i.e. each controller captures roughly 8% of the catchment's land area). In the left panel, partitions are shown in order of decreasing priority from dark to light (i.e. darker regions are partitioned first by the algorithm). The right panel shows the stacked width functions for each partition. The sum of the width functions from each partition reconstitute the original width function for the catchment. From the stacked width functions, it can be seen that the algorithm tends to prioritize the pruning of subgraphs that align with the peaks of the travel time distribution. Note for instance, how the least-prioritized paritions gravitate towards the low end of the travel-time distribution, while the most-prioritized partitions are centered around the mode. Controller placement schemes corresponding to different numbers of controllers are shown in Fig. S5 in the Supplementary Information.

3. Algorithm validation

To evaluate the controller placement algorithm, we simulate the controlled network using a hydrodynamic model, and compare the performance to a series of randomized controller placement configurations. Performance is characterized by the "flatness" of the flow profile at the outlet of the watershed, as measured by both the peak discharge and the variance of the hydrograph (i.e. the extent to which the flow deviates

Algorithm 1 Controller placement algorithm.

Data:
A directed graph described by adjacency matrix A;
A target vertex v_i with index i;
A desired number of partitions k;
The maximum accumulation at each controller, c;
Result: Generate partitions for a catchment

Let $\mathbf{q}$ be a vector representing all vertices in the graph;
Let k_c equal the current number of partitions;
Let τ represent a vector of travel times from each vertex to vertex v_i;
Let A represent the adjacency matrix of the system;
Let A_c represent the adjacency matrix for the current iteration;

$A_c \leftarrow A$;
$k_c \leftarrow 0$;
while $k_c < k$ **do**
  $H(t, v_i) \leftarrow (A_c^t \mathbf{1})(i)$;
  $C \leftarrow (\sum_{t=0}^{\infty} A_c^t \mathbf{1})$;
  $\mathbf{w} \leftarrow W(v_i, \mathbf{q})$;
  $C_w \leftarrow (\sum_{t=0}^{\infty} A_c^t \mathbf{w})$;
  if $C(v_i) > 0$ **then**
    $V_c \leftarrow \{v_m \in V | C(v_m) \leq c\}$;
    $V_s \leftarrow V_d(A_c, v_i) \cap V_c$;
    $v_{opt} \leftarrow \underset{v_s \in V_s}{\text{argmax}}(C_w(v_s))$;
    $A_c \leftarrow A_p(A_c, v_{opt})$;
    $k_c \leftarrow k_c + 1$;
  else
  end
end

from the mean flow over the course of the hydrologic response). To establish a basis for comparison, we simulate a “volume capture” scenario (Emerson et al., 2005), wherein roughly half of the total contributing area is controlled, and each controller completely captures the discharge from its respective upstream area. Additionally, we simulate a “delayed release” scenario in which each controller continuously releases water from a large, bottom-mounted orifice, thereby delaying rather than halting flows from the upstream channel network. These scenarios are chosen as bounding cases, given that most real-world reservoir operation will fall somewhere between these two regimes.

The validation experiment is designed to test the central premises of the controller placement algorithm: that synchronous cells can be identified from the structure of the drainage network, and that maximally capturing these synchronous cells will lead to a flatter overall hydrologic response. If these premises are accurate, we expect to see two results. First, the controller placement algorithm will produce flatter flows than the randomized control trials. Second, the performance of the algorithm will be maximized when using a large number of small partitions. Using many small partitions allows the algorithm to selectively target the highly-weighted cells that contribute disproportionately to the peak of the hydrograph. Conversely, large partitions capture many extraneous low-weight cells that don’t contribute to the peak of the hydrograph. In other words, if increasing the number of partitions improves the performance of the algorithm, it not only confirms that the algorithm works for our particular experiment, but also justifies the central premises on which the algorithm is based.

3.1. Experimental design

We evaluate controller configurations based on their ability to flatten the outlet hydrograph of a test watershed when approximately 50% of the contributing area is controlled. This test case is chosen because

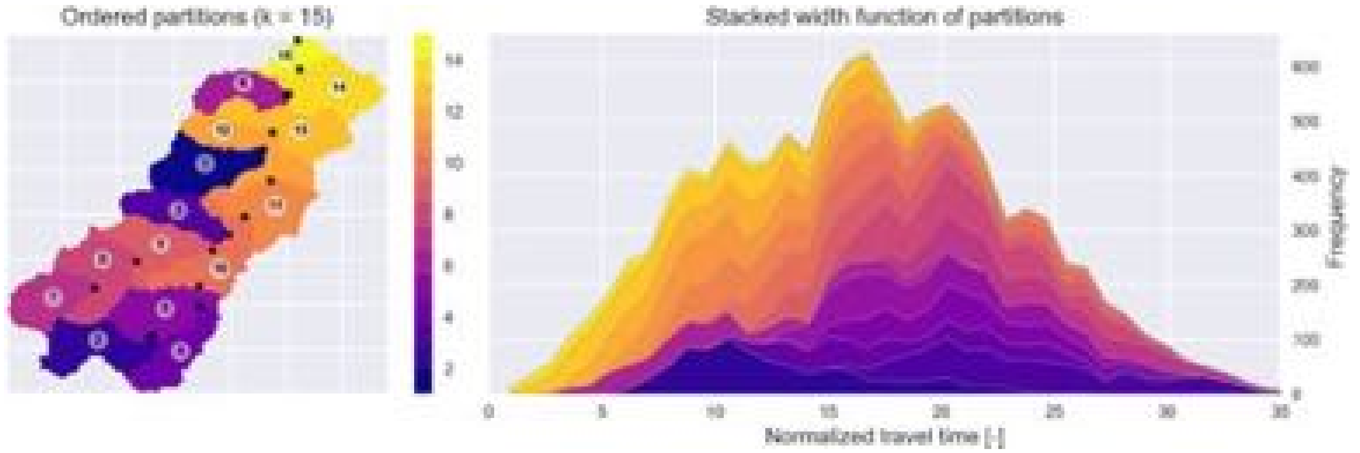


Fig. 4. Left: partitioning of the example watershed using the controller placement algorithm. Numeric labels indicate the order in which partitions are generated (from first to last). Right: stacked width functions for each partition. The brightness expresses the priority of each partition, with the darker partitions being prioritized over the brighter ones.

it presents a practical scenario with real-world constraints, and because it allows for direct comparison of many different controller placement strategies. For our test case, we use the Sycamore Creek watershed, a heavily urbanized creekshed located in the Dallas–Fort Worth Metroplex with a contributing area of roughly 83 km^2 (see Fig. 1). This site is the subject of a long-term monitoring study led by the authors (Bartos et al., 2018), and is chosen for this analysis because (i) it is known to experience issues with flash flooding, and (ii) it is an appropriate size for our analysis—being large enough to capture fine-scale network topology, but not so large that computation time becomes burdensome.

A model of the stream network is generated from a conditioned digital elevation model (DEM) by determining flow directions from the elevation gradient and then assigning channels to cells that fall above an accumulation threshold. Conditioned DEMs and flow direction grids at a resolution of 3 arcseconds (approximately 70 by 90 m) are obtained from the USGS HydroSHEDS database (Lehner et al., 2008). Grid cells with an accumulation greater than 100 are defined to be channelized cells, while those with less than 100 accumulation are defined as hillslope cells. This threshold is based on visual comparison with the stream network defined in the National Hydrography Dataset (NHD) (United States Geological Survey, 2013). Hillslope cells draining into a common channel are aggregated into subcatchments, with a flow length corresponding to the longest path within each hillslope, and a slope corresponding to the average slope over all flow paths in the subcatchment. Percent impervious area and land cover classifications for each subcatchment are obtained from the National Land Cover Database (Homer et al., 2015), allowing for overland flow velocities to be reasonably approximated (see Section S3 in the Supplementary Information). Channel geometries are assigned to each link within the channelized portion of the drainage network. We assume that each stream segment can be represented by a "wide rectangular channel", which is generally accurate for natural river reaches in which the stream width is large compared to the stream depth (Mays, 2010). To simulate channel width and depth, we assume a power law relation between accumulation and channel size based on an empirical formulation from Moody and Troutman (2002):

$$\begin{aligned} \omega &= 7.2 \cdot Q^{0.50\pm0.02} \\ h &= 0.27 \cdot Q^{0.30\pm0.01} \end{aligned} \tag{12}$$

where ω is stream width, h is stream depth, and Q is the mean river discharge. Knowing the width and depth of the most downstream reach, and assuming that the accumulation at a vertex is proportional to the mean flow, we generate channel geometries using the mean parameter values from the above relations. To simulate the effect of floodplain storage and prevent channel overflow, a trapezoidal floodplain section is added to the top of each channel (see Section S4 of the Supplementary Information for additional implementation details).

Controllers are implemented as retention basins regulated by outlet structures with controllable orifices. Orifices are mounted on the bottom of each outlet structure and have approximately 10% of the cross-sectional area of the upstream channel section. For the "volume capture" scenario the orifice is left closed, while for the "delayed release" scenario the orifice is left open. Retention basins are sized using a linear relationship between depth and surface area, and are checked against known real-world retention basins to ensure realistic storage capacities (for additional details, see Section S4 in the Supplementary Information).

Using the controller placement algorithm, control structures are placed such that approximately $50 \pm 3\%$ of the catchment area is captured by storage basins. To investigate the effect of the number of controllers on performance, optimized controller strategies are generated using between $k = 1$ and $k = 35$ controllers. The ratio of hillslope-to-channel travel times is estimated as $\phi = 50$ based on simulations of the catchment under uncontrolled conditions. We compare the performance of our controller placement algorithm to randomized controller placement schemes, in which approximately $50 \pm 3\%$ of the catchment area is controlled but the placement of controllers is random. For this comparison assessment, we generate 50 randomized controller placement trials, each using between $k = 1$ and $k = 24$ controllers.[5] For each randomized trial, the maximum and minimum accumulation that can be handled by each controller is selected, then controllers are placed sequentially until $50 \pm 3\%$ of the total catchment is upstream of at least one controller. This procedure is similar to the controller placement algorithm, except that at each iteration, the controller is placed at a random candidate cell (i.e. a cell with an accumulation in the appropriate range) instead of the candidate cell with the greatest weighted accumulation (see Section S2 in the Supplementary Information for a detailed description of the procedure).

We simulate the hydrologic response using a hydrodynamic model, and evaluate controller placement performance based on the flatness of the resulting hydrograph. To capture the hydrologic response under various rainfall conditions, we simulate small, medium and large rainfall events, corresponding to 11.2, 16.9, and 23.4 mm of rainfall delivered

[5] While the controller randomization code was programmed to use between 1 and 35 controllers, the largest number of controllers achieved was 24. This result stems from the fact that the randomization algorithm struggled to achieve 30 + partitions without selecting cells that fell below the channelization threshold (100 accumulation).

