The potential of knowing more – a review of data-driven urban water management

Sven Eggimann$^{1,2}$*, Lena Mutzner$^{1,2}$, Omar Wani$^{1,2}$, Mariane Yvonne Schneider$^{1,2}$, Dorothee Spuhler$^{1,2}$, Matthew Moy de Vitry$^{1,2}$, Philipp Beutler$^{1,2}$, Max Maurer$^{1,2}$

$^1$ Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland.

$^2$ Institute of Civil, Environmental and Geomatic Engineering, ETH Zürich, 8093 Zurich, Switzerland.

TOC/Abstract Art

Abstract: The promise of collecting and utilising large amounts of data has never been greater in the history of urban water management (UWM). This paper reviews several data-driven approaches which play a key role in bringing forward a sea change. It critically investigates whether data-driven UWM offers a promising foundation for addressing current
challenges and supporting fundamental changes in UWM. We discuss the examples of better rain-data management, urban pluvial flood-risk management and forecasting, drinking water and sewer network operation and management, integrated design and management, increasing water productivity, wastewater-based epidemiology and on-site water and wastewater treatment.

The accumulated evidence assembled from research documented in the literature points towards a future UWM that offers significant potential benefits thanks to increased collection and utilisation of data. The findings show that data-driven UWM allows us to develop and apply novel methods, to optimize the efficiency of the current network-based approach, and an extended functionality of today’s systems. However, generic challenges related to data-driven approaches (e.g. data processing, data availability, data quality, data costs) and the specific challenges of data-driven UWM need to be addressed, namely data access and ownership, current engineering practices and the difficulty of assessing the cost benefits of data-driven UWM.

1 Introduction

One of the earliest powerful demonstrations of how increased data availability can help to transform urban water management (UWM) is the work of John Snow in 1854.1 With the aid of precise spatial data about cholera victims in London, Snow provided evidence to support the hypothesis of a drinking-well being the source of the outbreak. His data-driven approach illustrates how the availability and interpretation of data can be for public hygiene, which is one of the essential UWM services along with the provision of safe drinking water, protection against flooding and water pollution control. Since the middle of the 19th century, major improvements in these services were made in many places in the world. However, today’s challenges in UWM have not been solved on a global scale.2,3
The various challenges (e.g. population change, ageing infrastructure or economic change) facing national water and wastewater infrastructures, bring to light an urgent need for innovation and a change in the UWM framework to provide more sustainable UWM services. Especially in the industrialised world, many challenges are endemic to the traditional and currently prevailing approach to UWM, which can be traced back to antiquity. This approach is based on the fundament of centralised conveyance infrastructure that provides drinking water for and evacuates waste and storm water from urban areas. In the majority of cases, this infrastructure is built to passively provide a defined transport capacity at a given peak load and has little flexibility to adapt to new demands. The most widely discussed changes affecting UWM infrastructure, such as urban development or climate change, albeit generally slow processes, can still cause inadequate system performance within the long lifespan of UWM infrastructure. Further consequential downsides of centralised infrastructure are its strong dependence on large water quantities, its vulnerability to excessive rainfall, high investment costs and vast and complex pipe networks.

Increasing evidence questions whether and in what form the prevailing UWM practice can be the best solution for the world, as it has been since the beginning of the 20th century. Today, both existing UWM services as well as the absence of such services are incurring increasing economic, social and environmental costs, even in countries that have successfully built up a functioning UWM infrastructure.

The aim of this review is threefold: First, to explore the potential of increased data availability to tackle today’s challenges, bringing about new and fundamental changes in the way UWM can be provided by data-driven UWM (Figure 1). Second, to review key challenges of data-driven UWM. Third, to relate data availability and its utility in UWM on a schematic...
basis. Data-driven infrastructure transformation is not limited to UWM, but is happening also in other infrastructure sectors such as e.g. the energy sector.\textsuperscript{18} Despite the potentially valuable insights that could be gained from a thorough comparison of UWM to other sectors, such an enterprise is beyond the scope of the current review.

Clearly, more and better data alone are not sufficient to solve today’s many UWM problems. However, the argument will be made that data are a necessary precondition for addressing some of the most pressing problems, be it by increasing efficiencies or by enabling novel and ‘smarter’ approaches.\textsuperscript{19,20,21} On the basis of the available literature, we will show that data-driven UWM, i.e. using different measurement techniques and collecting large volumes of data, carries the potential to expand the functionality of the UWM system beyond its traditional conveyance purpose.
Figure 1. Scope of UWM as illustrated by example variables of interest that will be affected by the current data and sensing revolution.

2 The current revolution in data and sensing
Whereas in the times of Snow, data collection was rudimentary and tedious, in today’s information age it is possible to generate and manage enormous amounts of data relatively cheaply and easily. This change has been mainly driven by UWM-independent advances in analytics, sensing, transmission, computing and data management. This section only touches on recent advances and does not provide a comprehensive overview of the literature:

Analytics and sensing: Data can originate from various sources and rely heavily on measuring devices. Advances in measurement techniques have made huge leaps in the last few years. Three different trends in the production of data can be observed: First, novel sensors and measurement devices have been developed; second, measurements devices are available at increasingly lower cost; and third, thanks to miniaturization, sensors can be mounted on mobile platforms. Examples in the field of UWM are autonomous drones used for surface characterisation, sensor mounted on mobile platforms floating downstream networks, fast counting and identification of bacteria by flow cytometry and lab-on-a-chip biosensors. However, sensing capabilities have been enhanced not only by direct sensing technology, but also by soft sensing via software-based data processing. Furthermore, today’s spatially enabled societies and ‘smart cities’ are increasingly equipped to actively or passively collect data on a voluntary or non-voluntary basis, and constitute thereby a fundamentally new type of resource. Examples of these new and upcoming data sources are crowdsourcing, using microwave links from telecommunication networks to identify rain intensities, harvesting social media or surveillance data to identify urban flood events.
Transmission: Independently of UWM, major investments are being made in various data transmission technologies and networks. They range from mobile communications, smart metering networks to the ‘internet of things’ and specific sensor networks. This phenomenon is highly relevant, as the possibility of piggybacking on these existing transmission infrastructures substantially lowers data costs.