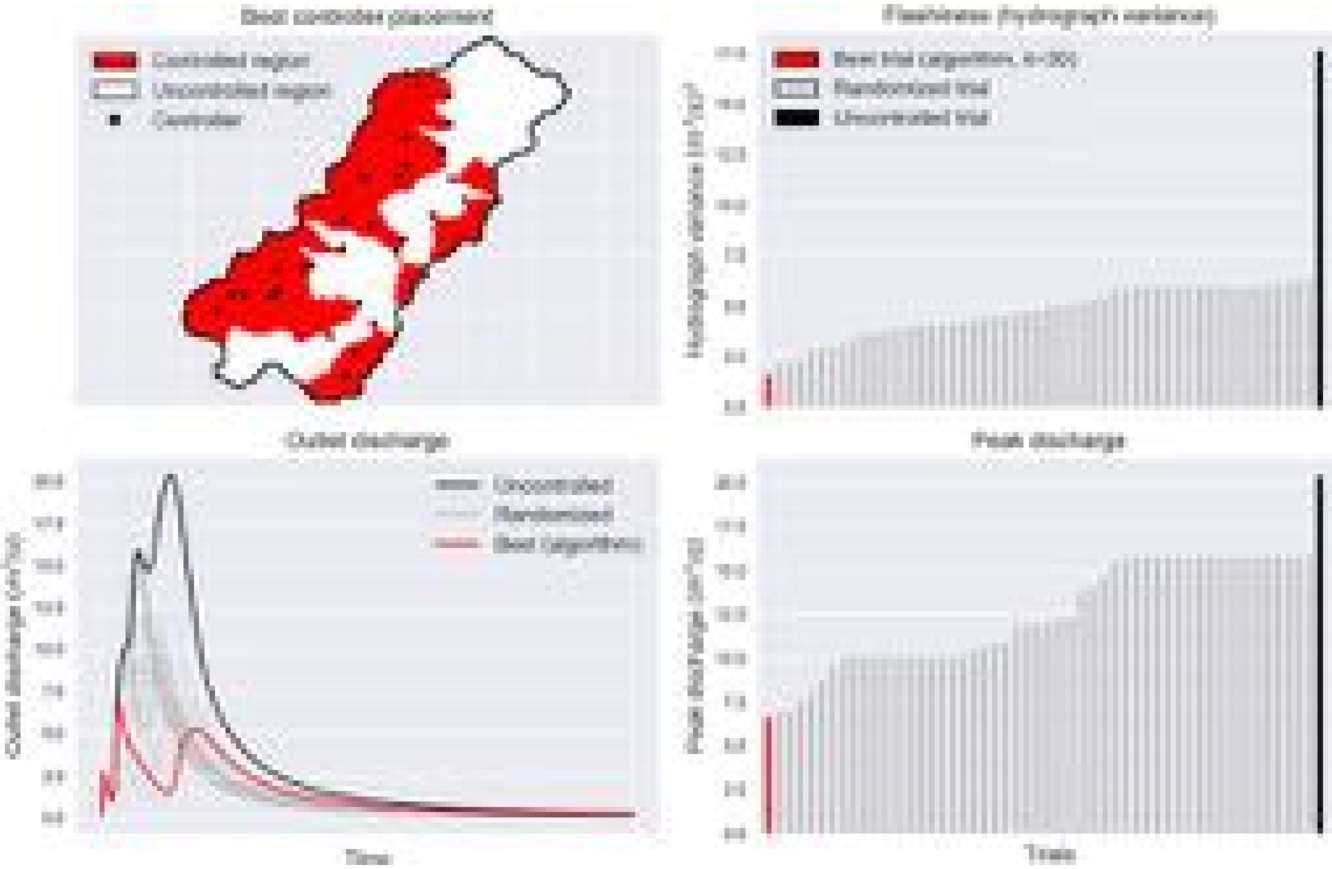


Fig. 5. Results of the hydraulic simulation experiment for the 1-year storm event (11.2 mm) under the volume capture scenario. Top left: best-performing controller placement ($k = 30$), with captured regions in red. Bottom left: hydrographs resulting from each simulation. The uncontrolled simulation is shown in black, while the optimized controller placement simulations are shown in red, and the randomized controller simulations are shown in gray. Right: the overall flashiness (variance of the hydrograph) and peak discharge for each simulation, using the same coloring scheme. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

instantaneously over the first five minutes of the simulation. These rainfall volumes are based on the 1-year, 10-year and 100-year design storms (5-minute duration) from intensity-duration-frequency curves for Tarrant county in Texas (iSWM, 2006). A hydrodynamic model is used to simulate the hydrologic response at the outlet by routing runoff through the channel network using the dynamic wave equations (United States Environmental Protection Agency, 2018). The simulation performance is measured by both the peak discharge and the total variance of the hydrograph. The variance of the hydrograph (which we refer to as "flashiness") is defined as:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (Q_i - \bar{Q})^2 \tag{13}$$

where Q is the discharge, $\bar{Q}$ is the mean discharge in the storm window, and N is the number of data points in the storm window. This variance metric captures the flow's deviation from the mean over the course of the hydrologic response, and thus provides a natural metric for the flatness of the hydrograph. This metric is important for water quality considerations like first flush contamination or streambed erosion—in which the volume of transported material (e.g. contaminants or sediments) depends not only on the maximum discharge, but also on the duration of flow over a critical threshold (Wong and Kerkez, 2016).

Note that the validation experiment is not intended to faithfully reproduce the precise hydrologic response of our chosen study area, but rather, to test the basic premises of the controller placement algorithm. As such, site-specific details—such as soil types and existing infrastructure—have been deliberately simplified. For situations in which these characteristics exert an important influence on the hydrologic response, one may account for these factors by adjusting the inter-vertex travel times used in the controller placement algorithm.

4. Results

The controller placement algorithm produces consistently flatter flows than randomized control trials. Fig. 5 shows the results of the hydraulic simulation assessment in terms of the resulting hydrographs (bottom left), and the overall flashiness and peak discharge of each simulation (right) for the 1-year storm event under the volume capture scenario. The best performance is achieved by using the controller placement algorithm with $k = 30$ controllers (see Fig. 5, top left). Comparing the overall variances and peak discharges, it can be seen that this controller placement produces flatter outlet discharges than any of the randomized controller placement strategies.[6] Specifically, the best controller placement predicted by the algorithm achieves a peak discharge that is roughly 29% of that of the uncontrolled case, while the randomized simulations by comparison achieve an average peak discharge that is more than 61% of that of the uncontrolled case. Similarly, the hydrograph variance of the best controller placement predicted by the

[6] Note that the controller placement algorithm results in a longer falling limb than the randomized trials. This result stems from the fact that the algorithm prioritizes the removal of grid cells that contribute to the peak and rising limb of the hydrograph, while grid cells contributing to the falling limb are ignored. In other words, the controller placement algorithm shifts discharges from the peak of the hydrograph to the falling limb.

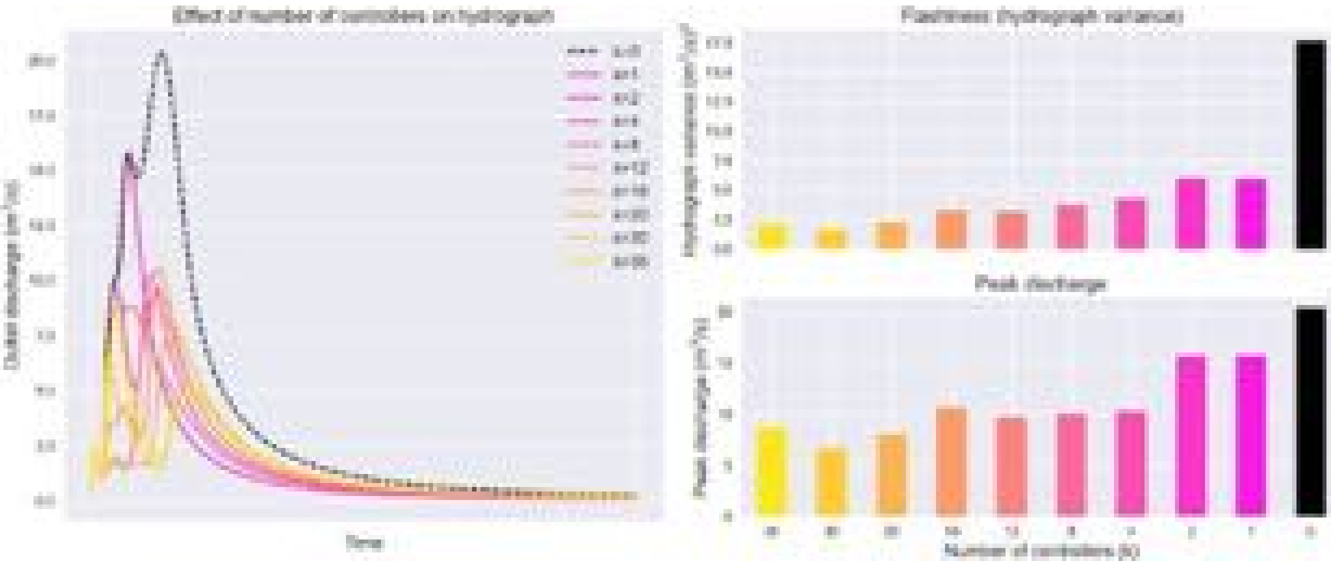


Fig. 6. Left: hydrographs associated with varying numbers of controllers (k), using the controller placement algorithm with 50% watershed area removal (1-year event, volume capture scenario). Right: hydrograph variance (top) and peak discharge (bottom) vs. number of controllers. In general, more controllers produces a flatter response.

algorithm is roughly 9.1% of that of the uncontrolled case, compared to 27% for the randomized simulations on average. Across all numbers of controllers considered, the controller placement algorithm yields results in approximately 20% lower variance and 15% lower peak discharge on average compared to the randomized placement strategy.

The performance of the controller placement algorithm holds under varying rainfall scenarios and reservoir operating rules. When tested against storm events of different sizes (10-year and 100-year storm events), the controller placement algorithm continues to outperform randomized control trials, with roughly 18% lower hydrograph variance and 13% lower peak discharge over all numbers of controllers considered (see Figs. S6–S9 in the Supplementary Information). Moreover, when tested under the delayed release scenario (in which each reservoir continuously releases water from a bottom-mounted orifice), the controller placement algorithm performs better than under the volume capture scenario. In particular, the algorithm achieves 28% lower hydrograph variance on average, while the best controller placement strategy obtained by the algorithm achieves 80% lower hydrograph variance than the average randomized placement (see Figs. S10–S13 in the Supplementary Information). While the performance of the algorithm holds under different rainfall and reservoir operation scenarios, it should be noted that the within-group performance varies with rain event size, which could result from the nonlinearities inherent in wave propagation speed (see the supplementary note in Section S7). Thus, while the optimized controller placement still produces flatter flows than randomized control trials, the performance of the controller placement algorithm could be further improved by tuning the assumed inter-vertex travel times to correspond to the expected speed of wave propagation.

Under the controller placement algorithm, the best performance is achieved by using a large number of small-scale controllers; however, more controllers does not generally lead to better performance for the randomized controller placement scheme. Given that increasing the number of controllers allows the algorithm to better target highly synchronous cells, this result is consistent with the central premise that capturing synchronous cells will lead to a flatter hydrologic response. Fig. 6 shows the hydrologic response when the controller placement algorithm is applied using varying numbers of controllers (left), along with associated hydrograph variances (top right) and peak discharges (bottom right). In all cases, roughly 50% of the watershed is controlled; however, configurations using many small controllers consistently perform better than configurations using a few large controllers. This trend does not hold for the randomized controller placement strategy (see Figs. S14 and S15 in the Supplementary Information). Indeed, the three best-performing randomized controller placements use a median of $k = 17$ controllers, while the three worst-performing randomized controller placements use a median of $k = 10$ controllers (where performance is measured in terms of the hydrograph variance). By comparison, when the controller placement algorithm is used, the three best-performing simulations use a median of $k = 30$ controllers, while the three worst-performing simulations use a median of $k = 2$ controllers. The finding that the controller placement algorithm converges to a (locally) optimal solution follows from the fact that as the number of partitions increases, controllers are better able to capture highly-weighted regions without also capturing extraneous low-weight cells. This in turn implies that the weighting scheme used by the algorithm accurately identifies the regions of the watershed that contribute disproportionately to synchronized flows. Thus, in spite of various sources of model and parameter uncertainty, the experimental results confirm the central principles under which the controller placement algorithm operates: namely, that synchronous regions can be deduced from the graph structure alone, and that controlling these regions results in a flatter hydrograph compared to randomized control trials.

In addition to demonstrating the efficacy of the controller placement algorithm, the validation experiments reveal some general principles for organizing hydraulic control structures within drainage networks to achieve downstream streamflow objectives. Overall, the controller placement strategies that perform best—whether achieved through optimization or randomization—tend to partition the watershed axially rather than laterally. These lengthwise partitions result in a long, thin drainage network that prevents tributary flows from “piling up”. Fig. 7 shows the partitions corresponding to the best-performing and worst-performing controller placement strategies with respect to peak discharge (left and center, respectively), along with the associated hydrographs (right). While the best-performing controller placement strategy evenly distributes the partitions along the length of the watershed, the worst-performing controller placement strategy controls only the most upstream half of the watershed. As a result, the worst-performing strategy removes the largest part of the peak, but completely misses the portion of the peak originating from the downstream half of the watershed. In order to achieve a flat downstream hydrograph, controller placement

Fig. 7. Left: Best controller placement in terms of peak discharge ($k = 30$ controllers, using the controller placement algorithm). Center: worst controller placement in terms of peak discharge ($k = 6$ controllers, randomized). Controller locations are indicated by black crosses, and controlled partitions are indicated by colored regions. Right: hydrographs associated with the best and worst controller placement strategies (1-year storm event, volume capture scenario).

strategies should seek to evenly distribute controllers along the length of the watershed.

5. Discussion

The controller placement algorithm presented in this study provides a tool for designing stormwater control systems to better mitigate floods, regulate contaminants, and protect aquatic ecosystems. By reducing peak discharge, optimized placement of stormwater control structures may help to lessen the impact of flash floods. Existing flood control measures often focus on controlling large riverine floods—typically through existing centralized assets, like dams and levees. However, flash floods may occur in small tributaries, canals, and even normally dry areas. For this reason, flash floods are not typically addressed by large-scale flood control measures, despite the fact that they cause more fatalities than riverine floods in the developed world (Doocy et al., 2013). By facilitating distributed control of urban flash floods, our controller placement strategy could help reduce flash flood related mortality. Moreover, by flattening the hydrologic response, our controller placement algorithm promises to deliver a number of environmental and water quality benefits, such as decreased first flush contamination (Wong and Kerkez, 2016), decreased sediment transport (Muschalla et al., 2014), improved potential for treatment in downstream green infrastructure (Bartos et al., 2018; Kerkez et al., 2016), and regulation of flows in sensitive aquatic ecosystems (Poresky et al., 2015).

5.1. Key features of the algorithm

The controller placement algorithm satisfies a number of important operational considerations:

- **Theoretically motivated**. The controller placement algorithm has its foundation in the theory of the geomorphological impulse unit hydrograph—a relatively mature theory supported by a established body of research (Gupta and Mesa, 1988; Gupta et al., 1986; Kirkby, 1976; Marani et al., 1991; Mesa and Mifflin, 1986; Rodriguez-Iturbe and Rinaldo, 2001; Troutman and Karlinger, 1985). Moreover, the algorithm works in an intuitive way—by recursively removing the subcatchments of a watershed that contribute most to synchronized flows. This theoretical basis distinguishes our algorithm from other strategies that involve exhaustive optimization or direct application of existing graph theoretical constructs (such as graph centrality metrics).
- **Generalizable and extensible.** Because it relies solely on network topology, the controller placement algorithm will provide consistent results for any drainage network—including both natural stream networks and constructed sewer systems. Moreover, because each step in our algorithm has a clear meaning in terms of the underlying hydrology, the algorithm can be modified to satisfy more complex control problems (such as systems in which specific regulatory requirements must be met).
- **Flexible to user objectives and constraints**. The controller placement algorithm permits specification of important practical constraints, such as the amount of drainage area that each control site can capture, and the number of control sites available. Moreover, the weighting function can be adjusted to optimize for a variety of objectives (such as the overall "flatness" of the hydrograph, or removal of flows from a contaminated upstream region).
- **Parsimonious with respect to data requirements**. The controller placement algorithm requires only a digital elevation model of the watershed of interest. Additional data—such as land cover and existing hydraulic infrastructure—can be used to fine-tune estimates of travel times within the drainage network, but are not required by the algorithm itself.
- **Fast implementation** For the watershed examined in this study (consisting of about 12,000 vertices), the controller placement algorithm computes optimal locations for $k = 15$ controllers in roughly 3.0 s (on a 2.9 GHz Intel Core i5 processor). While the computational complexity of the algorithm is difficult to characterize,[7] it is faster than other comparable graph-cutting algorithms, such as recursive spectral bisection or spectral clustering, both of which are $O(n^3)$ in computational complexity.

Taken together, our algorithm offers a solution to the controller placement problem that is suitable for research as well as for practical applications. On one hand, the algorithm is based in hydrologic and geomorphological theory, and provides important insights into the connections between geomorphology and the design of the built environment. On the other hand, the algorithm is fast, robust, and easy-to-use, making it a useful tool for practicing engineers and water resource managers.

5.2. Caveats and directions for future research

While our controller placement algorithm is robust and broadly-applicable, there are a number of important considerations to keep in mind when applying this algorithm to real-world problems.

[7] The computational complexity of the controller placement algorithm depends on the implementation of component functions (such as delineation and accumulation computation), which can in turn depend on the structure of the watershed itself.

- The controller placement algorithm implicitly assumes that rainfall is uniform over the catchment of interest. While this assumption is justified for small catchments in which the average spatial distribution of rainfall will be roughly uniform, this assumption may not hold for large (e.g. continent-scale) watersheds. Modifications to the algorithm would be necessary to account for a non-uniform spatial distribution of rainfall.
- The controller placement algorithm is sensitive to the chosen ratio of hillslope to channel speeds, ϕ. Care should be taken to select an appropriate value of ϕ based on site-specific land cover and morphological characteristics. More generally, for situations in which differential land cover, soil types, and existing hydraulic infrastructure play a dominating role, the performance of the algorithm may be enhanced by adjusting inter-vertex travel times to correspond to estimated overland flow and channel velocities.
- Our assessment of the algorithm's performance rests on the assumption that installed control structures (e.g. retention basins) are large enough to capture upstream discharges. The algorithm itself does not explicitly account for endogenous upstream flooding that could be introduced by installing new control sites.
- In this study, experiments were conducted only for impulsive rainfall inputs (i.e. with a short duration of rainfall). Future work should assess the performance of the distance-weighted controller placement strategy under arbitrary rainfall durations.
- Our analysis assumes that reservoirs are initially empty before each storm event. While some previous studies in distributed reservoir operation contend that an initially-empty condition is "the simplest and most defensible approach" (Goldman, 2001), other studies use random initialization of reservoir depths to simulate the effect of successive storm events when reservoir operation rules are unknown (Ayalew et al., 2013). While this latter approach may provide more realistic results under unknown reservoir operating conditions, we ultimately use an initially-empty condition due to the combinatorial difficulty of assessing the effect of random initial depths alongside varying numbers of controllers, controller placement strategies, and rainfall scenarios. Random initialization could potentially effect the results by inducing overflows under the volume capture scenario, or by quickening the rising limb under the delayed release scenario. Additional work is needed to understand how these effects would impact the performance of the controller placement algorithm compared to randomized control trials. With this in mind, it should be noted that new "smart" water systems are enabling more flexible control of distributed stormwater infrastructure (Kerkez et al., 2016), which may in turn strengthen the assumption of initially-empty storage conditions. Bartos et al., for instance, present a real-world case study in which real-time analytics and control are used to pre-emptively evacuate retention basins before a storm event, reducing the magnitude of the downstream hydrologic response (Bartos et al., 2018). Because initially emptying storage basins often leads to a favorable hydrologic response (Wong, 2017), assuming empty or near-empty initial storage conditions may be more realistic than assuming random initial depths for systems with real-time control capabilities.
- This study focuses primarily on event-based diagnostics of system performance—specifically, by measuring the flatness of the hydrologic response under independent storm events. However, it should be noted that water infrastructure may also be evaluated in terms of long-term performance—for instance, by measuring the response of the system to an extended stochastic rainfall time series. While not computationally feasible for the model used in this study, future work should investigate the performance of the controller placement algorithm under extended hydrodynamic simulations.

More broadly, future research should investigate the problem of sensor placement in stream networks using the theoretical framework developed in this paper. While this study focuses on the problem of placement of hydraulic control structures, our algorithm also suggests a solution to the problem of sensor placement. Stated probabilistically, the geomorphological impulse unit hydrograph (GIUH) represents the probability that a "particle" injected randomly within the watershed at time $t = 0$ exits the outlet at time $t = t'$. Thus, the peaks of the GIUH correspond to the portions of the hydrologic response where there is the greatest amount of ambiguity about where a given "particle" originated. It follows that the same locations that maximally de-synchronize flows may also be the best locations for disambiguating the locations from which synchronous flows originated. Future experiments should investigate the ability to estimate upstream states (e.g. flows) within the network given an outlet discharge along with internal state observers (e.g. flow sensors) placed using the algorithm developed in this study.

6. Conclusions

We develop an algorithm for placement of hydraulic control structures that maximally flattens the hydrologic response of drainage networks. This algorithm uses the geomorphological impulse unit hydrograph to locate subcatchments that dominate the peaks of the hydrograph, then partitions the drainage network to minimize the contribution of these subcatchments. We find that the controller placement algorithm produces flatter hydrographs than randomized controller placement trials—both in terms of peak discharge and overall variance. By reducing the flashiness of the hydrologic response, our controller placement algorithm may one day help to mitigate flash floods and restore urban water quality through reduction of contaminant loads and prevention of streambed erosion. We find that the performance of the algorithm is enhanced when using a large number of small, distributed controllers. In addition to confirming the central hypothesis that synchronous cells can be identified based on network structure of drainage basins, this result lends justification to the development of decentralized *smart* stormwater systems, in which active control of small-scale retention basins, canals and culverts enables more effective management of urban stormwater. Overall, our algorithm is efficient, requires only digital elevation model data, and is robust to parameter and model uncertainty, making it suitable both as a research tool, and as a design tool for practicing water resources engineers.

Declarations of interest

None.

Data availability

Code and data links are available at: https://github.com/kLabUM/hydraulic-controller-placement.

Acknowledgments

Funding for this project was provided by the National Science Foundation (grants 1639640 and 1442735) and the University of Michigan. We would like to thank Alex Ritchie for exploring alternative approaches to the controller placement problem and for his help with literature review. We would also like to thank Dr. Alfred Hero for his advice in formulating the problem.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2019.03.016.

References

Arora, M., Malano, H., Davidson, B., Nelson, R., George, B., 2015. Interactions between centralized and decentralized water systems in urban context: a review. Wiley Interdiscip. Rev. 2 (6), 623–634. https://doi.org/10.1002/wat2.1099.

Askarizadeh, A., Rippy, M.A., Fletcher, T.D., Feldman, D.L., Peng, J., Bowler, P., Mehring, A.S., Winfrey, B.K., Vrugt, J.A., AghaKouchak, A., Jiang, S.C., Sanders, B.F., Levin, L.A., Taylor, S., Grant, S.B., 2015. From rain tanks to catchments: use of low-impact development to address hydrologic symptoms of the urban stream syndrome. Environ. Sci Technol. 49 (19), 11264–11280. https://doi.org/10.1021/acs.est.5b01635.

Ayalew, T.B., Krajewski, W.F., Mantilla, R., 2013. Exploring the effect of reservoir storage on peak discharge frequency. J. Hydrol. Eng. 18 (12), 1697–1708. https://doi.org/10.1061/(asce)he.1943-5584.0000721.

Ayalew, T.B., Krajewski, W.F., Mantilla, R., 2015. Insights into expected changes in regulated flood frequencies due to the spatial configuration of flood retention ponds. J. Hydrol. Eng. 20 (10), 04015010. https://doi.org/10.1061/(asce)he.1943-5584.0001173.

Bartos, M., 2018a. Controller Placement Code. https://github.com/kLabUM/hydraulic-controller-placement.

Bartos, M., 2018b. Pysheds: Simple and Fast Watershed Delineation in Python. https://github.com/mdbartos/pysheds.

Bartos, M., Wong, B., Kerkez, B., 2018. Open storm: a complete framework for sensing and control of urban watersheds. Environ. Sci. 4 (3), 346–358. https://doi.org/10.1039/c7ew00374a.

Booth, D.B., Jackson, C.R., 1997. Urbanization of aquatic systems: degradation thresholds, stormwater detection, and the limits of mitigation. J. Am. Water Resour. Assoc. 33 (5), 1077–1090. https://doi.org/10.1111/j.1752-1688.1997.tb04126.x.

Cattafi, M., Gavanelli, M., Nonato, M., Alvisi, S., Franchini, M., 2011. Optimal placement of valves in a water distribution network with clp(fd). Theory Pract. Logic Program. 11 (4–5), 731747. https://doi.org/10.1017/S1471068411000275.

Coffman, L.S., Goo, R., Frederick, R., 1999. Low-impact development: an innovative alternative approach to stormwater management. WRPMD 1999. American Society of Civil Engineers. https://doi.org/10.1061/40430(1999)118.

Creaco, E., Franchini, M., Alvisi, S., 2010. Optimal placement of isolation valves in water distribution systems based on valve cost and weighted average demand shortfall. Water Resour. Manage. 24 (15), 4317–4338. https://doi.org/10.1007/s11269-010-9661-5.

Czuba, J.A., Foufoula-Georgiou, E., 2015. Dynamic connectivity in a fluvial network for identifying hotspots of geomorphic change. Water Resour. Res. 51 (3), 1401–1421. https://doi.org/10.1002/2014wr016139.

Dams and Development, 2000. A new framework for decision-making - the report of the world commission on dams. Routledge. https://doi.org/10.4324/9781315541518.

Doocy, S., Daniels, A., Murray, S., Kirsch, T.D., 2013. The human impact of floods: a historical review of events 1980–2009 and systematic literature review. PLoS Curr. https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a.

Emerson, C.H., Welty, C., Traver, R.G., 2005. Watershed-scale evaluation of a system of storm water detention basins. J. Hydrol. Eng. 10 (3), 237–242. https://doi.org/10.1061/(asce)1084-0699(2005)10:3(237).

Finkenbine, J.K., Atwater, J., Mavinic, D., 2000. Stream health after urbanization. J. Am. Water Resour. Assoc. 36 (5), 1149–1160.

Gelormino, M.S., Ricker, N.L., 1994. Model-predictive control of a combined sewer system. Int. J. Control 59 (3), 793–816. https://doi.org/10.1080/00207179408923105.

Goldman, D.M., 2001. Quantifying uncertainty in estimates of regulated flood frequency curves. Bridging the Gap. American Society of Civil Engineers. https://doi.org/10.1061/40569(2001)273.

Gupta, V.K., Mesa, O.J., 1988. Runoff generation and hydrologic response via channel network geomorphology — recent progress and open problems. J. Hydrol. 102 (1–4), 3–28. https://doi.org/10.1016/0022-1694(88)90089-3.

Gupta, V.K., Waymire, E., Rodríguez-Iturbe, I., 1986. On scales, gravity and network structure in basin runoff. In: Scale Problems in Hydrology. Springer Netherlands, pp. 159–184. https://doi.org/10.1007/978-94-009-4678-1_8.

Hajebi, S., Roshani, E., Cardozo, N., Barrett, S., Clarke, A., Clarke, S., 2015. Water distribution network sectorisation using graph theory and many-objective optimisation. J. Hydroinf. https://doi.org/10.2166/hydro.2015.144.