Computing and data management: Ubiquitous computing, automation and efficient data management have advanced by several orders of magnitude in the past decade. Computational power and data storage are becoming cheaper and data processing techniques are becoming much more powerful.

Various aspects of our daily lives have been deeply influenced by this ongoing revolution in data and sensing. In the following, we show that adopting a more data-driven approach offers the potential to enable radically different practices in UWM and serves as a precondition for a shift towards novel and possibly more sustainable UWM services.

3 Approaches to data-driven UWM
A number of data-driven approaches appear promising for initiating fundamental changes in UWM. In the following sections, we discuss these approaches with the help of key examples that illustrate the relevant challenges and opportunities in UWM. Although additional examples could potentially be identified, we argue that the selection represents decisive trends and shows the potential of data-driven UWM in an exemplary way.

3.1 Spatial and temporal variability of rainfall
Rainfall is one of the main agents that influences UWM. It determines the number and duration of sewer overflows or the hydraulic load of the drainage system and to a certain
extent also the performance of wastewater treatment plants (WWTP). Thus, challenges such as flooding of urban areas, sub-optimal WWTP operation and pollution of water bodies through sewer overflows are directly linked to rainfall data. Rainfall is highly variable in nature and varies significantly over time and space. One of the main hurdles to achieving smarter management of sewer systems is thus our limited ability to accurately measure and predict rainfall at relevant spatial (generally sub-kilometre) and temporal (around minute-resolution or less) scales.

There has been a surge in improving conventional rainfall measurement techniques and introducing novel sensing methods to account for the spatial variance of rainfall. The design of point measurement instruments like rain gauges has been improved, e.g. electronic floating-device rain gauges that capture rainfall intensities better, or the rain gauges have been supplemented with more sophisticated instruments such as disdrometers providing the drop size distribution of rainfall. Land and satellite-based radar has been increasingly used to provide a more spatially explicit picture of rainfall. Weather radars designed specifically for urban application have a shorter range and provide a finer resolution of rainfall measurements. Additional information on rainfall can be acquired by measuring the signal attenuation in telecommunications microwave links; the data of such commercial links have been used from several urban catchments, e.g. in Czech Republic, the Netherlands and Switzerland, to obtain path-averaged rainfall intensities. Furthermore, new attempts have also been made to increase the density of point measurements by crowdsourcing.

While new measurement techniques, such as radar systems and microwave links, are addressing the demand for finer spatio-temporal resolution of rainfall data, the measurement
accuracy of these devices is still somewhat insufficient and needs improvement so that their full potential for UWM can be harnessed.\textsuperscript{52,53}

The big challenge however, is integrating all these different sources of rainfall data into a coherent and usable data collection. Measurement and interpolation uncertainties need to be adequately considered,\textsuperscript{54,55} and large amounts of data need to be processed. Significant work is being done to merge rain-gauge and radar data to produce more reliable rainfall fields.\textsuperscript{56,57,58} Although rainfall measurements are of major importance to smarter management of urban drainage systems the ability to efficiently control urban drainage systems, needs to be addressed likewise (see Section 3.3). Additionally, the flexibility to control the drainage system is strongly constrained by the capacity of the existing infrastructure.

\textbf{3.2 Urban flood risk management}

Urban pluvial floods potentially have a high social and economic impact, but they are by their nature one of the least predictable challenges facing UWM.\textsuperscript{59} Consequently, flood-risk management can also be described as ‘\textit{a process of decision making under uncertainty}’.\textsuperscript{60} Additionally, urban flooding is expected to grow in importance in line with climate change and urbanisation.\textsuperscript{61} Besides infrastructure planning and flood risk assessment, early warning systems are also becoming more and more relevant for flood-risk management.

Hydrological and hydraulic modelling is the basis for urban pluvial flood-risk management, and relies on different forms of input data: i.) \textit{layout data} allows the topography of the model to be established, ii.) \textit{rainfall and flood data} represent observed or anticipated rainfall events and the system's reaction to the rainfall, respectively.

i. \textit{Layout data} consists of information about the drainage infrastructure and the terrain (e.g. elevation, surface characteristics). Whereas information about the drainage
infrastructure is often considered as given but may not be available with the desired accuracy and completeness.\textsuperscript{62} Terrain data are more widely available. The impact of elevation data resolution is however essential: For example, a certain resolution (~2-5 meters raster cell size in general) and accuracy is required for the realistic representation of surface flow between buildings or along the side of roads.\textsuperscript{24,63} Today, elevation data can be easily collected at very high resolutions due to advances in remote sensing, and their application to urban flood modelling has been shown to potentially improve flow path delineation.\textsuperscript{64} However, their use is still not widespread for two reasons. First, the computation time for larger catchments is inhibitive. Second, urban environments contain a large variety of mobile and temporary elements, such as vehicles or construction sites, which are visible in very high-resolution remote sensing data but are not modelled adequately.\textsuperscript{65} Relevant surface characteristics such as perviousness and roughness are typically estimated by classifying aerial images into land use types and using lookup tables to assign surface characteristic values to each land use class. Recent land use classification methods perform very well even with minimal human supervision, but the incremental gains at the level of the modelling results are not significant, in part because of the integrative nature of the rainfall-runoff models.\textsuperscript{66} While the spatial distribution of surface characteristics may be sufficiently resolved, we see development potential for methods that allow more direct inference of surface characteristics, in lieu of land use lookup tables.\textsuperscript{67}

\begin{enumerate}
\item[ii.)] \textit{High-resolution rainfall and flood data} are particularly important for urban flood modelling. Flood data are particularly difficult and costly to obtain, which in combination with the rarity of flood events, explains its scarcity.\textsuperscript{68,69} This has motivated the exploration of new measurement techniques such as robust and cheap binary
Moreover, researchers are now looking into social media as a platform for obtaining such data, for example by using Twitter to collect flood information for a decision support tool to decision-makers in Jakarta, or by using YouTube videos and crowdsourcing to estimate flash-flood volumes to better understand flood risks. While the above-mentioned data collection modes are still in a phase of development and do not solve the issue of flood event rarity, we find in the examples that they can provide valuable information about flood events that are challenging to measure with conventional sensors.

Detailed and accurate flood models will potentially provide improved flood hazard maps. By using historic rainfall and flood data, the uncertainty of model prognosis can be reduced through calibration. A Bayesian framework also allows this reduction to be expressed formally. Many studies have also shown clear increase in the performance of hydrological models with the increase in the layout detail and the quality of input data. Also, it has been demonstrated that for proper parameter estimation of models, longer time series of flood data, which captures more flood events, is desirable. A very attractive application of flood models is that of early flood-warning systems. The usefulness of a hydrological model for an early flood-warning system depends on how far ahead in time (lead time) a reasonably accurate prediction can be made. However, the prediction uncertainty of models increases with increasing lead times. Improved flood modelling will lead to higher lead times without a significant drop in the prediction accuracy.

It is important to stress that the increased availability of the above mentioned data can only improve the well-being of society (e.g. safer cities, fairer insurance policies or better flood
evacuation measures) if the required technologies and engineering methods are adopted by practitioners and society (cf. Section 4.1).

### 3.3 Drinking water and sewer network operation and management

Massive resources were invested to construct today’s UWM network infrastructures. However, investments are increasingly falling short, making it essential to increase the efficiency of existing systems. Only limited knowledge is currently available about the performance of water supply and sewer networks at any given moment in time and across the whole network. There is a strong and growing literature that supports the notion of using performance indicators, including customer satisfaction data, to guide and improve infrastructure management. In this review, we are assuming that this is the current state of the art and will explore more advanced possibilities of data collection and the potential of novel data for UWM.

In the following, we outline novel approaches aiming to improve network management and operation, namely i.) real-time control (RTC) modified with model predictive control (MPC), ii.) smart (water) metering, iii.) structural (pipe) health monitoring, and iv.) quality control through detecting source contamination.

i.) RTC and MPC allow us to go beyond considering sewers as passive transportation infrastructure since they can be used to increase the efficiency of existing infrastructure, thereby postponing or eliminating the need for investments in completely new infrastructure. RTC\(^{82,83}\) makes storm water systems ‘smart’\(^{23}\) and significantly improves utilisation of the existing drainage infrastructure.\(^{84}\) The release of untreated sewerage into receiving water bodies during rain events can be reduced by dynamically controlling the flow and retention volumes in sewers with sensor networks and automated valves, for example.\(^{85}\)
The high temporal and spatial variability of pollution flow would require RTC to distinguish between highly polluted and less polluted flows. Pilot studies show promising results using conductivity and turbidity sensors as real-time surrogates for pollution potential.\textsuperscript{86} Such information coupled with RTC, in the near future, would allow storm water discharge to be managed on the basis of the impact potential on the receiving water body.

\textit{MPC} is an advanced RTC technique, in which the optimisation is based not only on the knowledge of the current state of the system but also on its forecast state. Thus, MPC allows us to improve the monitoring process and to optimally utilise the storage capacities of rainwater reservoirs, detention ponds and in-sewer storage volumes by considering anticipated rainfall (thus e.g. reducing sewer overflows). Advanced control logic is being used in field studies to regulate storage capacities with information from water level sensors and real-time weather forecasts to create dynamically controlled rainwater reservoirs.\textsuperscript{87,88,89} Model results suggest that up to 92\% of releases to the combined sewer could theoretically be reduced with dynamically controlled cisterns.\textsuperscript{87}

The computational resources required for RTC and MPC have significantly improved in the last decade\textsuperscript{90,91} and studies confirm the robustness of MPC algorithms for complex (non-linear) systems\textsuperscript{92,93}

ii.) \textit{Smart (drinking water) metering}\textsuperscript{94,95,96} opens up promising new approaches to network management which rely on the availability of detailed water-related activity data from the customers. Smart metering allows frequent and high-resolution use patterns to be recorded at different positions in the network.\textsuperscript{97} Case studies suggest potential residential water use reductions of at least 10\% and monthly peak demand reductions of 10\%.\textsuperscript{96} Reduced water demand and peak water demand alleviation also means that
downsized water supply infrastructure could be built, which translates into potential capital cost savings were recently estimated to be between 11% and 51%. Excessive water usages, breakages, abuse or water theft, can then be detected anywhere in the network. Additionally, water end-use models can be created and the quantification of wastewater production improved. Another attractive feature of smart metering is its promise to switch from pure supply to demand management. This would allow the available infrastructure to be utilised more efficiently by influencing customer demand based on capacity. Examples include differentiated tariff structures designed to reduce peak hour and day demand by ‘peak’ or ‘drought pricing’. The alignment of prices along free market lines in real-time has been predicted to be especially effective for outdoor water use given its greater price elasticity. However, less research has focused on domestic indoor water use, where price elasticity is low. One option might be to provide customers with direct feedback, e.g. through on-site displays or by wireless communication via mobile phones or email. However, it is unclear how effective such measures are in the long run, i.e. there is little evidence how smart metering will affect long-term consumption behaviour.