Hawley, R.J., Vietz, G.J., 2016. Addressing the urban stream disturbance regime. Freshwater Sci. 35 (1), 278–292. https://doi.org/10.1086/684647.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 national land cover database for the conterminous united states–representing a decade of land cover change information. Photogramm. Eng. Remote Sens. 81 (5), 345–354.

iSWM design manual for site development, 2006. Appendix A: rainfall tables for North Central Texas. Technical Report. North Central Texas Council of Governments, Environment and Development Department.

Kerkez, B., Gruden, C., Lewis, M., Montestruque, L., Quigley, M., Wong, B., Bedig, A., Kertesz, R., Braun, T., Cadwalader, O., Poresky, A., Pak, C., 2016. Smarter stormwater systems. Environ. Sci. Technol. 50 (14), 7267–7273. https://doi.org/10.1021/acs.est.5b05870.

Kirkby, M.J., 1976. Tests of the random network model, and its application to basin hydrology. Earth Surf. Process. 1 (3), 197–212. https://doi.org/10.1002/esp.3290010302.

Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. Eos Trans. Am. Geophys. Union 89 (10), 93. https://doi.org/10.1029/2008eo100001.

Liu, Y.-Y., Barabási, A.-L., 2016. Control principles of complex systems. Rev. Mod. Phys. 88 (3). https://doi.org/10.1103/revmodphys.88.035006.

Liu, Y.-Y., Slotine, J.-J., Barabási, A.-L., 2011. Controllability of complex networks. Nature 473 (7346), 167–173. https://doi.org/10.1038/nature10011.

Mantilla, R., Gupta, V.K., Troutman, B.M., 2011. Scaling of peak flows with constant flow velocity in random self-similar networks. Nonlinear Process. Geophys. 18 (4), 489–502. https://doi.org/10.5194/npg-18-489-2011.

Marani, A., Rigon, R., Rinaldo, A., 1991. A note on fractal channel networks. Water Resour. Res. 27 (12), 3041–3049. https://doi.org/10.1029/91wr02077.

Mays, L., 2010. Water Resources Engineering. John Wiley & Sons.

Meerow, S., Newell, J.P., 2017. Spatial planning for multifunctional green infrastructure: growing resilience in detroit. Landsc. Urban Plan. 159, 62–75. https://doi.org/10.1016/j.landurbplan.2016.10.005.

Mesa, O.J., Mifflin, E.R., 1986. On the relative role of hillslope and network geometry in hydrologic response. In: Scale Problems in Hydrology. Springer Netherlands, pp. 1–17. https://doi.org/10.1007/978-94-009-4678-1_1.

Mollerup, A.L., Mikkelsen, P.S., Thornberg, D., Sin, G., 2016. Controlling sewer systems - a critical review based on systems in three EU cities. Urban Water J. 14 (4), 435–442. https://doi.org/10.1080/1573062x.2016.1148183.

Montestruque, L., Lemmon, M.D., 2015. Globally coordinated distributed storm water management system. In: Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks - CySWater 2015. ACM Press https://doi.org/10.1145/2738935.2738948.

Moody, J.A., Troutman, B.M., 2002. Characterization of the spatial variability of channel morphology. Earth Surf. Process. Landforms 27 (12), 1251–1266. https://doi.org/10.1002/esp.403.

Moore, I.D., Grayson, R., Ladson, A., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5 (1), 3–30.

Mullapudi, A., Bartos, M., Wong, B., Kerkez, B., 2018. Shaping streamflow using a real-time stormwater control network. Sensors 18 (7), 2259. https://doi.org/10.3390/s18072259.

Mullapudi, A., Wong, B.P., Kerkez, B., 2017. Emerging investigators series: building a theory for smart stormwater systems. Environ. Sci. 3 (1), 66–77. https://doi.org/10.1039/c6ew00211k.

Muschalla, D., Vallet, B., Anctil, F., Lessard, P., Pelletier, G., Vanrolleghem, P.A., 2014. Ecohydraulic-driven real-time control of stormwater basins. J. Hydrol. 511, 82–91. https://doi.org/10.1016/j.jhydrol.2014.01.002.

Norton, B.A., Coutts, A.M., Livesley, S.J., Harris, R.J., Hunter, A.M., Williams, N.S., 2015. Planning for cooler cities: a framework to prioritise green infrastructure to mitigate high temperatures in urban landscapes. Landsc. Urban Plan. 134, 127–138. https://doi.org/10.1016/j.landurbplan.2014.10.018.

O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Comput. Vis. Graph. Image Process. 27 (2), 247. https://doi.org/10.1016/s0734-189x(84)80047-x.

Perelman, L., Ostfeld, A., 2013. Application of graph theory to sensor placement in water distribution systems. World Environmental and Water Resources Congress 2013. American Society of Civil Engineers. https://doi.org/10.1061/9780784412947.060.

Petrucci, G., Rioust, E., Deroubaix, J.-F., Tassin, B., 2013. Do stormwater source control policies deliver the right hydrologic outcomes? J. Hydrol. 485, 188–200. https://doi.org/10.1016/j.jhydrol.2012.06.018.

Poresky, A., Boyle, R., Cadwalader, O., 2015. Piloting real time control retrofits of stormwater facilities: two oregon case studies and beyond. In: Proceedings of the Pacific Northwest Clean Water Association, Boise, ID, USA, pp. 26–27.

Riasi, M.S., Yeghiazarian, L., 2017. Controllability of surface water networks. Water Resour. Res. 53 (12), 10450–10464. https://doi.org/10.1002/2017wr020861.

Rodriguez-Iturbe, I., Rinaldo, A., 2001. Fractal River Basins: Chance and Self-organization. Cambridge University Press.

Rosenberg, E.A., Keys, P.W., Booth, D.B., Hartley, D., Burkey, J., Steinemann, A.C., Lettenmaier, D.P., 2010. Precipitation extremes and the impacts of climate change on stormwater infrastructure in washington state. Clim. Change 102 (1–2), 319–349. https://doi.org/10.1007/s10584-010-9847-0.

Ruths, J., Ruths, D., 2014. Control profiles of complex networks. Science 343 (6177), 1373–1376. https://doi.org/10.1126/science.1242063.

Sahagun, L., 2013. High Cost of Fighting Urban Runoff Examined in Report. LA Times.

Schilling, J., Logan, J., 2008. Greening the rust belt: a green infrastructure model for right sizing America's shrinking cities. J. Am. Plann. Assoc. 74 (4), 451–466.

Schubert, J.E., Burns, M.J., Fletcher, T.D., Sanders, B.F., 2017. A framework for the case-specific assessment of green infrastructure in mitigating urban flood hazards. Adv. Water Resour. 108, 55–68. https://doi.org/10.1016/j.advwatres.2017.07.009.

Shirin, A., Klickstein, I.S., Sorrentino, F., 2017. Optimal control of complex networks: balancing accuracy and energy of the control action. Chaos 27 (4), 041103. https://doi.org/10.1063/1.4979647.

Shreve, R.L., 1969. Stream lengths and basin areas in topologically random channel networks. J. Geol. 77 (4), 397–414. https://doi.org/10.1086/628366.

Strecker, E., Quigley, M.M., Urbonas, B.R., Jones, J., Clary, J., 2000. Determining urban stormwater BMP effectiveness. Proc. Water Environ. Federation 2000 (6), 2395–2412. https://doi.org/10.2175/193864700785150457.

Summers, T.H., Lygeros, J., 2014. Optimal sensor and actuator placement in complex dynamical networks. IFAC Proc. Vol. 47 (3), 3784–3789. https://doi.org/10.3182/20140824-6-za-1003.00226.

Tak, L.D., Bras, R.L., 1990. Incorporating hillslope effects into the geomorphologic instantaneous unit hydrograph. Water Resour. Res. 26 (10), 2393–2400.

Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resour. Res. 33 (2), 309–319. https://doi.org/10.1029/96wr03137.

Tejedor, A., Longjas, A., Zaliapin, I., Foufoula-Georgiou, E., 2015a. Delta channel networks: 1. A graph-theoretic approach for studying connectivity and steady state transport on deltaic surfaces. Water Resour. Res. 51 (6), 3998–4018. https://doi.org/10.1002/2014wr016577.

Tejedor, A., Longjas, A., Zaliapin, I., Foufoula-Georgiou, E., 2015b. Delta channel networks: 2. metrics of topologic and dynamic complexity for delta comparison, physical inference, and vulnerability assessment. Water Resour. Res. 51 (6), 4019–4045. https://doi.org/10.1002/2014wr016604.

Troutman, B.M., Karlinger, M.R., 1985. Unit hydrograph approximations assuming linear flow through topologically random channel networks. Water Resour. Res. 21 (5), 743–754. https://doi.org/10.1029/wr021i005p00743.

Tzatchkov, V.G., Alcocer-Yamanaka, V.H., Ortíz, V.B., 2008. Graph theory based algorithms for water distribution network sectorization projects. Water Distribution Systems Analysis Symposium 2006. American Society of Civil Engineers. https://doi.org/10.1061/40941(247)172.

United States Environmental Protection Agency, 2018. ORD Stormwater Management Model. https://github.com/USEPA/Stormwater-Management-Model.

United States Geological Survey, 2013. National Hydrography Geodatabase. https://viewer.nationalmap.gov/viewer/nhd.html?p=nhd.

Walsh, C.J., Roy, A.H., Feminella, J.W., Cottingham, P.D., Groffman, P.M., Morgan, R.P., 2005. The urban stream syndrome: current knowledge and the search for a cure. J. North Am. Benthol. Soc. 24 (3), 706–723. https://doi.org/10.1899/04-028.1.

Wang, L., Lyons, J., Kanehl, P., Bannerman, R., 2001. Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environ. Manage. 28 (2), 255–266. https://doi.org/10.1007/s0026702409.

Wines, M., 2014. Behind Toledo's Water Crisis, a Long-Troubled Lake Erie, 4. New York Times.

Wong, B., 2017. Real-time measurement and control of urban stormwater systems. Ph.D. thesis. University of Michigan.

Wong, B.P., Kerkez, B., 2016. Adaptive measurements of urban runoff quality. Water Resour. Res. 52 (11), 8986–9000. https://doi.org/10.1002/2015WR018013.

Wong, B.P., Kerkez, B., 2018. Real-time control of urban headwater catchments through linear feedback: performance, analysis and site selection. Water Resour. Res. https://doi.org/10.1029/2018wr022657.

Wright, J., Marchese, D., 2017. Briefing: continuous monitoring and adaptive control: the 'smart' storm water management solution. Proc. Inst. Civil Eng. 170 (4), 86–89. https://doi.org/10.1680/jsmic.17.00017.

Yan, G., Ren, J., Lai, Y.-C., Lai, C.-H., Li, B., 2012. Controlling complex networks: how much energy is needed? Phys. Rev. Lett. 108 (21). https://doi.org/10.1103/physrevlett.108.218703.

Yan, G., Tsekenis, G., Barzel, B., Slotine, J.-J., Liu, Y.-Y., Barabási, A.-L., 2015. Spectrum of controlling and observing complex networks. Nat. Phys. 11 (9), 779–786. https://doi.org/10.1038/nphys3422.

Yao, L., Chen, L., Wei, W., Sun, R., 2015. Potential reduction in urban runoff by green spaces in Beijing: a scenario analysis. Urban For. Urban Green. 14 (2), 300–308. https://doi.org/10.1016/j.ufug.2015.02.014.

Yazdani, A., Jeffrey, P., 2011. Robustness and vulnerability analysis of water distribution networks using graph theoretic and complex network principles. Water Distribution Systems Analysis 2010. American Society of Civil Engineers. https://doi.org/10.1061/41203(425)85.

Zellner, M., Massey, D., Minor, E., Gonzalez-Meler, M., 2016. Exploring the effects of green infrastructure placement on neighborhood-level flooding via spatially explicit simulations. Comput. Environ. Urban Syst. 59, 116–128. https://doi.org/10.1016/j.compenvurbsys.2016.04.008.

Zhang, B., di Xie, G., Li, N., Wang, S., 2015. Effect of urban green space changes on the role of rainwater runoff reduction in Beijing, China. Landsc. Urban Plan. 140, 8–16. https://doi.org/10.1016/j.landurbplan.2015.03.014.

Water Research 145 (2018) 697–706

Contents lists available at ScienceDirect

Water Research

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

Are all data useful? Inferring causality to predict flows across sewer and drainage systems using directed information and boosted regression trees

Yao Hu [a], Donald Scavia [b], Branko Kerkez [a, *]

[a] Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, United States
[b] School for Environment and Sustainability, University of Michigan, Ann Arbor, United States

ARTICLE INFO

Article history:
Received 11 May 2018
Received in revised form
3 August 2018
Accepted 4 September 2018
Available online 4 September 2018

Keywords:
Flow prediction
Causality
Directed information
Boosted regression trees
Data-driven model

ABSTRACT

As more sensor data become available across urban water systems, it is often unclear which of these new measurements are actually useful and how they can be efficiently ingested to improve predictions. We present a data-driven approach for modeling and predicting flows across combined sewer and drainage systems, which fuses sensor measurements with output of a large numerical simulation model. Rather than adjusting the structure and parameters of the numerical model, as is commonly done when new data become available, our approach instead learns causal relationships between the numerically-modeled outputs, distributed rainfall measurements, and measured flows. By treating an existing numerical model – even one that may be outdated – as just another data stream, we illustrate how to automatically select and combine features that best explain flows for any given location. This allows for new sensor measurements to be rapidly fused with existing knowledge of the system without requiring recalibration of the underlying physics. Our approach, based on *Directed Information* (DI) and *Boosted Regression Trees* (BRT), is evaluated by fusing measurements across nearly 30 rain gages, 15 flow locations, and the outputs of a numerical sewer model in the city of Detroit, Michigan: one of the largest combined sewer systems in the world. The results illustrate that the *Boosted Regression Trees* provide skillful predictions of flow, especially when compared to an existing numerical model. The innovation of this paper is the use of the *Directed Information* step, which selects only those inputs that are causal with measurements at locations of interest. Better predictions are achieved when the *Directed Information* step is used because it reduces overfitting during the training phase of the predictive algorithm. In the age of "big water data", this finding highlights the importance of screening all available data sources before using them as inputs to data-driven models, since more may not always be better. We discuss the generalizability of the case study and the requirements of transferring the approach to other systems.

1. Introduction

The need to understand and predict water flows across cities is important for predicting flash flooding, reducing sewer overflows, and designing infrastructure (Field and Tafuri, 2006; Morales et al., 2017; Paquier et al., 2015). The dynamics of flow across these systems are complicated by the combined influences of hydrology, infrastructure, and highly variable rainfall (Konrad, 2003). Presently, predictive approaches attempt to capture many of these features explicitly in the form of numerical models. These models, which are underpinned by physical laws and are often derived from first-order principles, represent the urban water systems at high resolutions and capture very specific characteristics, such as pipe dimensions, soil types, orifices, and subcatchment dynamics. For large cities, this can often lead to highly structured models that are difficult to parameterize and calibrate.

Simultaneously, many cities are scaling efforts to monitor assets in real-time, which means that more distributed sensor data are becoming available. For example, flow meters, water level sensors, and water quality sensors are now readily being deployed across urban water systems (Kerkez et al., 2016). The proliferation of

* Corresponding author. Department of Civil and Environmental Engineering, 2350 Hayward, 2044 GG Brown, Ann Arbor, MI, 48109-2125, United States.
E-mail addresses: huya@umich.edu (Y. Hu), scavia@umich.edu (D. Scavia), bkerkez@umich.edu (B. Kerkez).

https://doi.org/10.1016/j.watres.2018.09.009
0043-1354/

sensors seems promising, but more data may not always be helpful, especially if they do not exhibit a causal relationship with states being modeled. In those instances, two questions arise: *(1) Which emerging sources of data are actually useful in explaining flow across large urban water systems? (2) For those inputs that are deemed important, what quantity of data is required, and how can these data be rapidly ingested to improve predictions?*

Instead of relying on the recalibration of a numerical model, this paper presents a data-driven approach that combines all available data sources, including the outputs of an existing numerical model, into a holistic and automated prediction of water flows. In this way, the predictive skill embedded in a numerical model is retained when useful, while any additional sources of sensor data are ingested to further improve predictions. The fundamental contribution of the paper is a new method to predict flows in urban drainage systems, which: (1) selects useful (causal) inputs through a *Directed Information* algorithm, and (2) yields flow predictions with the selected inputs using *Boosted Regression Trees*. As more diverse data sources become available to decision makers, this approach will allow for the rapid and automated incorporation of emerging data into holistic predictions of flows. The approach is evaluated by fusing measurements from nearly 30 rain gages and 15 flow sensors with the outputs of a numerical sewer model in the city of Detroit, Michigan.

2. Background

2.1. Predicting through numerical models

A number of popular urban drainage models are presently in use, including the Stormwater Management Model (SWMM), MIKE URBAN and HEC-HMS, to name a few. These models couple hydrology and hydraulics, numerically computing processes such as infiltration and shallow water flow. Once calibrated, these models can be very effective at forecasting and decision making across fine spatiotemporal resolutions. The most common approach to model calibration seeks to adjust the model structure and its parameters so that the model output agrees with the measurements (Sun and Sun, 2015). This often includes a combination of manual parameter tuning that relies on the expertise of modelers or, in some cases, auto-calibration (Doherty, 2015). If knowledge is updated—due to changes in infrastructure, new measurements, or updated information on the watershed—model recalibration is often needed.

It is well known that standard parameter calibration methods are subject to the *curse of dimensionality,* where computational cost increases exponentially with the number of calibrating parameters (Sun and Sun, 2015). Given this complexity and resulting financial cost of recalibration, most water models hardly keep pace with urban change or the emergence of new data sources. In fact, it is not uncommon for many numerical water models in the United States to be over a decade. As such, more streamlined approaches are needed to keep pace with the emergence of new data sources and to ensure forecasts are made using the most relevant and up-to-date information.

2.2. Data-driven forecasting

In lieu of statistical models, a number of data-driven approaches are showing promise to model water systems. Instead of explicitly modeling physics, these approaches rely only on data, such as sensor measurements or features of a system, to make forecasts. While data-driven models cannot always provide insight into system behavior, they can enable streamlined and adaptive toolchains to rapidly ingest many data sources. Broadly, many of these approaches fall under the umbrella of supervised machine learning (ML) approaches, which use historical data to "learn" complex mappings between inputs and a target variable.

When modeling flows across urban water systems, traditional ML approaches, such as generalized linear regression, may not work well due to the nonlinearities and collinearities inherent in complex systems (Dormann et al., 2013). Nonlinear mappings can be learned through methods such as *Artificial Neural Networks* (ANN), but these often require a large amount of data and are computationally expensive (Schalkoff, 1997). For many applications, a supervised ML approach known as *Boosted Regression Trees*, has recently shown promise in balancing computational intensity with performance. According to Caruana and Niculescu-Mizil (2006), *Boosted Regression Trees* have shown the best overall predictive performance among supervised learning algorithms, while remaining immune to collinearity. Regardless of the choice of algorithm, it is well known that many data-driven approaches may be sensitive to overfitting due to inadequate input data selection, especially if irrelevant inputs are used during training (Ng, 1998). As such, the prospects associated with access to many new sensor measurements may be hampered by the realization that more is not always better. That is, not all data may be useful, and brute-force use of all available data may actually lead to worse forecasts.

3. Methods

Instead of forcing the choice between a numerical or data-driven approach, as is commonly done, our method seeks to strike a balance by combining the benefits of both. The toolchain uses *Boosted Regression Trees* to make forecasts by ingesting a large number of input features (Fig. 1). However, instead of using sensor data as the only input, we also treat an existing numerical model as another input data source – the idea being that even an out-of-date or poorly calibrated numerical model may still embed a significant amount of information, which may not be captured by sensor data alone. Depending on the location of interest, various combinations of input features may provide the best forecast. As such, an innovation in our approach is a preliminary step, which uses the criterion of *Directed Information*, to determine which inputs may lead to the best prediction. This forms a holistic and automated toolchain, which ingests all available data but reduces the risk of overfitting by selecting the most "useful" data for forecasting.

3.1. Feature selection using directed information

Given a set of random processes $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m\}$ and a target random process $\mathbf{Y}$, we seek to predict the process $\mathbf{Y}$ using the processes in $\mathbf{X}$. In our case, $\mathbf{Y}$ represents a time series of flow measurements at the location for which we would like to generate future predictions (namely, the target variable). The processes $\mathbf{X}$ include time series that could be used to derive a predictive model for $\mathbf{Y}$, such as rainfall measurements, as well as the outputs of a numerical model. In many cases, it may not be advantageous to use all processes from $\mathbf{X}$ as inputs to a predictive model because this may lead to a model that performs well in the training phase, but one that performs poorly in prediction (i.e. overfitting). Instead, the goal is to select a subset of processes that are "useful" in describing $\mathbf{Y}$. Statistically, this can be captured using *Granger Causality* (Granger, 1969). If predictions of $\mathbf{Y}$ are improved by using a process $\mathbf{X}_i \in \mathbf{X}$, we say $\mathbf{X}_i$ statistically causes $\mathbf{Y}$. More formally, $\mathbf{X}_i$ causes $\mathbf{Y}$, as measured by Granger Causality, if the past of $\mathbf{X}_i$ can help predict the future of $\mathbf{Y}$.

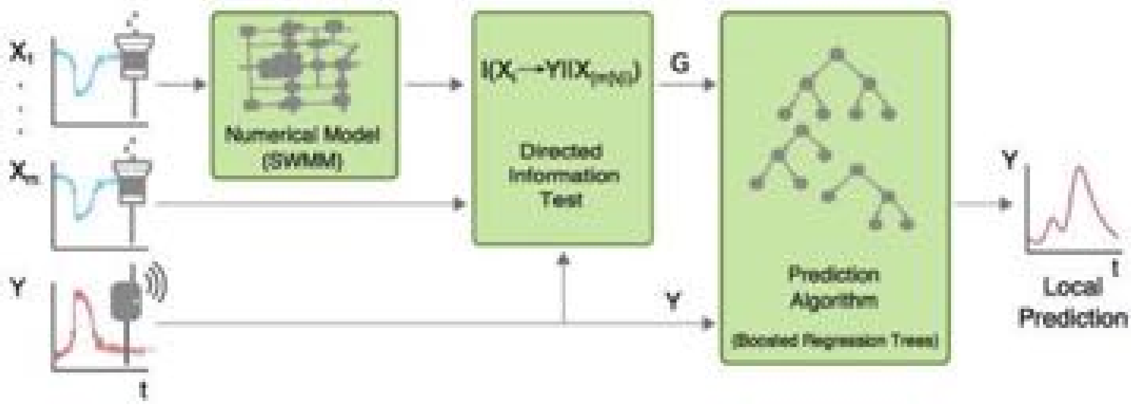


Fig. 1. Predicting flows ($\mathbf{Y}$) by combining inputs features (sensor data) and the outputs of a numerical model ($\mathbf{X}_1,\ldots,\mathbf{X}_m$). The Directed Information check is used to select only those input features that are statistically causal with the flow measurements ($\mathbf{Y}$). The selected features are then used as inputs to the Boosted Regression Trees prediction algorithm.

The metric of causality, known as *Directed Information* (DI), is an information theoretic quantity that measures the statistical causation. Given a set of random processes $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m\}$ and $\mathbf{Y}$, the *Directed Information* $\mathbf{I}$ from $\mathbf{X}_1$ to $\mathbf{Y}$ is defined as the time-averaged expected log-likelihood ratio between two conditional probability distributions of $\mathbf{Y}$ at time step t, $\mathbf{Y}^t$ (Equation (1)). This ratio is also known as Kullback–Leibler Divergence (Kullback and Leibler, 1951). For the conditional probability in the numerator, $\mathbf{Y}^t$ is conditioned on the past of $\mathbf{X}_1$, $\mathbf{X}_1^{1:t-1}$ (Marko, 1973; Kramer, 1998):

$$\mathbf{I}(\mathbf{X}_1 \to \mathbf{Y}||\mathbf{X}_2,\ldots,\mathbf{X}_m) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{E}_{P_{Y,X_1,X_2,\ldots,X_m}}\left[\log \frac{P_{Y^t|X_1^{1:t-1},X_2^{1:t-1},\ldots,X_m^{1:t-1}}}{P_{Y^t|X_2^{1:t-1},\ldots,X_m^{1:t-1}}}\right], \tag{1}$$

where $\mathbf{X}_i^{t_1:t_2}$ denotes the process $\mathbf{X}_i$ from time step t_1 to t_2. If the past of $\mathbf{X}_1$ can help predict the future of $\mathbf{Y}$, then the conditional probability with $\mathbf{X}_1^{1:t-1}$, $P_{Y^t|X_1^{1:t-1},X_2^{1:t-1},\ldots,X_m^{1:t-1}}$ is larger than the conditional probability without $\mathbf{X}_1^{1:t-1}$, and the expected log-likelihood of their ratio will be positive. Otherwise, if the past of $\mathbf{X}_1$ cannot help predict the future of $\mathbf{Y}$, then the two conditional probabilities are equal, in which case the expected log-likelihood of their ratio is zero. In other words, the future value of $\mathbf{Y}$ is conditionally independent from the past value of $\mathbf{X}_1, \mathbf{X}_1^{1:t-1}$ given the past value of the rest processes in $\mathbf{X}$, $\mathbf{X}_2^{1:t-1}, \ldots, \mathbf{X}_m^{1:t-1}$. For random processes $\mathbf{Y}$ and $\mathbf{X}_1$, the larger the *Directed Information* value, the more causal influence $\mathbf{X}_1$ has on $\mathbf{Y}$ – and hence the more "useful" $\mathbf{X}_1$ is on predicting $\mathbf{Y}$.