Furthermore, most applications of smart metering to date have been proposed in the context of industrialised countries and less applications can be found for low-income markets where a critical set of enabling factors need first to be addressed. Nevertheless, smart metering is not restricted to network-based water delivery and has already been applied in a pilot field study in the context of developing countries, e.g. by equipping hand pumps with sensors to monitor water levels and charge customers adequately.
iii.) **Structural (pipe) health monitoring** addresses UWM challenges such as groundwater contamination, local flooding, sinkholes and high water consumption. Pressure or vibration sensors and microphones are available to allow network management to be improved by remotely collecting data and thus monitoring the structural health of sewer and drinking water pipes.\(^\text{110,111}\) The miniaturization of sensors and the possibility of mounting them on mobile platforms are driving new network exploration methods (e.g. inline mobile sensor technology).\(^\text{112}\) Furthermore, we are witnessing the apparition of alternative monitoring approaches that rely on crowdsourcing to gather information about the network (e.g. water leaks or water levels) via smartphones.\(^\text{113}\) Increased knowledge about the condition of sewer and water-supply networks contributes to reduced water loss and prevents local flooding or sinkholes e.g. by automatic leak detection,\(^\text{114,115,116}\) and enables improved rehabilitation planning and modelling thanks to reduced uncertainty about pipe conditions.\(^\text{117}\)

iv.) **Quality control through source contamination detection** is essential in view of threats to human and environmental health due to discharges of untreated wastewater into receiving waters and wastewater intrusion into storm drainage and drinking water resources through leakages, wrong connections or intentional contamination (compare Section 3.5 in case of water reuse).\(^\text{118,119}\) However, it remains a challenge to monitor water quality and especially to identify the precise source of the pollution, as wastewater and water supply networks are highly distributed, have many entry points and have to deal with a wide variety of potentially harmful contaminants.

Reliable water quality control and rapid detection of contamination sources is particularly vital in water supply networks. It could be achieved by real-time monitoring of selected contamination indicators and optimal sensor placement – two highly topical
Approaches relying on simple indicators, such as a distributed temperature sensor cables, have been successfully used to identify incorrect wastewater connections to storm water sewers. Panasiuk et al. conclude in a comprehensive review of human waste monitoring methods that the most promising approaches are based on chemical indicators and biological markers (cf. Table 1 in Section 3.6 for potential wastewater indicators). Nevertheless, further research is needed on innovative sampling methods, such as online-monitoring or passive sampling. Additionally, the monitoring of biological markers via e.g. quantitative polymerase chain reaction requires further development before being field ready. Contamination warning systems in water supply systems that combine real-time monitoring data with crowdsourcing using customer feedback on water quality via smartphones have been tested but require extended pilot field studies before implementation. Further, remote mobile sensors moving along the pipe with the flow could increase the likelihood of detecting contamination and fully-functional prototypes should be available in the near future.

The data-driven approaches above outlined allow the management and operation of network-based infrastructures to be improved by saving water and protecting environmental and human health. Generally, the main aim is to minimally delay premature rehabilitation (e.g. by identification of pipe health) or even replace the creation of infrastructures (e.g. by switching from supply to demand management). To date, the literature does not provide us with clear proof that such a strategic shift from hardware to data actually leads to the desired savings, flexibility or long term increase in performance. However, the cited literature does show many promising examples that these hopes might be justified.
3.4 Water productivity

Urbanisation in combination with climate change is expected to influence the spatial pattern of water availability and demand in complex ways. One related challenge is enhancing efficiency and optimising water allocation and usage in order to address water scarcity, especially in growing cities and regions affected by drought. The successful construction of water sensitive cities relies on increased data availability via measurement devices. In order to use less water, water productivity needs to be improved, i.e. the value of goods and services produced per unit of water used. Water can be saved in many different manners: by conservation, efficiency, sufficiency, substitution, reuse, recycling or harvesting. Not all of these water saving strategies can be influenced by data-driven approaches to the same degree: For example, the practice and acceptability of water conservation is generally culturally dependent. In the following, we therefore only outline strategies for which the availability and leveraging of data plays a significant role: i.) increasing water use efficiency ii.) reusing and reclaiming water and iii.) rainwater harvesting.

i.) Increasing efficiency: The simplest way to increase water productivity is to use less water for the same service or to incentivise and inform on water-saving behaviour, e.g. by means of smart metering (cf. Section 3.3). Another data-driven way to reduce water consumption is ‘smart irrigation’, where irrigation systems are operated on the basis of humidity sensors, soil moisture sensors and real-time weather forecast data to allow precise watering depending on actual watering needs. Whereas smart irrigation is driven mainly by the agricultural sciences, these technologies and approaches are equally suitable for residential areas and are already available on the market.

ii.) Water productivity may be further increased by reusing or reclaiming water. Currently, only 1.7% of the urban water use is reused. The (direct or indirect) potable
or non-potable reuse of water from various sources requires robust and comprehensive monitoring and control systems for risk management to ensure human and environmental health. Successful large-scale implementation beyond niche applications of non-potable water in drought-affected regions such as California is data-driven and will rely amongst many other factors on detailed real-time water quality assessments measuring pathogens and chemical constituents. Whereas software for the timely detection of treatment failures has recently been successfully developed such as the CANARY event detection software, direct potable water reuse still lacks reliable monitoring techniques.

iii.) Finally, water productivity is not increased only by saving water but also by harvesting rainfall. When harvested rainwater is put to a different use (e.g. for toilet flushing, cloth washing, irrigation or drinking after adequate treatment), data are drawn from various sources (e.g. tank-level sensors or rain forecasts). Another further way in which the revolution in data and sensing revolution has innovated rainwater harvesting is by providing better information about estimates of harvesting sites and harvesting potential. Quigley and Braun showed in several demonstration projects that using rain forecasts to control the filling level of retention structures can optimise rainwater utilisation and harvesting without compromising the retention capabilities of the investigated ‘green infrastructure’ elements.

Overall, the data and sensing developments enabling these different approaches promise to improve water productivity and could potentially lead to the acceleration of the trends for increasing water productivity.
3.5 **Infrastructure design and management**

Socio-economic changes frequently outrun the lifespan of existing water infrastructures. Thus estimates indicate that the global urban areas will grow by 60% to 200% between 2000 to 2030. Good design concepts that maintain flexibility and can achieve overall goals over the long term are and will continue to be important. It is not surprising that several such concepts can be found in literature and in practice. Examples are low impact urban design and development (LIUDD), sustainable urban drainage systems (SUDS), water sensitive urban design (WSUD) and integrated urban water management (IUWM), all of which emphasize the importance of a more integrated or holistic approach to UWM.

These approaches demand significant integrated modelling. This quantifies the interaction between sub-systems (i.e. freshwater production, storm water runoff, wastewater treatment) within the wider urban system (e.g. water, energy, solid waste, agriculture) and assesses their performance under changing conditions. However, integrated modelling requires highly consistent data, of the kind commonly processed by geographic information systems. It needs spatially explicit input data of very diverse type and from many different sources (e.g. local climate, soil characteristics, elevation data, groundwater depth, land use, geology, water needs, system performance, water quality, sewer system characteristics etc.). The increased availability of manifold data will therefore allow the reliability of integrated models to be improved.

Highly integrated models play an important role as input for decision support systems (see also Figure 2). More comprehensive ones such as structured decision-making or multi-criteria decision analysis in particular make very high demands on the volume of the data and modelling results. Another interesting approach in this field is the use of artificial intelligence for decision support: the integration of numerical or statistical models achieves higher accuracy and reliability, but imposes a corresponding toll on the amount and quality of the underlying data. Figure 2 indicates the different layers of data demand for modelling
and decision-making. An illustration is provided by ‘green infrastructure designs’ (e.g. green roofs, constructed wetlands, mixed-use storm water collection and infiltration sites) which have a broad range of effects, such as altering the quality or the amount of storm water runoff, effects on groundwater, air quality and ambient temperatures (cf. heat island effect). An Australian example of spatially explicit modelling suggest that 10% of park area allocation may result in 62% of nitrogen reduction from stormwater, and experimental studies show positive effects of vegetated roofs on stormwater retention by reducing rainwater volume (52-95%). However, the influence of green roofs on urban microclimate remains negligible unless they are combined with vegetation areas at street level: integrated models revealed a temperature cooling up to 2°C.

Integrated design concepts and modelling encounter many obstacles. Various obstacles (e.g. conflicting objectives, reluctance in investigative data collection or confidentiality of existing data) to integrated planning have already been described and most of them are still highly relevant today. Nevertheless, the adoption of integrated modelling seems to be progressing as the growing number of modelling tools indicates. Data availability for the calibration of such models is certainly not the only bottleneck, but it is an important precondition for their useful application.
3.6 Wastewater-based epidemiology
Wastewater is a reflection of the society producing it. It contains a mixture of human excreta and substances used in households for daily living. Novel technological opportunities such as quantitative polymerase chain reaction\textsuperscript{123} lab-on-a-chip biosensors,\textsuperscript{27,156} in-situ chemical analysers\textsuperscript{157} or fluorescence spectroscopy\textsuperscript{158} will allow us to harvest information from this complex sewage matrix (almost) in real-time. With these new data sources, the opportunity arises to use wastewater to gain insight into i.) the prevention of epidemics, ii.) the assessment of human health and lifestyle and iii.) the monitoring of drugs via indicator substances (Table 1).

i.) Pathogen monitoring would allow early warning systems for infectious diseases to be implemented. Successful attempts have been made to identify and quantify viruses,
protozoa and bacteria in wastewater. However, a review article about the applicability in practice also concluded that ‘the molecular techniques available today and those under development would require further refinement in order to be standardized and applicable’. 123

The potential benefits could be substantial, as many pathogens are excreted before symptoms of infections appear and can be detected if only a small percentage of a population is infected.159 Monitoring studies show the potential of detecting infectious viruses, for example at the inflow of WWTPs.160,161,162 Improvements in measurement technology, such as high-throughput mass sequencing or biosensors for rapid and on-site monitoring of disease biomarkers, can facilitate the detection process.163 Biosensors hold the promise to develop into a cheap and easy-to-handle option for developing countries, where few suitable laboratories are available and hygiene is a major concern.164,165 However, successful development and implementation require the consideration of various barriers present in real-world settings, and must be field tested (cf. Section 4.1).166

ii.) Next to the prevention of epidemics, the status and change of human health and lifestyle characteristics can be assessed. First research attempts are the monitoring oxidative stress indicators167,168 or the measurement of microbial gut communities in sewage to predict the obesity levels of cities.169 If sewer systems exist, such human health and lifestyle indicators (cf. Table 1) can be identified on different spatial scales, e.g. on a city or single household level. New sensor technologies might one day enable real-time monitoring of vital health parameters in smart toilets, e.g. as described in a patent entitled ‘Toilet device with system for inspecting health conditions’.170 This would enable users to assess their personal state of health and share information via online
networks or report it to physicians. In addition, valuable information about human lifestyle is gained by monitoring household-specific discharge patterns (see Section 3.3).

iii.) By analysing wastewater for drugs (illicit or legal), the variety of knowingly and unknowingly consumed drugs may be assessed. Novel approaches enable tobacco and alcohol usage in sewage\textsuperscript{171,172} to be monitored or drug consumption across European cities to be compared via lab-based analysis of 24-hour composite samples.\textsuperscript{173,174} Such information can be used to develop prevention campaigns or to assess their success for entire populations without depending on strongly biased and time lagged-surveys.\textsuperscript{175} While many real-time sensors for monitoring water quality are still at the research and development phase, the assessment of drug consumption does not necessarily need this data in high temporal resolution.