To reduce potential overfitting, we use a model complexity penalty known as minimum description length (MDL; Grünwald, 2007):

$$\mathbf{MDL} = \mathbf{h}\frac{\log_2(\mathbf{n})}{2\mathbf{n}}, \tag{2}$$

where $\mathbf{h}$ denotes the Markov order and $\mathbf{n}$ denotes the sample size of $\mathbf{X_i}$ used for model fitting. The use of this penalty term ensures that only the random processes with *Directed Information* values larger than the MDL are considered as causal. This is summarized in Algorithm 1, which seeks to store all these causal processes in a new subset $\mathbf{G}$, which will be subsequently used as input to a prediction algorithm (modified from Quinn et al., 2015). When selecting candidate features for the *Directed Information* test, non-deterministic relationships among all features need to be guaranteed—that is, no feature can be derived directly from the others.

Algorithm 1: Assessment of causal influence with directed information

```
Input  : m random processes: X={X_1, ..., X_m}
         Target random process: Y
Output: Causal influence processes: G
begin
    for each i ∈ [m] do
        G(i) ← ∅ ;
    end
    for each i ∈ [m] do
        if I(X_i → Y||X_[m]\{i}) ≥ MDL then
            G(i) ← G(i) ∪ {j} ;
        else
            continue;
        end
    end
    return G
end
```

3.2. Prediction using boosted regression trees

Once the causal features are selected using the *Directed Information* (DI) approach, they can be used to train a predictive model, which in our case takes the form of *Boosted Regression Trees* (BRT). Instead of learning one regression tree, *Boosted Regression Trees* learn multiple trees and weigh them to describe the relationship between the target variable and the features. *Boosted Regression Trees* rely on boosting methods that create an ensemble of regression models to improve the accuracy of model fitting (Elith et al., 2008). Given a regression problem, *Boosted Regression Trees* assign individual weights to every sample point of the training data set. A single regression tree is then constructed and evaluated using the data. A loss function (Wald, 1950), which describes the deviance between the measurements and predicted values, is used to update the individual weights on the tree. The data points with larger deviance are assigned larger weights in the next step. After the updated weights are assigned to individual data points, a new regression tree is constructed. The procedure repeats until the number of the iterations M is reached (Algorithm 2).

Through a forward and additive fashion, *Boosted Regression Trees* gradually optimize predictive performance by using linear combination of all individual trees. Like many supervised ML approaches, *Boosted Regression Trees* can still be subject to overfitting if too many features are used. To this end, the *Directed Information* is used to select only causal features, thereby reducing the potential for overfitting in the final *Boosted Regression Trees*.

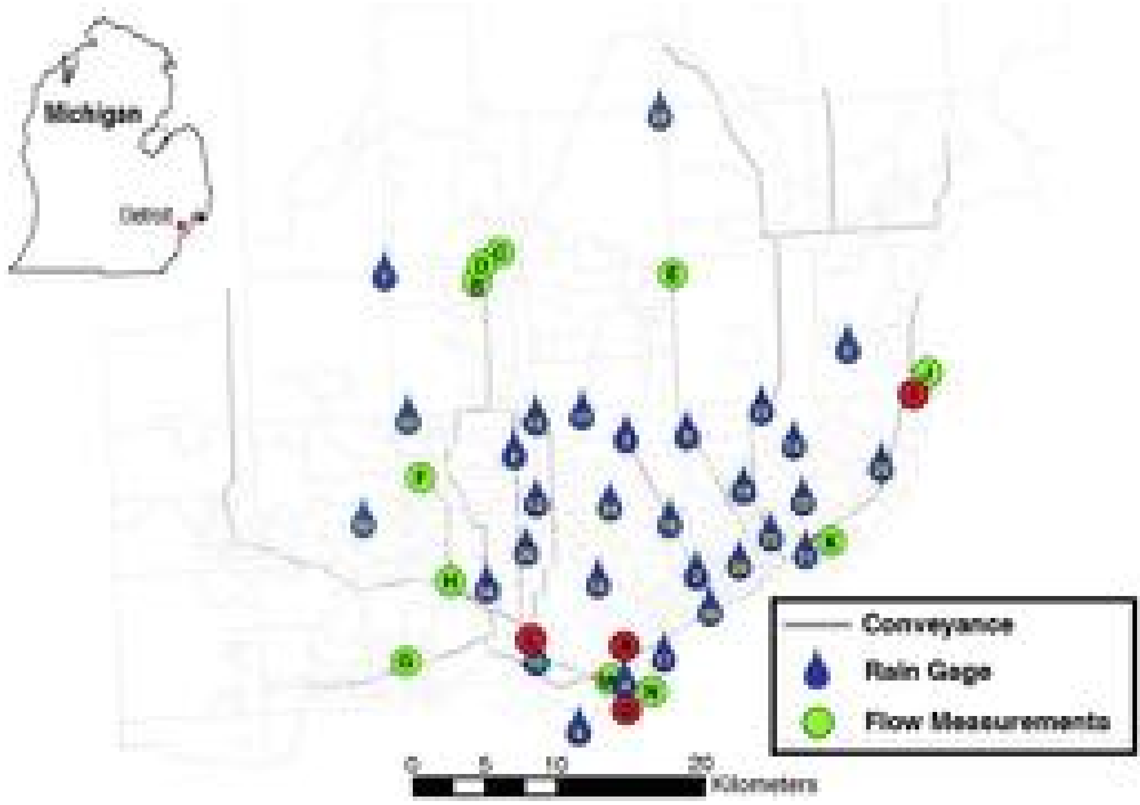


Fig. 2. Detroit sewer collection system service area, showing location of 30 rain gages and 15 prediction points of interest. Site A measures inflow volume to the Wastewater Treatment Plant and sites B-O measure combined sewer overflow volume.

Algorithm 2: Boosted Regression Trees

Input : Features: $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m\}$
Target variable: $\mathbf{Y}$
Number of iterations: M

Output: Ensemble of regression trees: $\mathbf{T}$

begin
  Initialize weights w_i^0 for the measurements of $\mathbf{Y}$, $\{Y_i, i = 1, \ldots, n\}$;
  for each $j \in [1, \ldots, M]$ **do**
    1. Use features $\mathbf{X}$ to fit a regression tree, T_j to the weighted measurements using weights, w_i^{j-1} and yield the predicted values, $\hat{y}^j$;
    2. Compute the loss between $\hat{y}^j$ and the measurements of $\mathbf{Y}$;
    3. Update the weights based on the loss from step 2, w_i^j.
  end
  Build $\mathbf{T}$ through the linear combination of T_j;
  return $\mathbf{T}$
end

4. Case study and implementation

4.1. Study area

Our case study concerns the prediction of flows in large combined sewer systems. Specifically, we focus on the city of Detroit, Michigan, one of the largest combined sewer collection systems in the United States (Fig. 2). We seek to predict daily flows at locations of interest using a large number of available data sources. Detroit and its surrounding suburbs are the largest urban source of total phosphorus to the river system connecting lakes Huron and Erie (Maccoux et al., 2016), and because phosphorus load from this system is driven primarily by flow, predicting and controlling the occurrence of combined sewer overflows is important. While dozens of new measurements (e.g., rain gages) have become available across the city, they have not yet been used in a predictive model. A physically-based numerical model is available but was updated over 6 years ago. As such, this case study presents a great opportunity to apply our approach to fusing new sensor data with the expertise embedded in the existing numerical model.

4.2. Data source: numerical model

In 1998, a physically-based hydrological and hydraulic model was developed using the EPA Storm Water Management Model (SWMM) for the Detroit sewer collection system's 1963.2 km^2 service area (Tenbroek et al., 1999) (Fig. 2). The first version of the model was initially calibrated with available flow data (Santini et al., 2001). Since then, it has been updated several times to reflect new facilities, including 14 major combined sewer overflow outfalls (sites B through O in Fig. 2). The latest version of the model was released in 2012, which is the model that has been shared with the authors. Initial inspection revealed that while the model did represent some of the larger, downstream flows adequately (e.g. flows at the final outlet of the system), it generally overestimated daily flows across smaller, upstream locations.

4.3. Data source: sensor measurements

Hourly flow measurements were made by sensors at the terminal node of the system (Fig. 2, site A from April, 2014 to July, 2014), representing the inflows into the Wastewater Treatment Plant. Event-based, combined sewer overflow volume measurements from May 2013 to October 2015 were also obtained from the Michigan Department of Environmental Quality[1] for 38 storm events during this period. Most combined sewer overflow events occurred within one day; however, when an event spanned multiple days, a daily average was obtained by dividing the total discharge by the number of days of the event. Hourly precipitation data from 2013 to 2015, which served as input into the numerical model, were also obtained from 30 distributed rain gages in the service area (Fig. 2).

4.4. Implementation

The objective of the evaluation was to predict daily inflows to the wastewater treatment plant and sewer overflows at 15 sites. For the purpose of this study, predicting inflows to the treatment plant tests the ability of the approach to describe large-scale, continuous flows, while predicting combined sewer overflows captures the ability to predict smaller-scale, more dynamic events. Data from all rain gages were used as inputs to the SWMM model, after which SWMM outputs were used as input into the *Directed Information* test. The rainfall data were manually quality controlled by gap filling measurements or saturation points through interpolation with neighboring gages. This was intended to ensure that the highest possible quality inputs were used as inputs to the SWMM. For each site, the *Directed Information* test was first used to select causal input features from available rain gages and co-located SWMM outputs. While feasible, upstream flow measurements were not used as inputs to downstream predictions since it was assumed that upstream dynamics are implicitly captured by SWMM. To meet the non-determinism criterion of the *Directed Information* test, one of the gages (site 2) was randomly removed from the data set before the *Directed Information* algorithm was executed. Once causal features were selected, they were forwarded to the *Boosted Regression Trees* algorithm (60/40% split for training and validation). Through iteration, the number of trees was set to 500, while the tree depth was set equal to 4 and the learning rate equal to 0.1.

To promote transparency, experimental repeatability, and broader adoption, all of the source code for this paper is shared in an open source web repository (http://github.com/kLabUM/DIBRT). The entire approach has been implemented in MATLAB. Due to security considerations and data agreement with the Great Lakes Water Authority (owner of the data), the authors are unable to share the SWMM model and sensor data. However, an anonymized dataset has been provided in the same web repository to allow others to evaluate the general functionality of our approach.

4.5. Evaluation

To evaluate performance, two fit metrics were used, R-squared (R^2) and Nash-Sutcliffe efficiency (NSE):

$$R^2 = \frac{\left(\sum_{i=1}^{n} \widehat{Y}_i Y_i - n\overline{\widehat{Y}}\overline{Y}\right)^2}{\left(\sum_{i=1}^{n} \widehat{Y}_i^2 - n\overline{\widehat{Y}}^2\right)\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right)} \text{ where } R^2 \in [0,\ 1]$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(\widehat{Y}_i - Y_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \times 100 \text{ where } NSE \in (-\infty,\ 1], \quad (3)$$

where n is the sample size; Y_i and $\widehat{Y}_i$ are the measurements and predicted values, respectively; $\overline{Y}$ and $\overline{\widehat{Y}}$ are the mean value of Y_i and $\widehat{Y}_i$. For interpretation, an R^2 value closer to 1 indicates a good model fit, while a NSE of 1 indicates perfect fit. To determine if the *Directed Information* step of our approach actually improves final forecasts made by the *Boosted Regression Trees*, two scenarios were evaluated. The first used the *Directed Information* step, as described previously, and the other did not, instead directly feeding all available inputs to the regression tree algorithm. The improvement in R^2 and NSE scores was calculated to determine the benefits of using *Directed Information*.

5. Results

Given the amount of data in the study, four sites have been selected to illustrate and visualize the performance of the approach. The predictive performance across all sites is summarized below, while detailed information and additional figures are provided in the Supplementary Information of this paper.

5.1. Performance

The performance of the algorithm at site A is shown in Fig. 3, from April 2014 to July 2014. Measured flows are compared to those predicted by the numerical model (SWMM), as well as those predicted by our algorithm. While the SWMM model performed relatively well at this location compared to other sites (NSE = 0.17 and R^2 = 0.59), it nonetheless had a positive bias, tending to over-predict peak flows. For this location, which corresponds to the largest conduit in the system (inflow to the treatment plant), the *Directed Information* algorithm selected 7 of the 30 features as inputs to the *Boosted Regression Trees*. Out of these features, SWMM model output was selected as the major source of information, followed by six rain gages (Table 1). There was no clear correspondence between the causal influence of gages and their proximity to the modeled site. Once trained on two months of data, the *Directed Information Boosted Regression Trees* algorithm predicted flows well (NSE = 0.52/R^2 = 0.58). The predictive ability was quite pronounced especially during rainfall events, during which the *Boosted Regression Trees* were able to accurately reconstruct both the magnitude and dynamics of the flows.