Sewage is full of valuable information and represents a dynamic picture of human societies in time and space. Monitoring wastewater data in conjunction with spatial and demographic data are particularly promising. Wastewater-based epidemiology has so far been applied mostly in the field of illicit drug consumption, but it has the potential to assess the improvement of overall human health.\textsuperscript{176} For wider applications, there are still open questions to be solved, such as obtaining information about the behaviour and stability of substances and pathogens of interest in the sewers\textsuperscript{177} or the accurate estimation of the number of people at a given time in the sample catchment for normative purposes.\textsuperscript{178} Even more importantly, while the use of chemical monitoring is in a field-applicable phase, further research and development on microbial markers is needed.\textsuperscript{118,123,124} In a review on new microbial markers assays, the authors considered the field of microbial markers as still rapidly evolving.\textsuperscript{124}
<table>
<thead>
<tr>
<th>Category</th>
<th>Category indicator</th>
<th>Selected prominent examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidemic</td>
<td>Virus</td>
<td>Rotavirus, Smallpox, Norovirus, Hepatitis A and E</td>
</tr>
<tr>
<td>Bacteria</td>
<td></td>
<td>Human fecal and antibiotic resistance (e.g. escherichia coli, intestinal enterococcus)</td>
</tr>
<tr>
<td>Health &amp; lifestyle</td>
<td>Oxidative stress</td>
<td>Isoprostanes</td>
</tr>
<tr>
<td>Pharmaceuticals</td>
<td></td>
<td>Carbamazepine, Diclofenac, Antibiotics (e.g. sulfamethoxazole)</td>
</tr>
<tr>
<td>Personal care products</td>
<td></td>
<td>Anti-microbial disinfectant (e.g. triclosan), Nail polish and hair spray (e.g. di-n-butyl phthalate), Laundry and dishwasher detergents (e.g. nonylphenol)</td>
</tr>
<tr>
<td>Nutrition</td>
<td></td>
<td>Obesity indicators (human fecal oligotypes), Synthetic sweeteners (e.g. sucralose, acesulfame)</td>
</tr>
<tr>
<td>Drugs[a]</td>
<td>Legal</td>
<td>Alcohol (e.g. ethyl sulfate), Tobacco (e.g. nicotine, cotinine)</td>
</tr>
<tr>
<td>Illicit</td>
<td></td>
<td>Cocaine, Benzoylecgonine, Heroin, Tetrahydrocannabinol (THC), Amphetamines, 3,4-methylenedioxymethamphetamine (MDMA)</td>
</tr>
<tr>
<td>Other</td>
<td>Surface runoff</td>
<td>Pesticides and biocides (e.g. mecoprop, carbendazim, diazinon, diuron)</td>
</tr>
<tr>
<td></td>
<td>Wastewater marker</td>
<td>Household indicators (e.g. caffeine)</td>
</tr>
</tbody>
</table>

[a] The distinction between legal and illicit is arbitrary and only matters in that there may be alternative ways to quantify the consumption of most legal substances, e.g. sales data.

Table 1. A non-exhaustive selection of prominent examples of wastewater substances measured for retrieving information.
3.7 On-site drinking and wastewater treatment
Most of the examples for data-driven UWM so far address challenges in highly urbanised environments with extensive and complex network infrastructures. Truly impressive as the outlined application examples are, due to limitations of the traditional networked approach, UWM service provision is and will not be network-based everywhere. On-site treatment - also called distributed or decentralised treatment - is so far used for niche applications, in areas with low population densities or fast demographic change, or as a stop-gap solution in densely populated unserved settlements.\(^{182}\) However, on-site treatment has been suggested as a complement or fundamentally different alternative to conventional centralised network-based systems.\(^{13,147,148,183,184}\) We argue that data-driven UWM allows on-site treatment to go beyond niche applications and constitute a widespread customised and flexible engineering solution. We do not contrast on-site and centralised solutions here (cf. Section 4), but review the potential for improving on-site technologies by the revolution in data and sensing. Three main on-site applications can be found in the literature:

i.) On-site or point-of use treatment of water is the most effective way to provide safe drinking water in areas with insufficient water quality.\(^{185,186}\) This approach does not require extensive infrastructure investments and only a fraction of the total water consumption needs to be treated.\(^{187,188}\) The main challenge from a sensing and data point of view is to reliably identify the treatment performance (mainly hygiene, but also contaminants such as arsenic) and to detect the point for rehabilitation or replacement of critical decentralised infrastructure elements.

ii.) On-site water treatment for reuse is an attractive option to increase water productivity (cf. Section 3.5).\(^{187}\) The US-NRC identified several research areas that hold significant potential to advance the reuse of municipal wastewater.\(^{188}\) They include the
identification of ‘better indicators and surrogates that can be used to monitor process performance in reuse scenarios and develop online real-time or near real-time analytical monitoring techniques for their measurement’.

iii.) On-site wastewater treatment. Even though on-site WWTPs are already widely applied in OECD countries, they are often seen as a stop-gap. On-site systems are perceived to have low technical reliability and their overall treatment performance is often unconvincing or unknown. Kaminsky and Javernick-Will however concluded, that ‘system software is more likely to be the root cause of system failure than the hardware itself’, implying that on-site systems are not limited by the available technology but by institutional and organisational factors.

Barriers towards widespread successful implementation of on-site systems are multi-faceted and or a technological transfer demanding (see Section 4.1). However, there are two important barriers for which data-driven approaches are especially promising, namely the lack of information about the treatment performance and the cost of operation and maintenance. The latest developments in sensor technology, data acquisition and computation are already applied to some centralised approaches, but they could also pave the way with respect to performance and costs for on-site systems. In the following, these two issues are specifically discussed for on-site wastewater treatment. However, they also apply in a similar way to on-site water treatment:

Treatment performance: Most on-site treatment plants are scarcely monitored today - e.g. in Germany small-scale WWTP are monitored twice or three times annually. The treatment performance of on-site WWTPs could be improved significantly by more frequent monitoring, which serves as a basis for detecting malfunctioning systems.
monitoring would additionally improve the understanding and optimisation of on-site treatment systems.\textsuperscript{197} Furthermore, improved monitoring techniques may enable the application of more complex and demanding treatment technologies such as autotrophic ammonium oxidation. Continuous monitoring approaches to centralised WWTPs exist, but are not yet sufficiently low-maintenance for direct implementation on on-site WWTPs. First promising tests for on-site systems have been made with SAC\textsubscript{254} (the spectral absorption coefficient at wavelength 254) and dissolved oxygen measurements, while turbidity showed no correlation with biological oxygen demand.\textsuperscript{198} Soft sensing may be another promising way to generate useful information out of easy-to-collect data.\textsuperscript{199,200}

\textit{Maintenance costs:} High maintenance costs are a common problem in massively modular infrastructure,\textsuperscript{201} because of the costs of regular inspections, cleaning, sludge collection or repair.\textsuperscript{202} Their operation will consequently benefit substantially from developments in sensoring and automation.\textsuperscript{203} For centralised infrastructures, lower costs can typically be achieved with scheduled, planned and pro-active maintenance.\textsuperscript{204,205} This generally also holds true for on-site systems: For example, tank-filling sensors could lower the cost of managing sludge and scum collection management by enabling the implementation of scheduled management schemes.\textsuperscript{206} Greater knowledge about maintenance requirements would allow operation and maintenance to be targeted and would optimise the allocation of personnel and resources, i.e. a shift towards pro-active and planned maintenance. Lower maintenance would then lead to lower overall decentralised treatment costs.