Site O is one of the most downstream combined sewer overflows. The frequency of the difference between modeled and predicted overflow volumes for 2013–2015 is compared for SWMM and *Boosted Regression Trees* (Fig. 4a). Since many storms did not result in overflow events, we compare the frequency of prediction residuals (difference between measurements and prediction) rather than a time series comparison. This site had more overflows compared to other locations, even during small storm events. For this location, the SWMM model vastly over-predicted the overflow volumes, by nearly an order of magnitude (NSE = −24.1 and R^2 = 0.39). However, the *Boosted Regression Trees* predictions showed much better agreement with the measurements (NSE = 0.62 and R^2 = 0.61). For this location, the *Directed Information* algorithm selected 10 total features as inputs to the *Boosted Regression Trees* Algorithm. Interestingly, even though SWMM alone performed poorly, its outputs were still selected as the most informative feature for the predictive model (Table 1). Compared to SWMM, the algorithm reduced the overestimation bias by nearly a factor of 8.

The performance of the algorithm at site I is illustrated by Fig. 4b, which corresponds to a sewer overflow location in the system. For site I, the SWMM model generally over-predicted flows (NSE = −2.71 and R^2 = 0.65) and the *Directed Information* step did

[1] Michigan Department of Environmental Quality (MDEQ): http://www.michigan.gov/deq/.

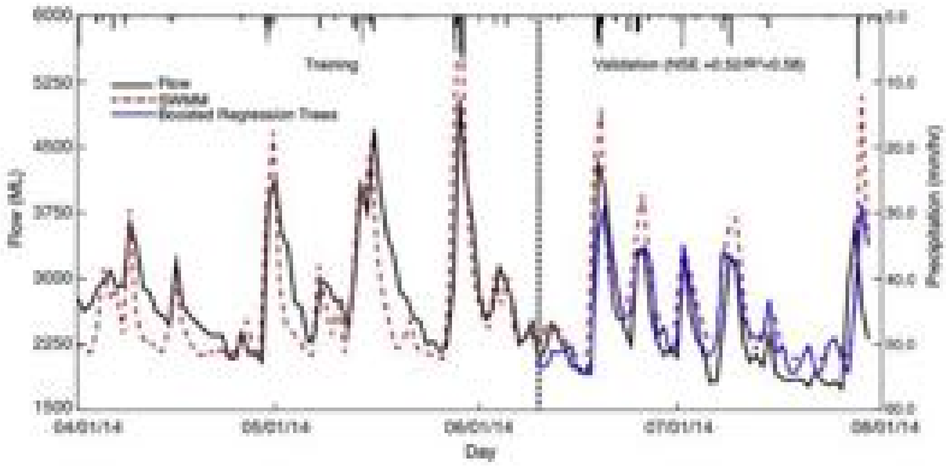


Fig. 3. Comparison of daily flows (million liters, ML) between April, 2014 and July, 2014 at site A (black-solid: measured inflow; red-dashed: SWMM prediction; blue-solid: Boosted Regression Trees prediction). The gray dashed line separates the training and validation phases for the Boosted Regression Trees. The upper part of the figure shows the average precipitation (mm per hour; mm/hr) during this period. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
List of influential features on the flow measurements and their DI values.

Site	Influential Feature	DI Value[a]
Site A	SWMM-modeled flow	0.99
	Gage 10	0.08
	Gage 12	0.06
	Gage 19	0.06
	Gage 25	0.06
	Gage 9	0.04
	Gage 18	0.04
Site O	SWMM-modeled flow	0.72
	Gage 30	0.05
	Gage 13	0.04
	Gage 26	0.03
	Gage 34	0.03
	Gage 3	0.02
	Gage 10	0.02
	Gage 21	0.02
	Gage 8	0.01
	Gage 24	0.01
Site I	Gage 19	0.14
	Gage 14	0.04
	Gage 32	0.03
	Gage 18	0.02
	Gage 9	0.01
	Gage 12	0.01
Site L	Gage 30	0.12
	SWMM-modeled flow	0.03
	Gage 3	0.03
	Gage 7	0.01
	Gage 11	0.01

[a] For site A, features with DI value no less than MDL = 0.03 were considered as influential; for sites O, I and L, features with DI value no less than MDL = 0.01 are considered as influential.

not select it as an influential input feature. As such, the *Boosted Regression Trees* were trained only on six selected gages, yielding an improved predictive performance (NSE = 0.69 and $R^2 = 0.8$) compared to SWMM.

The measurements at site L (Fig. 4c) showed very few overflows across the 2013–2015 study period. Here, neither the SWMM model (NSE = −0.07 and $R^2 = 0.0$), nor the *Boosted Regression Trees* (NSE = 0.14 and $R^2 = 0.16$) performed well. While the *Directed Information* algorithm selected the SWMM outputs as one informative feature, a number of rain rages were deemed much more informative (Table 1).

5.2. Selection of informative inputs

Overall, the *Boosted Regression Trees* approach, when combined with *Directed Information* feature selection, was able to predict flows at 10 of the 15 sites well, as measured by NSE or R^2 scores (>0.4). The SWMM model was selected as an informative input for 11 of the 15 sites (Fig. 5). The number of inputs selected varied from site to site, with no clear relationships to physical features, such as distance to the input rain gages. Performance was mainly related to the magnitude and variability of flows at the target location. The algorithm generally performed better at locations with more non-zero measurements (e.g. active flows or overflows during every storm). Many of the lower-performing sites generally had mostly no flows or overflows during storms. The variability of flows also played a role in predicative performance. Flow at sites with highly variable flows or overflows (measured by deviation from mean) was more difficult to predict.

The performance of the approach across all sites is summarized in Table 2 as a comparison of the quality of the predictions with and without *Directed Information*. Overall, the use of the *Directed Information* step reduced the fit during the training phase of the *Boosted Regression Trees*, but improved its performance during validation, as quantified by an improvement in NSE and R^2 scores. The use of the *Directed Information* step improved the predictive performance at almost all locations and improved the performance significantly (increased NSE or R^2 by at least 0.05) for more than half of the sites.

6. Discussion

The use of data-driven prediction techniques, such as *Boosted Regression Trees*, shows a good potential for predicting complex and nonlinear flows across large water systems. As more data become available, these methods will offer an automated and efficient way to rapidly ingest and adapt to new sources of information. As shown here, new data sources are not limited to new sensors, such as rain gages. Rather, existing numerical models can serve as valuable inputs. For a given location, if the underlying numerical model already captures flows accurately, the *Boosted Regression Trees* will still improve predictions, but may not outperform the numerical model. This was the case for site A (Fig. 3) where the SWMM model already had fairly strong performance. In such instances, *Boosted Regression Trees* offer a rapid way to ingest new

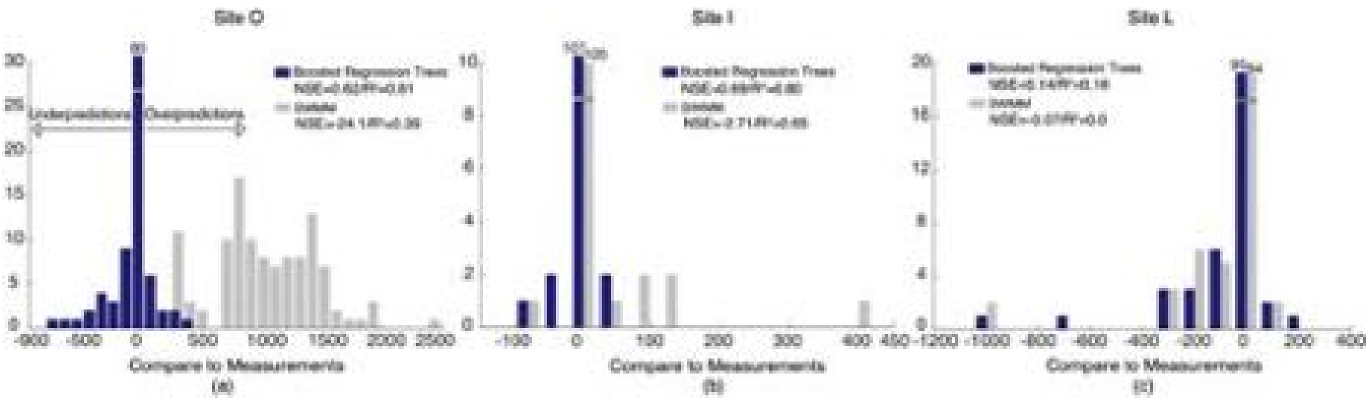


Fig. 4. Histogram of the difference between the combined sewer overflow volume measurements (million liters, ML) to predictions made by the Boosted Regression Trees and the numerical model for (a) site O, (b) site I, and (c) site L, from May and October from 2013 to 2015. Values were obtained by calculating the difference between each prediction and measurement. Similar plots for all other sites are included in the supplementary information section of this paper.

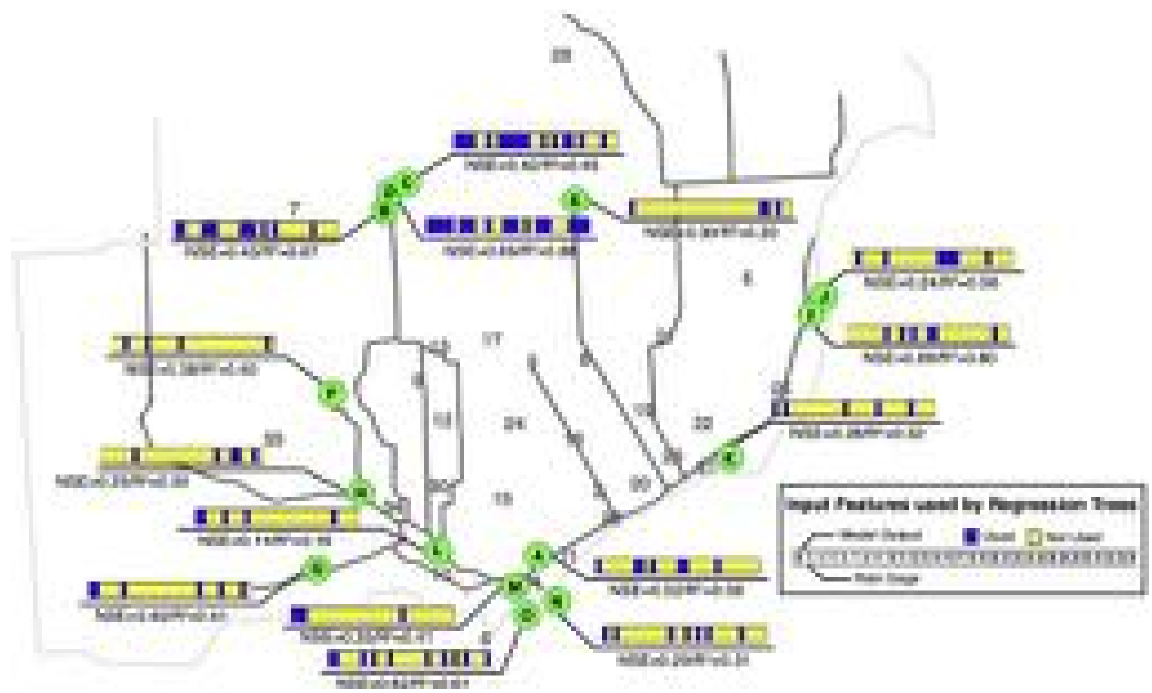


Fig. 5. Evaluation of Boosted Regression Trees performance, showing the fit metrics obtained for each site. The boundary of the service area is outlined in light gray, the Detroit sewer collection system is marked in dark gray, the locations of rain gages are marked by numbers. The input features used by the algorithm (selected by Directed Information criterion) are indexed in the bar connected to each site. The color-coded bar indicates which features are selected (blue) and which are not (yellow). The first element (M) indicates if the output of the SWMM model was used as an input to the regression trees, while the other elements indicate the number of the rain gage. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

data sources to “nudge” the outputs of the numerical model to match observations more closely.

Even when the outputs of the numerical model show strong bias or inaccuracy, they may still prove useful when used as inputs to *Boosted Regression Trees* approach. Numerical models, even those that could be considered “out of date”, still embed a lot of information and domain expertise. For example, a numerical model that may not be correct in regard to absolute flow values, may still be correct in regard to timing of flows and their relative magnitudes. This was the case for site O (Fig. 4a), where the outputs of the SWMM model were heavily biased, but were nonetheless selected as the most causal feature. In such cases, the role of *Boosted Regression Trees* is analogous to correcting these biases by using additional sources of sensor data. By extension, if the underlying numerical model improves as the result of better model inputs or improvement of model structures, or another model becomes available in the future, the data-driven approach should immediately benefit since it does not need to be altered to account for these changes.

As illustrated, the use of more data does not necessarily lead to better predictions. This important point appears to run counter to conventional wisdom on water data, which often assumes that data-driven techniques can arrive at the best answer by ingesting and optimizing around as much data as possible. Rather, ensuring statistical causality between inputs and outputs is important. The use of *Directed Information* provides a reliable and automated way to accomplish this. In our case study, when input features were selected using the *Directed Information* criterion, the performance

Table 2
Measures of model fitting and overfitting by NSE and R^2 values for all outfalls.

Site	DI Test	NSE(tr)	NSE(val)	R^2(tr)	R^2(val)
A[a]	Yes	0.55	0.52	0.93	0.58
	No	0.56	0.5	0.94	0.57
	Improvement	−0.01	0.02	−0.01	0.01
B[a]	Yes	0.92	0.43	0.94	0.67
	No	0.97	0.32	0.98	0.33
	Improvement	−0.05	0.11	−0.04	0.34
C[a]	Yes	0.76	0.42	0.82	0.45
	No	0.82	0.41	0.88	0.47
	Improvement	−0.06	0.01	−0.06	−0.02
D[a]	Yes	0.94	0.69	0.95	0.88
	No	0.93	0.68	0.95	0.88
	Improvement	0.01	0.01	0	0
E[a]	Yes	0.39	0.3	0.43	0.3
	No	0.45	0.19	0.53	0.34
	Improvement	−0.06	0.11	−0.1	−0.04
F	Yes	1	0.38	1	0.4
	No	1	−0.38	1	0
	Improvement	0	0.76	0	0.4
G[a]	Yes	0.65	0.4	0.75	0.41
	No	0.71	0	0.83	0.11
	Improvement	−0.06	0.4	−0.08	0.3
H	Yes	0.67	0.25	0.83	0.3
	No	0.73	−0.37	0.91	0.09
	Improvement	−0.06	0.62	−0.08	0.21
I	Yes	1	0.69	1	0.8
	No	1	0.59	1	0.76
	Improvement	0	0.1	0	0.04
J	Yes	0.46	0.24	0.72	0.56
	No	0.49	0.2	0.74	0.62
	Improvement	−0.03	0.04	−0.02	−0.06
K[a]	Yes	0.56	0.28	0.74	0.32
	No	0.64	0.35	0.84	0.38
	Improvement	−0.08	−0.07	−0.1	−0.06
L[a]	Yes	0.69	0.14	0.77	0.16
	No	0.73	0.15	0.81	0.17
	Improvement	−0.04	−0.01	−0.04	−0.01
M[a]	Yes	0.98	0.35	0.98	0.41
	No	1	0.28	1	0.31
	Improvement	−0.02	0.07	−0.02	0.1
N[a]	Yes	0.58	0.29	0.76	0.31
	No	0.62	0.26	0.82	0.28
	Improvement	−0.04	0.03	−0.06	0.03
O[a]	Yes	0.84	0.62	0.88	0.61
	No	0.87	0.52	0.9	0.52
	Improvement	−0.03	0.1	−0.02	0.09

[a] Model output from the site is selected as an influential feature by DI test; DI test (NO) means all the candidate features are used for training and validation rather than the influential ones selected by DI test (Yes); NSE(R^2)(tr) and NSE(R^2)(val) are NSE(R^2) values for model training and validation; Improvement indicates the difference of NSE(R^2) values w and w/o DI test.

of the *Boosted Regression Trees* improved. In many cases, only half or fewer of the available data sources were actually selected for use in the predictions.

The role of the *Directed Information* step in improving predictive performance may be best explained when interpreting the results of the training phase of the *Boosted Regression Trees* (Table 2). The use of more input features may lead to an improved fit during the training phase since more data are available to explain the variability in the target variable (Ng, 1998). However, some of this variability may only be temporary or the inputs may not exhibit causality with the target variable. As such, strong fit during training may lead to worse predictive performance during the validation phase since the predictive algorithm becomes sensitive to non-informative inputs. As opposed to selecting all possible input features, the chance of overfitting during training will thus be reduced when using only informative inputs. While this will lead to a reduction of fit during the training phase, it will often translate to an improvement in fit during the validation phase, as seen in our study.

This result suggests that the concept of model complexity should be considered more broadly. Complexity of a model is often tied to notions of model structure. As suggested by our case study, when modeling water systems the amount of input data used should also be considered, where more input data may lead to overfitting if not screened ahead of time. As such, the temptation to use all available data when training data-driven water models should be accompanied with a keen appreciation of unintended overfitting.

The SWMM model output was selected as an influential feature by the *Directed Information* algorithm in the majority of the study locations. This is intuitive since the numerical model does embedded a significant amount of information regarding the connectivity and nonlinearities of the system. However, aside from *Directed Information*-based causality, no clear physiographic features explain why some rain gages were selected over others for use in the prediction (Fig. 5). Neither gage proximity to the modeled site nor connectivity via the drainage system were identified as factors that could explain why one gage may have been selected over the others. The challenge in identifying informative gages without the use of a tool such as *Directed Information* may be rooted in the operational complexity of the Detroit sewer collection system. As one of the largest combined sewers in the world, this system contains a large number of control points, in the form of pumps, gates, and valves, which are represented in the numerical model but often operated based on operators' discretion. As such, stochastic uncertainty is embedded in measurements of flow, which limits the ability to deterministically trace the inputs of any given rain input. As such, many observations of flow may thus often be explained by statistical relationships between the input and output data. This, however, plays to be the strength of our approach, which blends statistically-, physically-, or numerically-based mappings.

Through this case study, a number of requirements become apparent when assessing the ability of our approach to work across other systems. First and foremost, the approach will benefit from as many input data as possible – not all, of course, of which will be used. This will improve the likelihood of finding locations that will be causal with the output. Since the *Directed Information* pre-processing step is computationally efficient, ingesting many data inputs can be conducted seamlessly. Once the most informative features are selected, the length of the time record and variability in the output measurements will become important. In general, the time record is a proxy for number of available training storms. While having more storm observations is always better to capture any statistical variability, the size of the storms plays an important role as well. A short time record (a few months or less) will suffice in training the algorithm if the output signal shows a proportional response to a broad range of inputs. For example, site A (Fig. 3) shows a proportional response to a large number of storms, which allowed the *Boosted Regression Trees* to explore a broad output space. The length of the time record will become important especially when predicting sewer overflows. Unlike continually measured sites (e.g. site A), which generally exhibit many non-zero flows, measurements of overflow will primarily be populated with many zero-flow observations. This highly nonlinear behavior challenges a data-driven prediction algorithm because many rain inputs may not be large enough to cause any response. For some sites, large storm events may be less frequent and may thus results in few, but highly variable outputs (e.g. site L). In these instances, the *Boosted Regression Trees* does not have enough relevant training data unless a longer time record is available. For very few sites, this may require years of observations, which were not available in our study. This data requirement does not, however, change the implementation of our toolchain, as more data can simply be ingested as they are measured.

Another important consideration when applying this approach relates to the temporal granularity of predictions. In our study, daily flow and volume measurements were available. This placed a bound on the temporal resolution of the predictions. This daily resolution still has utility in our case study since Detroit's sewer system is one of the largest in the world. The system dynamics play out over relatively long time scales, as can be seen in Fig. 3. Storms often lead to flow responses or overflows that can last multiple days, which means that daily forecasts have utility in treatment planning, collection system dewatering, and overflow operations. Higher-order dynamics are obscured or averaged at such resolutions, which may be important for some smaller sewer system or other applications. If higher resolution forecasts are desired, our approach would require higher resolution flow data. While an additional analysis would be required to assess performance across these time scales, the toolchain could be applied to these data without requiring modifications. This is particularly true about the *Directed Information* step of the approach, which can still be used to determine which input-output relationships are casual. Depending on the dynamics in higher resolution measurements, the boosted regression trees could be replaced with more dynamically-based predictive approaches. This presents good opportunities for future studies, which will be carried out across smaller and more rapidly changing systems.

Overall, the approach presented in this study stands to provide a number of benefits to decision makers and modelers. From an operational perspective, the toolchain provides an automated method by which cities and municipalities can leverage all their existing and emerging data sources to improve forecasts of flows. This will be particularly useful in operational situations where an existing numerical model may not provide sufficiently fine-grained warnings of floods or impending overflows. However, the resulting predictive model should not be used for infrastructure planning purposes (changing pipe diameters, evaluating new designs, etc.) since all relationships are statistical and inherent to the data of the existing system. For these purposes, the numerical model will rather need to be updated and recalibrated. To this end, our approach can also serve as a tool to guide model calibration. The use of *Directed Information* can serve as an alternative to traditional metrics, such as NSE or R^2, providing insight not only on fit but rather the causality between the model structure and observation. The *Directed Information* criterion could then be used to help determine which inputs may be most important to the numerical model, which may reduce amount of inputs and time spent on calibration.

Finally, the *usefulness* of any particular data source must be viewed holistically. A non-causal relationship may suggest that an input data source is not relevant or of sufficient quality to explain a particular output. However, utility and causality are two-way properties. Namely, a flow measurement (the output) may not be informative to begin with. If this is the case, the inputs may still be useful for forecasting at other locations. A level of user discretion should thus be exercised when evaluating input and output pairings. The *Directed Information* step will help in this regard, as it will provide a first check to determine if certain input-output pairings should even be considered before a predictive model is constructed. Ultimately, the quality of the final prediction, which can be evaluated using more classic fit metrics or specific requirements of an application, will remain a good proxy for utility of any particular forecast and thus, implicitly, the utility of the underlying data. While not conducted in this case study due to the temporal granularity of data and the assumption that SWMM model captures upstream dynamics, future studies could also consider the value of flow measurements as predictors for other flow measurements.

7. Conclusions

This paper introduced a holistic, data-driven toolchain based on *Directed Information* and *Boosted Regression Trees* to provide flow forecasts across urban drainage systems. More broadly, this methodology should also work well for other types of water systems where many data or numerical models are available. It was demonstrated that the use of more data is not always advantageous and may often lead to worse predictions. Rather, a computationally-efficient pre-processing step (*Directed Information*) will be important in selecting only those input data that are informative to the overall prediction. The approach based on *Boosted Regression Trees* was also shown to be effective at learning complex and non-linear mappings between rainfall inputs and flow. More importantly, it was demonstrated that the outputs of a numerical model could also be used as an important input to the data-driven approach. Even if a numerical model is no longer fully calibrated due to aging or changes in the system, it still embeds valuable information that can improve the predictive performance of the regression trees. This will provide a rapid and automated way for city managers to use a diverse set of information, which may be at their disposal, without requiring the often-expensive recalibration of numerical models. Naturally, if a numerical model does improve, so will the predictions of our approach. This discovery will be important as the push for “smart” water systems and “big water data” continues. Future work should be carried out to determine how consistent our findings are across other study areas and other types of water systems.

Acknowledgements

This work was funded by the Fred A and Barbara M Erb Family Foundation grant number 903 and the University of Michigan Graham Sustainability Institute. We appreciate the SWMM model and data provided by the Great Lakes Water Authority.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.watres.2018.09.009.

References

Caruana, R., Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 161–168.
Doherty, J., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. Watermark Numerical Computing.
Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36 (1), 27–46.
Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77 (4), 802–813.
Field, R., Tafuri, A.N., 2006. The Use of Best Management Practices (BMPs) in Urban Watersheds. DEStech Publications, Inc.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica J. Econometric Soc. 424–438.
Grünwald, P.D., 2007. The Minimum Description Length Principle. MIT press.
Kerkez, B., Gruden, C., Lewis, M., Montestruque, L., Quigley, M., Wong, B., et al., 2016. Smarter Stormwater Systems. ACS Publications.
Konrad, C.P., 2003. Effects of Urban Development on Floods. US Geological Survey.
Kramer, G., 1998. Directed Information for Channels with Feedback. Eidgenossiche Technische Hochschule Zurich.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.
Maccoux, M.J., Dove, A., Backus, S.M., Dolan, D.M., 2016. Total and soluble reactive phosphorus loadings to Lake Erie: a detailed accounting by year, basin, country, and tributary. J. Great Lake. Res. 42 (6), 1151–1165.
Marko, H., 1973. The bidirectional communication theory–a generalization of

information theory. IEEE Trans. Commun. 21 (12), 1345–1351.
Morales, V.M., Mier, J.M., Garcia, M.H., 2017. Innovative modeling framework for combined sewer overflows prediction. Urban Water J. 14 (1), 97–111.
Ng, A.Y., 1998. On Feature Selection: Learning with Exponentially Many Irrevelant Features as Training Examples. Massachusetts Institute of Technology.
Paquier, A., Mignot, E., Bazin, P.-H., 2015. From hydraulic modelling to urban flood risk. Procedia Eng. 115, 37–44.
Quinn, C.J., Kiyavash, N., Coleman, T.P., 2015. Directed information graphs. IEEE Trans. Inf. Theor. 61 (12), 6887–6909.
Santini, A., Brink, P., Sherman, B., TenBroek, M., 2001. An equivalent rainfall technique to identify dry weather flow data for city of Detroit flow balance analysis. *Models and Applications to urban water systems*. Monograph 9, 259.
Schalkoff, R.J., 1997. Artificial Neural Networks, vol. 1. McGraw-Hill, New York.
Sun, N.-Z., Sun, A., 2015. Model Calibration and Parameter Estimation: for Environmental and Water Resource Systems. Springer.
Tenbroek, B.M., Bunyan, R.J.T., Whiting, G., Redman-White, W., Uren, M.J., Brunson, K.M., Edwards, C.F., 1999. Measurement of buried oxide thermal conductivity for accurate electrothermal simulation of SOI device. IEEE Trans. Electron. Dev. 46 (1), 251–253.
Wald, A., 1950. Statistical Decision Functions.