The revolution in data and sensing is thus breaking new ground for the application of on-site technologies in regions where a rethinking of the most suitable UWM system is needed due to infrastructure costs and resource efficiency.\textsuperscript{207} Performance monitoring is not only
imperative to ameliorate the poor reputation of on-site systems, but also to provide a defined level of performance and to reduce overall costs by enabling a more targeted and efficient maintenance and repair regime.

4 Discussion
So far, the focus of this review has been on approaches and examples showing where more data have the potential to substantially improve the efficiency of current UWM practice. This would enable a switch from pure infrastructure generation to the active operation and management of UWM systems and leading to enhancing the utility of existing infrastructures (e.g. RTC of sewer overflows) or the active management of risk and threats (e.g. urban pluvial flood warnings or water supply contamination). This is especially relevant in a changing world where demand generally changes faster than the lifespan of critical infrastructure elements.12,208 However, the most promisingly prospect is that a data-driven UWM could fundamentally novel ways of offering water related services. The ability to provide high levels of water productivity and sanitation with fewer investments in network-based infrastructures could be a key tool in a strongly urbanising world. This is particularly true for areas which are becoming depopulated and areas with low population densities.209 However, careful multi-criteria evaluations need to accompany the implementation of appropriate on-site systems, and the optimal treatment scale needs to be determined with respect to costs, energy consumption, greenhouse gases, drinking water reuse or resources recovery.147,210,211,212,213 Another exiting prospect is that wastewater-based epidemiology can bring added value to operating and monitoring sewer systems and permit intimate insights into societal health and wellbeing.
However, the availability of more data also raises concerns that need to be addressed. The aim of Section 4.1 is to summarise the most pressing concerns of data-driven UWM rather
than giving a complete picture of all conceivable hurdles. In Section 4.2, the role of data and its utility for UWM is discussed in more generic terms.

### 4.1 Challenges of data-driven UWM

Challenges relating to generic data and technology, i.e. successfully taking advantage of data-driven UWM, include turning data into information in a timely manner (data processing), making useful information available to users and utilities (data availability), addressing the issues of data-quality and uncertainty of different data sources (data quality), achieving low operation and maintenance costs for sensors and measurements (data costs) and the lack of general standards and protocols or data management.\(^{37,96,97,214}\) The increased demand on data processing and modelling, e.g. for detailed forecast generation or detailed flood risk assessments is more specific (but also not necessarily exclusive) for UWM. It requires advanced computational power.\(^{90,91}\) Advances in technology and methods are catching up with these needs - improved hardware and new mathematical techniques including the introduction of emulators, have been introduced to bypass inhibitive computational times.\(^{215,216}\)

These challenges are more or less common to all types of ‘smart’ initiatives, such as smart electricity grids.\(^{217}\) In addition to these generic issues, we suggest that three challenges need special consideration, namely i.) data access and ownership, ii.) changes in engineering and management practice, and iii.) the trade-off between benefits and costs. Although these three issues may also apply to other thematic fields, they manifest themselves in specific ways in UWM:

- **i.)** Privacy is a key societal issue surrounding data-driven UWM, namely the ownerships of the data and who has (open) data access. The end of privacy has been declared,\(^{218}\) but in view of the examples outlined above, there is a need to reflect on Orwellian critiques in
UWM. Of course, not all data are sensitive, i.e. rainfall data. However, the availability of more personalised data increases its potential for abuse: thus detailed water consumption patterns could help burglars to identify currently non-inhabited homes, or wastewater-based epidemiology would allow geospatially explicit health information to be revealed that might be exploited by insurance companies. Furthermore, a dependence on ‘smart’ or ‘intelligent’ systems increases the potential vulnerability to cyber-crime. Safe UWM systems therefore need to function reliably and be cyber-secure, e.g. to prevent the unauthorized manipulation of critical infrastructure elements. Introducing common standards is one way of dealing with the widespread scepticism, although this needs to be addressed by a legitimation portfolio approach. Ethical guidelines and legal regulations must also be developed, such as suggested for (illicit) drug monitoring. Whereas this is of lower importance for aggregated data collection (e.g. in the main sewer), data collection at household level is especially sensitive.

ii.) Data-driven UWM requires a change in practices ranging from network operation to decision-making. Professionals working in UWM are generally not used to dealing with an abundance of data and dynamic systems. The adaptation to increased data availability thus requires a change in engineering and management practices which considers the adaptation of new types of models and takes into account uncertainty and risks. However, the general tendency to risk aversion and the restricted time available to utility managers in UWM are not conducive to promoting innovative change. The introduction of new practices is obviously complex and time-consuming and will require changes such as institutional transitions. However, the complex and multidimensional process of changing dominant professional cultures and forming new
industries with different practices such as on-site treatment or RTC needs to be discussed elsewhere. Nevertheless, it is interesting to note that in the electricity sector for instance, IT and communications technology companies, and not the utilities themselves, are a key driver for the transition towards smart grids.

Data acquisition is not free and its costs need to be justified with the potential benefits it provides. For example, the RTC of drainage systems requires reliable actuators, energy provision, sensors, transmission, SCADA (supervisory control and data acquisition) equipment and corresponding operation and maintenance efforts. This can be compared directly with the more traditional approach of investing in static retention tanks and conveyance capacity.

However, the trade-off between the potential benefits and the necessary investments is often difficult to quantify, as the forecasting benefits of increased data are not easily foreseeable and their advantages, such as greater flexibility, are difficult to monetize.

This difficulty in assessing the benefits of change is also discussed in other infrastructures analogous to UWM: thus Moretti et al. e.g. concluded in a review about the environmental and economic benefits of smart electricity grids, that they ‘are energy efficient and reduce greenhouse gas emissions’ but also warn that ‘investments in smart grid systems may not yield any benefits’. In contrast, Niesten and Alkemade reviewed the literature about value creation in smart grids and concluded that corresponding business models could be profitable.

The above listed issues demonstrate that the application of data-driven approaches in UWM is not limited to techno-economic innovation. A whole-system perspective is necessary for
the transformation of UWM systems and technologies need to be adapted to local conditions. Very different barriers need to be addressed as a myriad of factors affect innovation processes and they ‘typically depend on the co-development of new socio-technical configurations, new market structures, new actors and new institutional settings.’

4.2 The utility of data-driven UWM
With respect to data, more are not necessarily better – especially considering the costs involved. It is necessary to decide on which data to collect and to which resolution in view of a specific purpose. The relationship between the amount of data and the utility derived from it may differ for different data-driven applications. Different trajectories can be described (two characteristic examples are shown in Figure 3). The three data regions presented there are characterised as follows: in the data-scarce region, the utility of the data is small despite the cost and effort involved. An example is rain data, where a few high-resolution events do not provide much useful information for the design of the system. After several years of collecting rain events, the data moves into the optimal region, where it offers enough variability to allow design decisions to be made (trajectory A). In the data-abundant region, the additional gain in information and therefore the increase in utility with greater data availability is comparable small. Data-driven UWM should aim to reach the region of optimal data availability.

Another simplified didactic example for this relationship can be taken from the assessment of urban flood risk: better elevation data increases the gain in information, especially if only a limited resolution is so far available (data-scarce region). Once the resolution of the elevation data is good enough to detect streets, a substantial gain in utility can be achieved, resulting in a 'sigmoid' relationship (exemplary trajectory B in Figure 3). After a certain resolution is reached (optimal region), the additional information gain of higher resolution levels out. As
discussed above, the exact shape of the trajectory may vary for different data-driven applications.

While elevation data often offers the desired data availability (see Section 3.2), we suggest that most utilities are still in the data scarce-region as regards most other UWM applications.

**Figure 3.** Schematic representation of two example trajectories demonstrating the relationship between data and utility with respect to different data regions.

### 5 Conclusions

This review has explored the possibility and benefits of creating more data-driven UWM by utilising the ongoing push for smart infrastructures. The main findings are:

- The compiled evidence indicates that i.) considerable research documented in the literature points towards more data-driven UWM, and ii.) the potential benefits are significant.

- Increased data availability from conventional and new data sources is a precondition for many novel methods developed within the UWM literature. Examples include efficiency gains through real-time control, improved risk assessment though sensor-based contamination detection or credible early pluvial flood-warning systems.
• Utilisation of the current network-based UWM can be substantially optimised. Making use of more data improves its efficiency and expands its functionality beyond the passive conveyance of water and wastewater.

• Novel sensoring and smart meter technologies can lead to radically different approaches to UWM. Prominent examples are on-site water and wastewater treatment and the introduction of demand management in water supply.

• Wastewater-based epidemiology can add value to operating and monitoring sewer systems and permit intimate insights into societal health and wellbeing.

• Diverse, consistent and spatially explicit data are essential for integrated modelling and deciding on more sustainable design and management of UWM infrastructures.

• The provision of a high level of water productivity and sanitation with less network-based infrastructure requires more data-driven UWM. This allows defined levels of performance and lower costs for on-site systems to be achieved.

• An evolution towards data-driven UWM faces several obstacles extending from privacy concerns to the requirements of institutional changes in the legal and operative framework. Most importantly, clear evidence for a beneficial cost-benefit ratio that would justify widespread implementation of a more data-driven UWM is generally missing.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +41 58 765 6851. E-mail: sven.eggimann@eawag.ch

Author Contributions
The manuscript was written with contributions from all its authors. All authors have given
approval to the final version of the manuscript. All co-authors contributed equally to this
work.

**Funding Sources**

Omar Wani received funding from the European Union’s Seventh Framework Programme for
research, technological development and demonstration under grant agreement no. 607000.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENT**

We thank Ann-Kathrin McCall and Francesca Cecinati for their valuable comments and
Richard and Margarete Michell for proof reading the text. We acknowledge the valuable
feedback of three anonymous reviewers.

**ABBREVIATIONS**

IUWM, Integrated Urban Water Management; LIUDD, Low Impact Urban Design and
Development; MPC, Model Predictive Control; RTC, Real Time Control; SAC$_\lambda$, Spectral
Absorption Coefficient at wavelength $\lambda$; SUDS, Sustainable Urban Drainage Systems; UWM,
Urban Water Management; WSUD, Water Sensitive Urban Design; WWTP, Wastewater
treatment plant.

**REFERENCES.**


2. Sadoff, C.W.; Hall, J.W.; Grey, D.; Aerts, J.C.J.H.; Ait-Kadi, M.; Brown, C.; Cox,


132. ATSE. Drinking Water Through Recycling: The Benefits and Costs of Supplying Direct to the Distribution System; Australian Academy of Technological Sciences and Engineering; Melbourne, Australia, 2013.


145. Ellis, J.B. Sustainable surface water management and green infrastructure in UK urban catchment planning. *Journal of Environmental Planning and Management* 2013, 56 (1), 24–41.


205. Fontecha, J.E.; Akhavan-Tabatabaei, R.; Duque, D.; Medaglia, A. L.; Torres, M. N.; Rodríguez, J.P. On the preventive management of sediment-related sewer blockages:


