The UWO dataset – long-term observations from a full-scale field laboratory to better understand urban hydrology at small spatio-temporal scales

Frank Blumensaat[1,2,3], Simon Bloem[1], Christian Ebi[1], Andy Disch[1], Christian Förster[1], Max Maurer[1,2], Mayra Rodriguez[1], Jörg Rieckermann[1]*

[1]Department of Urban Water Management, 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
[3]Landesdirektion Sachsen, Stauffenbergallee 2, 01099 Dresden, Germany

Correspondence to: Jörg Rieckermann (joerg.rieckermann@eawag.ch)

Abstract.

Urban drainage systems are integral infrastructural components. However, their monitoring poses considerable challenges owing to the intricate, hazardous nature of the process, necessitating substantial resources and expertise. These inherent uncertainties act as deterrents, discouraging active involvement of researchers and sewer operators in the rigorous monitoring and utilization of data for a comprehensive understanding and efficient management of drainage-related processes. Consequently, a notable absence of openly available urban drainage datasets hampers exploring their potential for engineering applications, scientific analysis, and societal benefits. In this study, we present a distinctive dataset from the Urban Water Observatory (UWO) in Fehraltorf, Switzerland. This dataset is unique in terms of its completeness, consistency, extensive observation period, high spatio-temporal resolution and its availability in the public domain. The dataset comprises coherent information from 124 sensors that observe rainfall-runoff processes, wastewater and in-sewer atmosphere temperatures. Of these 124 sensors, 89 transmit their signals via a specifically set-up wireless network using long-range, low-power transmission technologies. Sensor data have a temporal resolution of 1-5 minutes and covers a period of three years from 2019-2021. To make the data interpretable and re-useable we provide systematically collected meta-data, data on sewer infrastructure, associated geo-information including a validated hydrodynamic rainfall-runoff model. Basic data quality checks were performed, and we motivate future research on the dataset with five selected research opportunities from detecting anomalies in the data to assessing groundwater infiltration and the capability of the low-power data transmission. We conclude that robust automated data quality checks, standardized data exchange formats, and a systematic meta-data collection are needed to boost interpretability and usability of urban drainage data. In the future, ontologies and knowledge graphs should be developed to expand the application of sewer observation data in solving scientific and practical problems.

1 Introduction

Urban drainage systems are essential for public health and sanitation (Ferriman, 2007) because they not only safely transport wastewater and reduce the risk of waterborne diseases, but also protect groundwater and prevent flooding in populated areas. Thus, a profound understanding of the urban hydrological processes is crucial to develop urban areas to more liveable, more sustainable cities. Innovation in this field comes from first principles and data (Eggimann et al., 2017; Blumensaat et al., 2019). However, collecting data in urban hydrology is expensive, because sewers are a hazardous environment that requires specialized training and equipment (Nedergaard Pedersen et al., 2021). Although sharing datasets among research groups has always happened on a personal or project basis (Deletic et al., 2011; Lepot et al., 2016; Caradot et al., 2013; Ochoa-Rodriguez et al., 2015), it is not well established, neither in the research community nor among practitioners.

Understanding urban hydrological processes require high-resolution data on both, the input - such as rainfall - and the output - such as wastewater flows and pollution. However, reliably collecting minute-by-minute rainfall data over many years can be challenging (Bianchi et al., 2013); monitoring stormwater runoff, wastewater flows and pollution processes at a minute scale is resource-consuming. Specialized equipment, training, and software (Mourad and Bertrand-Krajewski, 2002; Dürrenmatt et al., 2013) as well as considerable investments are required to collect and manage the data, as well as to maintain the sensors (Hoppe et al., 2016; Blumensaat et al., 2019). Arguably, sensor maintenance to ensure good data quality is one of the biggest challenges in urban monitoring (Mourad and Bertrand-Krajewski, 2002; Nedergaard Pedersen et al., 2021). Data quality is often dubious due to the low data literacy of the sewer workforce and lack of incentives to use data for evidence-based management of urban drainage systems (Manny et al., 2021). In addition, the lack of standards and meta-data makes it difficult to work with existing or historical data.

To address these issues, there is a need to provide examples of open datasets (Nedergaard Pedersen et al., 2021), which can be used to develop procedures for data quality management and better understand the highly dynamic processes of rainfall-runoff and water quality. Unfortunately, the urban drainage research community has not yet fully embraced Open Science principles and openly available datasets of rainfall, runoff and water quality, as well as topological data of the urban drainage network and a domain-specific description of the catchment, i.e. including land use data and terrain models, are virtually lacking (Nedergaard Pedersen et al., 2021).

In recent times, the demand for open datasets in the field of urban hydrology has been fueled by four primary factors. First, advancements in low-power electronics, data transmission (Ebi et al., 2019) and sensor application (Mathis et al., 2022; Boebel et al., 2023) have significantly reduced the effort of data collection. Second, meta-data efforts and OGC standards have facilitated data sharing among researchers and practitioners (Taylor et al., 2013; Bustamante et al., 2021). Third, scientific data collection efforts, such as the data set “Catchment Attributes and Meteorology for Large-sample Studies (CAMELS)”

(Newman et al., 2015; Addor et al., 2017), have truly revolutionized the field of hydrology through the use of advanced data-driven models (Kratzert et al., 2018). Fourth, the public demands greater transparency of urban infrastructure performance, including reducing environmental pollution (Benyon, 2013; Giakoumis and Voulvoulis, 2023). In this regard, sharing data from urban water management and making it available to the public will become increasingly important soon given the impact of climate change. Therefore, regulatory bodies, such as those in the UK (Environment Act, 2021) and in the EU (EC, 2022), are and will be demanding more monitoring. At the same time industry initiatives, such as STREAM[1], further emphasize the importance of collecting and sharing water company data.

Recently, some openly available datasets have been highlighted, some of which are just available for visual exploration (Spraakman, 2023). For other available datasets[2], no historical data are available. One promising example is the Bellinge dataset from Denmark (Nedergaard Pedersen et al., 2021), which includes rainfall-runoff data from 17 level sensors on the hydraulic behavior of the systems, such as levels, pump power and flows and 3 rain gauges, as well as X-band and C-band radar and air temperature. It is particularly strong on the asset data, which even includes CCTV footage and two different hydrodynamic models. Unfortunately, given the complexity of the system, it only provides data from comparably few level and flow sensors and limited meta-data on sensors: "[…] *exact documentation of sensor maintenance has not been a high priority over all the years, and it is therefore presently not possible to give an overview of when and where sensors have been repaired, been replaced or received some sort of maintenance.*" (Nedergaard Pedersen et al., 2021, p.4786). Thus, important sensor-related meta-data, as well as information on wastewater and stormwater quality is missing.

In this publication, we present curated data from the Urban Water Observatory (UWO) field lab in and around the municipality Fehraltorf, Switzerland. The UWO dataset is unique due to its high spatial and temporal resolution, dense network coverage, and rich meta-data. It consists of four main data packages (A1-A4) (Figure 1), accompanying information (package B) and tools to explore and access the data (C and D). Specifically, the data packages contain 14 sources of precipitation and other meteorological variables (A1), 70 hydraulics sensors, e.g. flow, water level, overflow detection (A2), and 40 temperature measurements of wastewater and the sewer atmosphere (A3). Additionally, the dataset contains information on the behavior of 89 wireless sensor nodes (A4), partly from underground locations, which is unprecedented, to the best of our knowledge. The dataset has a temporal resolution of one to a few minutes and spans a period of three years from 2019-2021.
The dataset is enriched with detailed geoinformation, including geographical and topological data, as well as a hydrodynamic rainfall-runoff model implemented in SWMM (B). Additionally, we provide tools for visual data exploration (C) and sample scripts for accessing monitoring data (D). Notably, the dataset exhibits high data completeness with minimal outages, as shown in Figure 4 (right), and maintains a consistent quality. Thus, the UWO dataset offers extensive research opportunities, ranging

[1] https://streamwaterdata.co.uk/
[2] https://www.kaggle.com/datasets/new-york-state/nys-combined-sewer-overflows-csos; https://theriverstrust.org/key-issues/sewage-in-rivers

from enhancing our knowledge of urban drainage processes to evaluating the effectiveness of process-based and data-driven methods. It also allows for the assessment of wireless sensor network performance in underground applications.

In the following, we describe the catchment of Fehraltorf, the sensors and datasets that are available as well as the methods used to collect, process and explore the data. Finally, we highlight five exemplary research opportunities in the prospect of the UWO dataset (Figure 1).

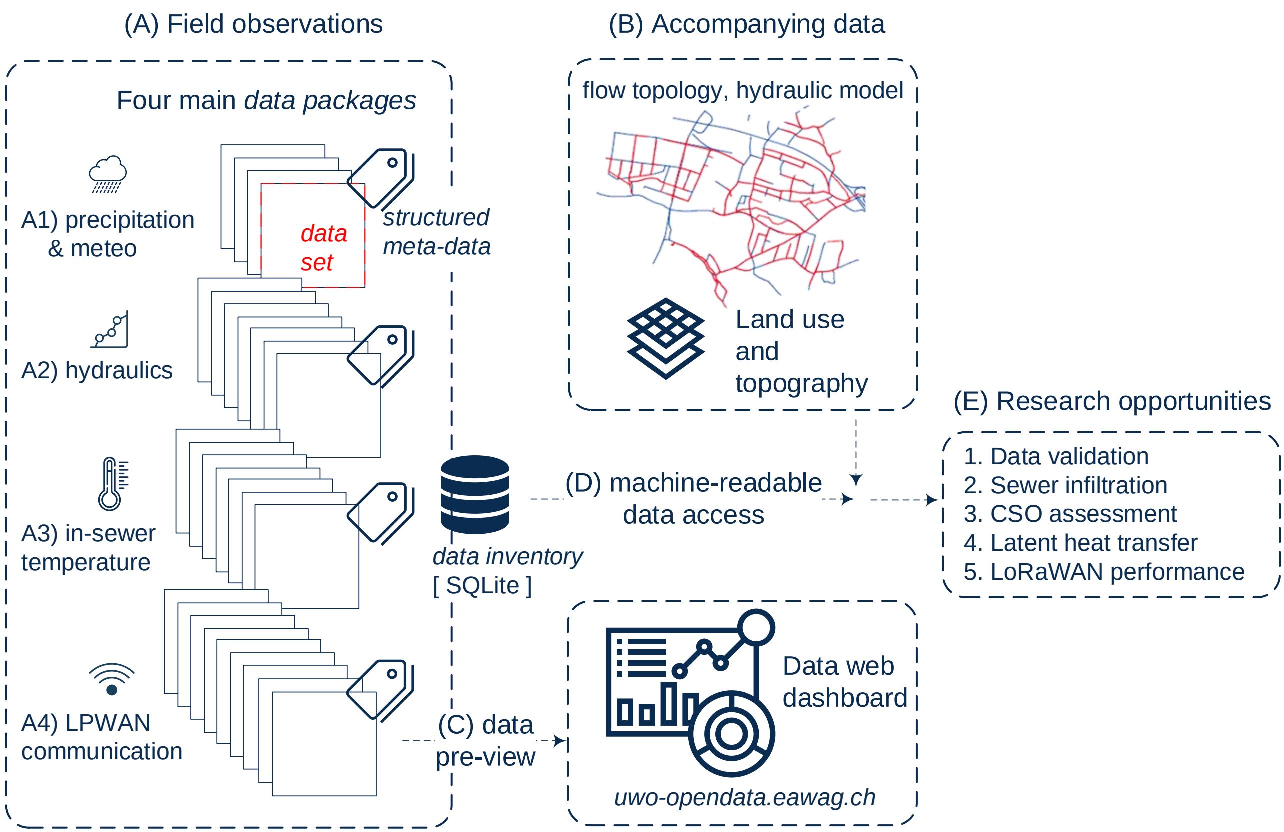

Figure 1: Graphical representation of data provided (A-B), the current data viewer (C), scripts to access the data (D), and examples to highlight future research opportunities in the prospect of the UWO dataset (E).

2. Material

2.1 Description of the catchment

Fehraltorf is a municipality located in the vicinity of Zurich and Winterthur, at an elevation of 530masl with a population of 6,292 inhabitants in 2015 (HBT, 2016) and 6,578 in 2020. In terms of administrative boundaries the municipality covers an

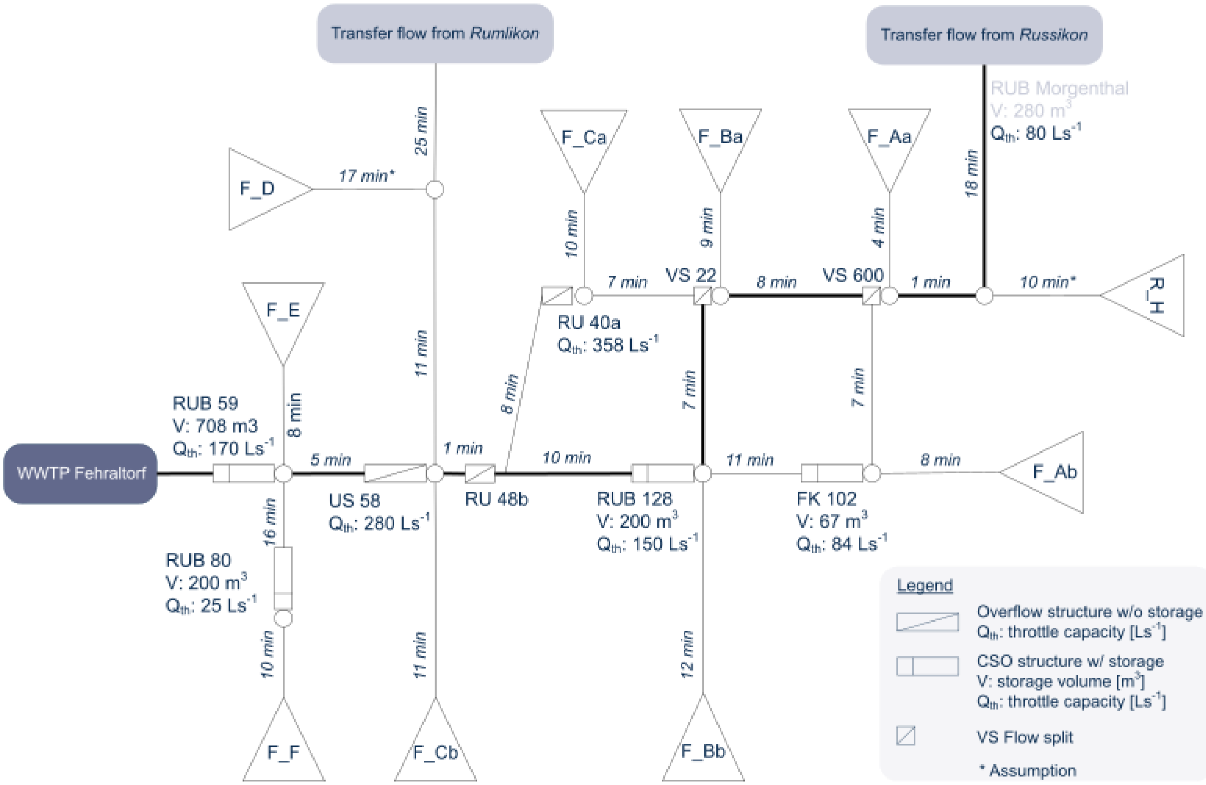

Figure 2, Top: Overview of the Fehraltorf catchment, including urban drainage network, geography and land use. Bottom: The urban drainage flow scheme of Fehraltorf, illustrating the combined sewer network of Fehraltorf. The main flow path through the network is from top right to the left (bold line). Triangles represent aggregated sub-catchments. The relevant characteristics can be found in the supporting information. The here indicated flow times are estimated assuming a flow velocity in the sewer of 1 ms $^{-1}$.

area of 950 hectares with over half of it being used for agriculture (52.7 %), 26.9 % being forested, 13.4 % being developed for settlements, and 5.7 % being designated for transportation purposes (as of 2007) (Figure 2).

The climate in Fehraltorf is a typical humid continental climate,[3] with warm summers and cold winters. The average temperature in the warmest month is above 10 °C and the average temperature in the coldest month is just below -3 °C. The mean annual precipitation (1981-2010: 1334 mm) is distributed throughout the year, with an average rainfall depth in February of 78 mm and the heaviest amount falling in the summer months (ca. 140 mm - see Supporting Information (SI), Section 1 for details). The climate region has a high variability in weather patterns and the frequency of storms and extreme weather events.

[3] According to the Köppen-Geiger climate classification, the climate in Fehraltorf is classified *Dfb* (*cf.* https://skybrary.aero/airports/lszk - last access 23.04.2023)

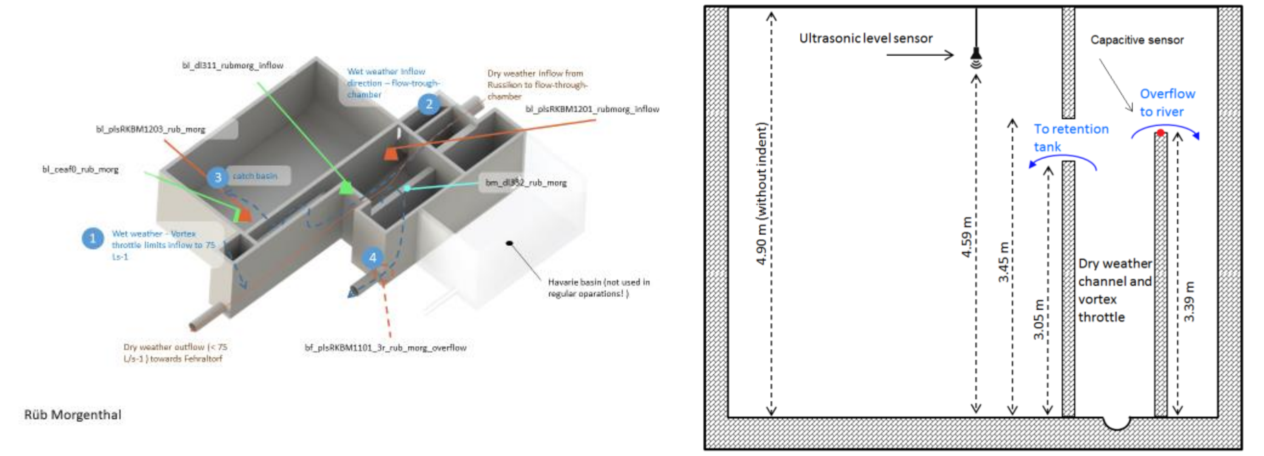

Figure 3 left, CAD model of the combined sewer overflow (CSO) tank "RUB Morgental" including inflow chamber, detention tank, overflow to receiving water and installed sensors. Right: Cross-section through the overflow structure at the sensor bl_plsRKBM1201_rubmorg_inflow (not drawn to scale). Detailed information of all detention basins and flow control structures is provided in the SI.

2.2 The urban wastewater system and the River Luppmen/Kempt

The drainage system of Fehraltorf is mostly a combined sewer system, which consists of 13 km of combined sewers, 4.6 km of foul sewage pipes, and 10.9 km of stormwater pipes. The municipality's runoff-efficient, i.e. reduced area is 40 hectares, as reported in the general drainage plan (HBT, 2016). The system has six overflow structures of which four have a notable retention volume; the area-specific storage volume is about 36 m^3 per hectare runoff-efficient area. Two additional flow split structures (VS22 and VS600) provide hydraulic relief within the system during large storm events but do not spill into the environment (see Figure 2, where the bold line depicts the main flow path).

The nearby villages of Russikon (3320 inhabitants) and Rumlikon (451 inhabitants) are also largely connected to Fehraltorf's wastewater treatment plant (WWTP) (HBT, 2016) (Figure 2). Also included in the dataset is the RUB Morgental (Figure 3), the catchment-concluding combined sewer overflow (CSO) structure in Russikon that limits the transfer flow to 75 Ls-1 towards Fehraltorf. A significant portion of the drainage network infrastructure is located below the average groundwater table. Thus, some parts of the network are affected by sewer infiltration. Depending on the season infiltrating groundwater contributes to the WWTP inflow; the contribution varies from 35 % up to 55 % (HBT, 2016). The overall performance of the sewer system has been evaluated as satisfactory by the consultant engineers Hunziker Betatech (HBT, 2016). More details on the drainage system and the special structures is given in Sections 2 and 3 of the SI.

The Fehraltorf WWTP is a modern facility that utilizes various stages of physical treatment and activated sludge processes to remove solids, organic pollutants, and nitrogen from wastewater. The facility has a maximum hydraulic capacity of 170 Ls^{-1} and contributes a major share to main watercourse, the River Luppmen. As there is considerable industry and commerce in Fehraltorf, e.g. chemical, paint, metal, pharmaceutical and others, the WWTP is substantially influenced by industrial emissions (AWEL, 2021). It is currently being upgraded to eliminate micropollutants with adsorption to powdered activated carbon.

The River Luppmen is the main watercourse, which flows through Fehraltorf, where it changes its name to Kempt at the junction with the Wildbach tributary. The ecomorphology of the Luppmen is heavily influenced by human activity and is partly described as artificial (HBT, 2016). The minimum residual flow (Q_{347}) in the Luppmen amounts to 46 Ls^{-1} upstream and 72 Ls^{-1} downstream of Fehraltorf (HBT, 2016), which means that the proportion of treated wastewater can be greater than the natural baseflow of the river, especially in dry summer periods. The intensive use of groundwater in the last 30-40 years has led to a continuous lowering of the groundwater level and, as a result, to a reduction of groundwater discharge into the Luppmen in the residential area of Fehraltorf. During prolonged periods of dry weather, stretches upstream of the confluent with the Wildbach dry out completely (Krejci et al., 1994).

In 2015 Fehraltorf was chosen as the location for Eawag's Urban Water Observatory (UWO), because the prior knowledge on the urban wastewater system (Krejci et al., 1994; Rossi et al., 2009), its proximity to Eawag, and its similarity in size and system characteristics to many settlements on the Swiss Plateau. Monitoring started in May 2016, reached a peak in 2022 with 124 sensors. Today, the UWO still operates with a reduced number of sensors.

2.4 Sensor data

The dataset provided encompasses information gathered between 2019 and 2021, strategically selected to ensure the most comprehensive and consistently high-quality data. This dataset is rich in sensor signals with a remarkable temporal resolution of typically 1-5 minutes, covering diverse aspects such as precipitation and meteorological data (Figure 4, A1), wastewater hydraulics (A2), temperature of wastewater and sewer atmosphere (A3), and metrics evaluating the performance of the LoRaWAN[4] network (A4). In our definition, a data source refers to a combination of a sensor and logger. This implies that a single source, like a rain gauge, can transmit various signals, including instantaneous precipitation, accumulated volumes, or battery voltage. The dataset not only incorporates signals from our own instruments but also integrates selected sensors from the utility's SCADA System. The data collection process involves automatic retrieval from a range of sources such as FTP servers, databases, and manufacturers' platforms, as detailed in Section 3.1.

In addition to the primary measurement data, the dataset is enriched with crucial meta-data, including logbook entries detailing maintenance activities or operational malfunctions, as well as detailed images of installations (refer to SI Section 4 for further details). Faced with the absence of established standards, we devised a specific naming convention, aligning with the structuring principles outlined in the norm on industrial monitoring EN 81346 (IEC 81346-2, 2019), as elaborated in the SI in Section 5.

[4] LoRa®, ("Long-Range") represents a wireless modulation technique based on a proprietary communication technology. LoRaWAN is the protocol built on top of LoRa modulation delineating the communication and system structure. LoRaWAN has gained recognition as an official standard, ITU-T Y.4480, within the International Telecommunication Union (ITU).

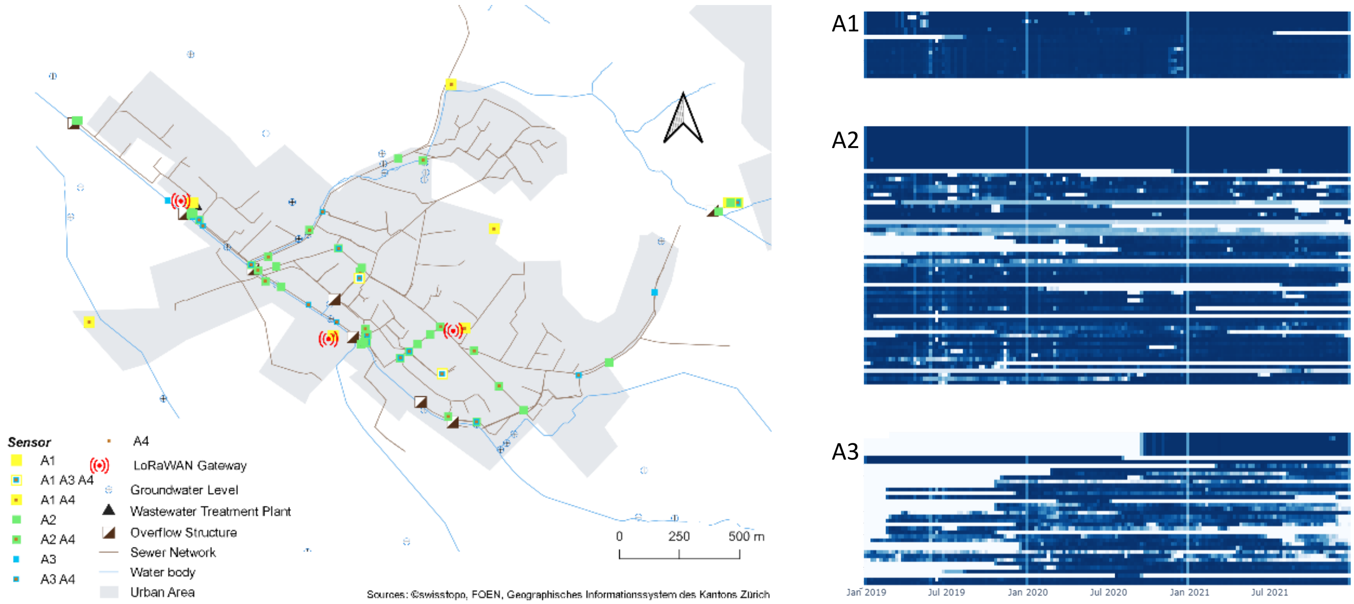

Figure 4, Left: Locations of sensors and sensor nodes. Right: Data completeness from 01.01.2019 to 31.12.2021. A1= precipitation sensors, A2= hydraulics sensors, A3= temperature data. A4 (LPWAN sensor nodes) is not shown. The color saturation indicates the degree of data completeness on a weekly basis. Dark blue indicates periods with 100 % data completeness , i.e. all data points are collected according to the sensor node configuration; white indicates periods with no data. Light blue indicates a reduced number of data points, either through sensor maintenance, sensor outage or incomplete transmission. We provide a dynamic plot in package C, which can be used to interactively explore the data availability and view details. See SI Section 6 for details.

Table 1: Overview of the installed sensor types and some important characteristics.

	A1 - Precipitation and further meterological variables										
Brand	Ott	RM Young	Davis	Lufft							
Model	Pluvio2L	52202	Rain Gauge	WS700							
Number of sources	4	7	1	2							
Temporal resolution [min]	1	1	1	1							
Data transmission	Cellular	LoRaWAN	LoRaWAN	Cellular							
	A2 - Hydraulic measurements										
Brand	Decagon (Meter)	Decagon (Meter)	Decagon (Meter)	Hach	Keller	Maxbotix	Nivus	Nivus	Sommer	STS	PLS (SCADA)
Model	10 HS	5TM	ECTM	Flo-Dar/AV9000	PR36XKY	MB7369/7389/7386	CSM	POA	SQ-3	DLN70	various
Number of sources	1	10	1	3	2	35	2	2	2	1	11
Temporal resolution [min]	5	5	5	5	5	5	5	5	5	5	1
Data transmission	LoRaWAN	LoRaWAN	LoRaWAN	Cellular	LoRaWAN	LoRaWAN	Cellular	Cellular	Cellular	Offline	SCADA system
	A3 - Temperature measurements										
Brand	Maxim Integrated	Sensirion	UIT GmbH	PLS (SCADA)							
Model	DS18B20	SHT 21/35	TSIC	various							
Number of sources	28	4	6	2							
Temporal resolution [min]	1	5	30	1							
Data transmission	LoRaWAN	LoRaWAN	Cellular	SCADA system							

2.4.1 Precipitation and further meteorological variables (A1)

We operated four weighting rain gauges (OTT Pluvio^{2}L, catch area = 400 cm^2), which transmit via Cellular network and eight tipping bucket rain gauges (7x RM Young 52202, catch area= 200 cm^2 and 1x Davis Rain Gauge, catch area= 200 cm^2), which transmit their data via LoRaWAN. To avoid vandalism, we installed them preferentially on utility property, e.g. pumping stations or the enclosure of CSO facilities, or on rooftops. In contrast to the data of the Bellinge dataset (Nedergaard Pedersen et al., 2021), these gauges are not part of the national network operated by MeteoSwiss[5] and their measurements have not been quality controlled a priori. Nevertheless, they operated reliably during the three-year period and measured the data in 1 min intervals. The OTT Pluvios were checked twice per year. Thus, they provide a comparably complete picture of the variability of liquid precipitation. The Davis rain gauge was chosen deliberately as a low-cost rain gauge operated at the site RUB Morgenthal. The meteorological variables have been collected by two multi-parameter weather stations (LUFFT, WS700), which simultaneously measure air temperature, humidity, pressure, precipitation, solar radiation and wind. Verified climatological information from MeteoSwiss is available for the stations *Kloten* and *Fluntern*, which are 15 km (Northwest) and 14 km (West) away from Fehraltorf. However, retrieval of actual precipitation and weather radar data is restricted to research and educational purposes and are not included in this publication.

2.4.2 Hydraulic measurements (A2)

In the study area, a total of 70 sensors have been implemented that survey the hydraulic behaviour of the network during dry and wet weather (Table 1). These sensor nodes measure and transmit data on hydraulic conditions at a temporal resolution of one to five minutes. Among these sensors, thirty-five are equipped with ultrasonic level sensors (MB7369/7389/7386, Maxbotix). Flows are recorded with two correlation wedge sensors (NIVUS), and four non-contact radar flow meters (2x Sommer, 2x Flo-Dar, MarshMcBirney). Additionally, twelve di-electric conductivity sensors (Meter, formerly known as Decagon) were installed as combined sewer overflow detectors (see Section 4.3) which provide redundant information on sewer spills (Blumensaat et al., 2017).

The “backbone” of the rainfall/runoff monitoring consists of four high-quality weighting rain gauges (OTT Pluvio^{2}L) and four industry-grade flow monitors (Nivus POA, Flo-Dar, Sommer SQ-3), which were deployed at strategic locations in Fehraltorf (see Figure 4, left) to provide reliable information on the functioning of the collection system during dry and wet weather. Flow monitors serve as a crucial anchor due to their ability to enable rainfall/runoff analyses and assist with flow balancing. They also survey relevant upstream boundary conditions, i.e. wastewater inflows from upstream sub-catchments: a) WW inflow from the Rumlikon district in the Northern part (F02), and b) WW inflow from the municipality Russikon is connected at the Northeast (F03). At two further locations (F08, F12, [F07, F10]) – redundant at the same site – we observe flow dynamics within the drainage network.

[5] MeteoSwiss - Swiss Federal Office for Meteorology and Climatology

Level sensors and threshold detectors were mainly used to complement the flow meters, e.g. in smaller sewers and to better describe the filling, spilling and emptying behaviour of CSO tanks. For example, all overflow structures had been instrumented with at least one level sensor (MaxBotix and PLS) inside the CSO tank and one capacitive sensor (Meter) installed on top of the weir crest to act as a binary spill detector (SI Section 4.4).

Calibration of the hydraulic sensors was carried out at least once a year, but rather on demand than based on a structured maintenance schedule. Mechanical cleaning of the sensors, as well as reference measurements and visual checks (e.g., for perpendicular alignment of the non-contact sensors) were carried out regularly, besides the routine maintenance work (e.g., changing batteries) (see further details in the SI Section 4). The utility operates a process monitoring at certain structures, namely water level in retention basins as well as two flow measurements (Venturi and inflow WWTP). In the basins themselves, tank levels were monitored redundantly, mainly using low-power ultrasonic sensors (MB 7369/7389/7386, MaxBotix).

Some of the sensors were specifically tailored to sensor- units (nodes) to enable Long-Range-Wide-Area-Networks (LoRaWAN). We used an operating voltage of 5 V or less and a current consumption of a few milliamperes for sensors and radio modules, which simplified the implementation of intrinsically safe devices to protect against ignition protection in the (per-se) explosive sewer environment.

2.4.3 Temperature measurements (A3)

Temperature can be used as a natural tracer to provide information on hydraulics (Dürrenmatt et al., 2013), groundwater infiltration (Schilperoort et al., 2013; Panasiuk et al., 2022) and sediments (Regueiro-Picallo et al., 2023). Also, net-zero considerations require a detailed understanding of urban energy and heat fluxes e.g., for energy recovery with heat exchangers (Hadengue et al., 2021), or to validate wastewater heat exchange predictions (Figueroa et al., 2021). Therefore, the in-sewer temperature was monitored using dual sensors that simultaneously record the temperature of the wastewater stream and the corresponding sewer headspace. Above ground, the ambient air temperature has been monitored at four locations across the catchment. To characterize surface runoff during wet weather, three additional temperature sensors were installed in gully inlets. The positions of the individual monitors are shown in Figure 4, left. This experimental design generates a consistent long-term data set, enabling the analysis of temperature dynamics within a hydrological context. It allows for studying temperature variations across different compartments at network scale, considering various seasonal and loading conditions.

2.4.4. Telemetry performance data of the sensor network (A4)

Nowadays, the Internet of Things (IoT) is flourishing, and its prospects are promising, although early implementations were hindered by high costs, a lack of standards and technological challenges (van Kranenburg and Bassi, 2012). However, with the emergence of Long-Range Low-Power technologies like LoRaWAN, Sigfox or NBIoT[6], implementing and operating wireless sensor networks, i.e. of such Low-Power Wide Area Networks (LPWANs) has become straight forward. The increasing availability and sophistication of these technologies has enabled data collection, also in the field of environmental engineering, at an – so far – unseen density and scale.

Data transmission based on the LoRa® technology (Semtech Corporation, 2015) was a key factor in the UWO monitoring initiative. The UWO sensor network comprises i) 89 LoRa-enabled sensor nodes, ii) 3 LoRaWAN base stations, or gateways, and iii) network management elements (Blumensaat et al., 2017; Ebi et al., 2019). Data are transmitted between local sensor nodes and base stations using low-power, sub-gigahertz wireless communications, primarily utilizing the LoRaWAN protocol. Here we provide telemetry performance data including Received Signal Strength Indicator (RSSI), Signal to Noise Ratio (SNR) and the Spreading Factor (SF) for further analyses. According to our experience, LoRaWAN enables two-way wireless communication between independent sensor nodes and base stations with a moderate range of up to 20 km (above ground, subject to line of sight and weather conditions). Underground, we find typical transmission ranges of 500 m, which can be substantially extended with our LoRa-based mesh technology (Ebi et al., 2019).

As sensor nodes, we deployed 89 industrial-grade nodes (DL-MBX, Decentlab) as well as custom prototypes (based on Libelium Waspmote). The DL-MBX nodes are powered by two standard LR20 alkaline-manganese monozinc cells (1.5 V, 18,000 mAh) and achieve a battery life of three to six years when transmitting every five minutes. Our custom prototypes are powered by standard lithium-polymer batteries (3.7 V, 6,700 mAh) and reach a battery life of about six months (Blumensaat et al., 2017). To connect the radio modules and sensors, we used digital data communication. This prevented interferences and facilitated seamless integration with the nodes' microcontrollers. We deliberately chose to connect the sensors to the radio modules via sensor cables, sometimes of several meters, which made it possible to optimize positioning of i) the sensors regarding the flow and ii) the radio module with regard to connectivity to the gateway.

To transmit the data from the sensor nodes, we installed three standard gateways (Kerlink Wirnet Station 868). The gateways receive signals from the sensor nodes and transmit the data via cellular to a network server. As network manager (middleware), we use commercial services provided by LORIOT (Switzerland). This software manages the communication channels and

[6] Sigfox, LoRa, and NB-IoT are all types of low-power, wide-area network (LPWAN) technologies that are used on the Internet of Things (IoT) ecosystem. While LoRa and Sigfox are classified as non-cellular IoT technologies, NB-IoT is categorized as a cellular IoT technology.

data rates of the sensor nodes, sorting and forwarding the data packets to our internal data server. Although these services are subject to charges, they are far less costly than those for data transmission via SIM-based internet connectivity.

In evaluating the appropriateness of radio coverage, we initially conducted a series of signal strength tests. As a result, our investigations demonstrated that deploying three outdoor gateways on elevated structures presented a cost-effective and efficient solution, achieving a balance to ensure sufficient signal strength throughout the entire urban catchment area. This region covers approximately three kilometers by three kilometers in extension. Nevertheless, it is crucial to acknowledge that signal strength is not a static parameter, and the network's performance shows temporal and spatial variations, as elaborated in Section 4.5.

The LoRaWAN network data is automatically transmitted through a data pipeline to our internal data server. Subsequently, it is consolidated with signals from other sources and stored in our "Datapool" data management system, as outlined in Section 3.1.

3 Methods

To collect and curate the data from a large sensor network it is important to establish a data pipeline that provides automated data transfer, automatically stores the sensor data, performs automated and semi-automated quality checks and provides the data in the correct format and aggregation for the required services (Figure 5).

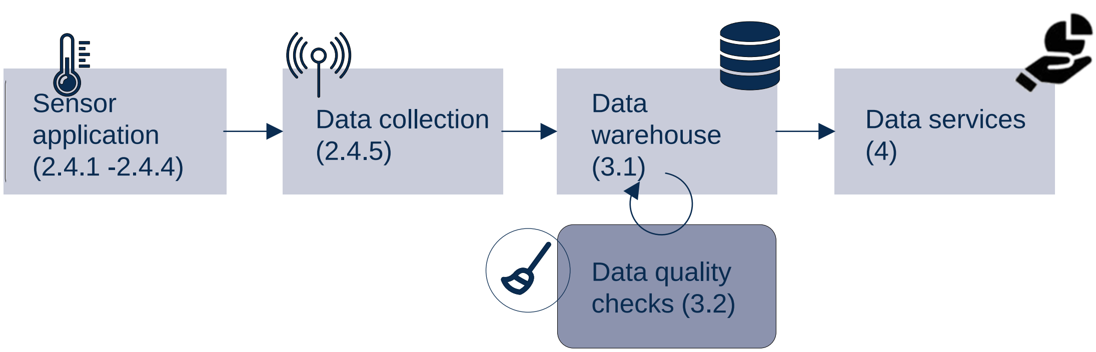

Figure 5: The individual components of the data pipeline for collecting, storing, and using UWO data from the sensor to the data service. The numbers refer to the corresponding Sections.

3.1 Implementation of the data pipeline: Data collection, processing and storage

The "Datapool" is an advanced data warehouse developed at Eawag with implementation support by ETH Zurich's Scientific IT services. It utilizes open-source software, including PostgreSQL, PostGIS, TimeScaleDB, and a Python wrapper to manage structured data, such as time series, meta-data, signal quality information, binary data, raster data, and laboratory data. It allows for fast query processing, even with large amounts of data, and offers a clear separation of data management roles (see SI Section 5). The "Datapool" is based on open-source techniques, ensuring seamless interoperability with other software. To enhance simplicity, the UWO sensor data, totalling around 120 GB (comprising packages A1 to A4), has been divided into three segments, each corresponding to a specific year, and exported in SQLite format.

3.2 Ensuring consistent data quality

Long-term field monitoring constantly requires a high level of attention to ensure that the data are consistent, accurate, and reliable. By continually validating the data, we can identify any type of issues that affect the data collection routine, adjust and maintain system performance, and thus ensure high data quality.

3.2.1 Automated flagging with range and gradient checks

Automatized data plausibility checks of raw observations have been accomplished in quasi real-time. Although data-driven modelling holds great promise to improve the data quality in the future, initial trials required supervised data to yield satisfactory results (Russo et al., 2019; Disch et al., in prep.; Disch and Blumensaat, 2019). Presently, two plausibility check routines, namely a range and a gradient check are executed regularly (currently every 24 hours) and return flags to the "Datapool". The range check is an effective measure to consider prior knowledge of the geometry of the sewer and the physical measurement principle in quality checking (Bertrand-Krajewski et al., 2008; Clemens et al., 2021). A range check is necessary to ensure that the measured values fall within a valid range (min, max). Values outside of this range may indicate an issue with the measurement or the sewer system itself and are flagged. Gradient checks are useful to detect sudden changes in a time series, which can help to identify events of malfunctioning or abnormal system behaviour (Clemens et al., 2021; Russo et al., 2019). The latter involves looking for abrupt changes in the order of one magnitude from one data point to the next. By detecting these artefacts, it may be possible to identify the beginning of an event, such as a sudden spike or drop in a measurement. However, it is important to validate the results of gradient checks to ensure that the detected changes are meaningful and not just random fluctuations in the data. Despite the fact that many sensors are similar, parameterization of both routine checks is specific for each sensor.

The results are intentionally stored as additional information "flags" (to the raw data). This flagging helps to avoid redundancies, improves reproducibility, and saves computing resources. The data user can retrieve data with or without flags from the integrated SQLite database using one of the provided data-access packages (DAP) provided in package D. DAPs are available for different programming languages, e.g. Julia, Python and Matlab/Octave. The result of the automated flagging is either *True* or *False* for each data point. As data are recorded, these results are visualised in an internal maintenance dashboard and trigger maintenance alerts as appropriate. Only data flagged as *True* is shown in the web viewer (C) (see Section 3.4).

3.2.2 Regular consistency and homogeneity checks

To enhance data quality, we performed regular checks to ensure data consistency and homogeneity. We also identified and rectified any significant errors resulting from device failures, monitoring configuration changes, clogging, cross-section alterations, and other potential factors, thus ensuring unambiguous data.

In addition to regular data consistency and homogeneity checks, advanced data validation methods, such as anomaly detection, have been proposed in literature (Russo et al., 2019, 2021; Clemens et al., 2021). However, these are a highly challenging topic of current research (Deheer, 2022; Disch et al., in prep.; Disch, 2022) and no standard methods are available yet for sewer data. We highlight this as a great opportunity to do further research with this dataset in Section 4.1.

3.3 Rainfall-runoff model

In addition to sensor data, we provide a hydraulic model implemented in EPA-SWMM, which entails information on underlying infrastructure, flow topology, network engineering, operation and functioning during dry and wet weather. It has been adjusted to the monitoring data using optimization and likelihood-based approaches and describes the rainfall-runoff process largely satisfactorily, with a Nash-Sutcliffe Efficiency (NSE) of 0.7 in calibration and a NSE of 0.4 for predictions of the WWTP inflow. Initially, the model was implemented in MikeUrban (DHI, 2015) by a consultancy for the drainage master plan. It was then converted to EPA-SWMM implementation to allow open access. This further referred to as base model only covers the combined sewer system, i.e. it ignores storm sewers and the associated drainage area, which represents about one-third of the total drained area. It consists of 246 sub-catchments that drain into 427 junction nodes connected by 431 links, with six CSO structures. Hydrological and hydraulic parameters were adjusted after the conversion to EPA-SWMM, while the structure and hydraulic characteristics of the pipe system remained unaltered. Flow balancing has been conducted to validate the model for dry weather conditions. The base model accommodates groundwater infiltration, and it incorporates two methods (base-flow infiltration; rain-dependent inflow). Wani et al. (2022) performed a spatially differentiated calibration using the Bayesian parameter inference approach and the NSE as the different points as objective functions. They concluded that the calibration of the model using spatially distributed data did not lead to better parameter estimates in Fehraltorf. The SWMM model structure files as well as input files, flow patterns are available in package B. As with any other drainage system, the sewer network in Fehraltorf is subject to changes. For instance, with the installation of new flow-limiting hardware in 2019-2020, modifications were made to the network that are not yet considered in the currently provided model structure. Further information on the rainfall-runoff model, as well as on the accompanying data can be found in the SI.

3.4 Accessing and exploring the monitoring data

To enable open access to data, we have adopted the FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Wilkinson et al., 2016). To make the data findable, we ensure that our data is described and labelled using appropriate keywords, standardized formats, and persistent identifiers (see Section above). In this way, researchers and users can easily find and access the data they need. To make them accessible, we make our data available through ERIC/open (Eawag Open Research Data Institutional Repository) (see Section 6). We also provide documentation on how to access and use the data. To make the data interoperable, we structure our data using open, standardized formats that can be easily integrated into existing workflows and systems (SQLite, CSV, ascii). We also provide meta-data on how the data was collected, processed, and analysed to facilitate reproducibility and reuse in table "*source_meta_data*" (SI Section 5, Figure S 5-2). To make the data reusable, i.e. to ensure that data can be used for multiple purposes beyond the original research project, we place them in the public domain with a CC0 license, which waives all copyright and related rights to the extent allowed by law. This means that others are free to use, modify, and distribute it without any restrictions (Section 6).

Unfortunately, the ERIC/open implementation provides access to the data, but yet only very limited exploration capabilities. Therefore, similar to previous work (Špačková et al., 2021), we provide a web-based dashboard[7] as a pre-viewer to the data, which has been implemented in R-Shiny and is an efficient way to explore and filter the data, e.g. by signal type, source or location (Figure 6).

[7] https://uwo-opendata.eawag.ch/, last accessed 15.01.2024

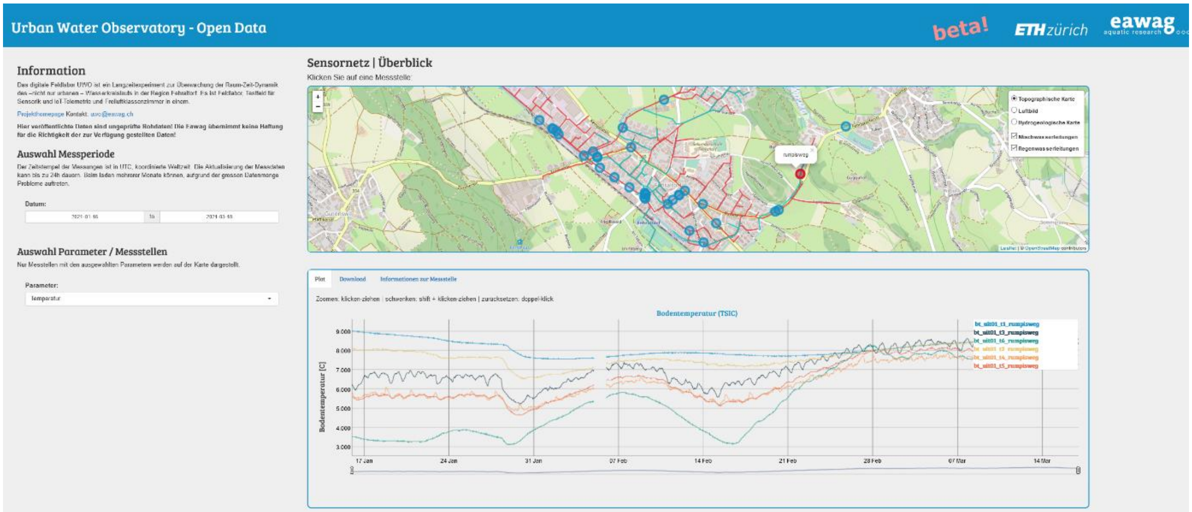

Figure 6: Online pre-viewer to the UWO dataset, which includes information on the location of the data source, as well as time series and meta-data on the characteristics of the sensor and the monitoring site. Web access: https://uwo-opendata.eawag.ch/

4. Research opportunities

The UWO dataset is valuable for researching urban drainage processes and evaluating process-based and data-driven methods with real-world data. Here, we will briefly highlight five research opportunities using the UWO dataset to: i) develop automated data quality checks, ii) assess spatio-temporal variations in groundwater infiltration, iii) improve event-duration monitoring of combined sewer overflows, iv) investigate sewer heat transfer and v) assess wireless sensor network performance. More details on the individual opportunities are given in SI Section 9.

4.1 Anomaly detection of sewer monitoring data using semi-automated machine learning approaches

Advancements in sensor technology and data transmission have transformed process monitoring in drainage systems through a wealth of data (Ruggaber et al., 2007; Kerkez et al., 2016). However, traditional manual data preparation is insufficient for extracting useful information. Automated approaches are needed for real-time sensor data validation, if needed with human intervention to improve data quality. Machine learning, especially unsupervised methods, can enhance real-time data preprocessing and ensure comprehensive data quality assessment. Detecting anomalies in recorded data is crucial, but complex data-driven methods often fail with urban drainage data (Deheer, 2022). In this study, we compared three data validation methods (ARIMA, Autoencoder, One-class Support Vector Machine (OCSVM) using UWO dataset time series of levels, flow, and temperature. We employed filtering, smoothing, and imputation preprocessing steps and generated a reference time series with synthetic errors. We used the popular F1 score to evaluate the performance to correctly detect anomalies (see SI Section 9.1). Our results suggest that the simultaneous analysis of related signals enables anomaly detection, and more preprocessing with human intervention improves the F1 score (Figure 7).

The F1 scores for monitoring data with synthetic errors during wet weather (Figure 7, right: D_{wet}) are substantially better (ARIMA: 0.71, OCSVM: 0.65, Autoencoder: 0.73) than for real measured data (average: 0.43). As expected, preprocessing (denoising, smoothing, imputing) aids anomaly detection and ARIMA fails with incomplete series. We also find that partitioning the data by weather enhances anomaly detection. ARIMA and Autoencoder methods show promise with synthetic data, but their superiority over OCSVM diminishes with real measurement data.

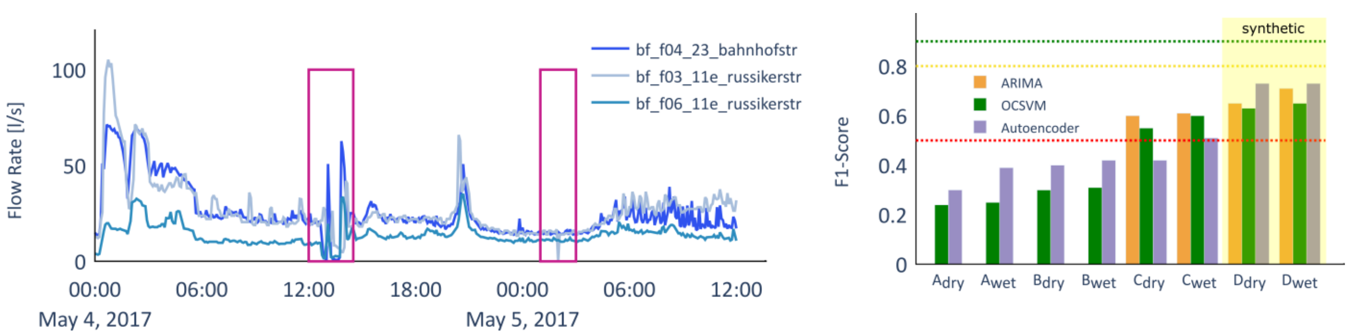

Figure 7, left: Observations with potential anomalies. The highlighted data show the benefit of multiple sensors in detecting anomalies. While the high variability on the left is reflected in all sensors, the one on the right only appears in a single sensor. Right: F1 scores for different methods for pre-processing (A-D). An F1 score of 0.5-0.8 (between the red and yellow dashed lines) is considered medium quality, 0.8-0.9 good and above 0.9 excellent. For real-world data (A-C), the performance increases with increasing levels of pre-processing. For real-world data, ARIMA performs best and the Autoencoder never reaches the performance on synthetic data (D).

4.2 Quantification of groundwater infiltration

Infiltration in urban drainage systems has detrimental effects on wastewater treatment efficiency and increases costs, because it dilutes sewage, reducing biological treatment efficiency, and overloads the drainage system during rain (Staufer et al., 2012). To estimate groundwater infiltration (GWI) rates in the Fehraltorf sewer network, we analyzed long-term flow recordings, focusing on dry weather night-minimum flows and excluding rain-induced infiltration. Using the night-minimum flow, GWI rates were estimated at the catchment outlet and two transfer flows in the Fehraltorf sewer network.

GWI rates ranged from 10 to 15 Ls^{-1}, depending on the season (see SI Section 9). The detailed monitoring data revealed spatial variations in GWI with changing seasons. For example, in April 2018, a high groundwater table affected 256 out of 459 manhole inverts, while in October 2018, only 100 manholes were impacted (Figure 8). Future work can incorporate the spatially differentiated GWI rates into the hydraulic sewer network model to improve the predictive capabilities.

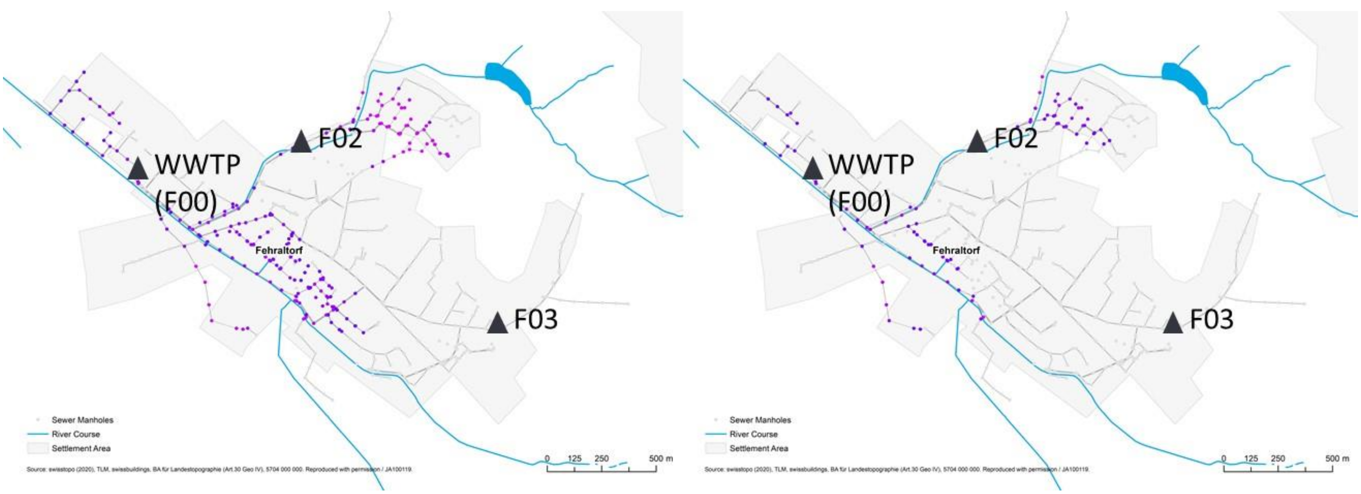

Figure 8 left: 256 sewer manholes in the Fehraltorf network were affected by groundwater in April 2018. The darker the colour the more are manholes submerged in GW. Right: 100 sewer manholes in the Fehraltorf network were affected by groundwater in October 2018. The darker the colour the more are manholes submerged in GW.

4.3 The value of redundant sensors in event-duration monitoring

Assessing CSO activity through tank-level monitoring is crucial for quantifying pollution and optimizing sewer networks. To date, more and more countries have implemented data-based compliance assessments of CSOs and make spill data, e.g. from overflow event duration monitoring, available to the public (Rieckermann et al., 2021; EC, 2022). However, ensuring data quality is a challenge and research explores various monitoring techniques to make data-based compliance assessment more reliable.

In the UWO, we equipped some CSO tanks with multiple ultrasonic and capacitive sensors to independently monitor overflow duration to investigate how redundant signals would reduce the uncertainty of CSO event-duration monitoring (SI Section 9).

The results shown in Figure 9 (right) demonstrate that comparing the capacitive sensor signal with the level signal greatly improves data confidence and allows for post-calibration. With only the tank level sensor, incorrect level sensor processing severely underestimated CSO activity and resulted in 50 % less cumulative overflow duration in a period of 1’077 days.

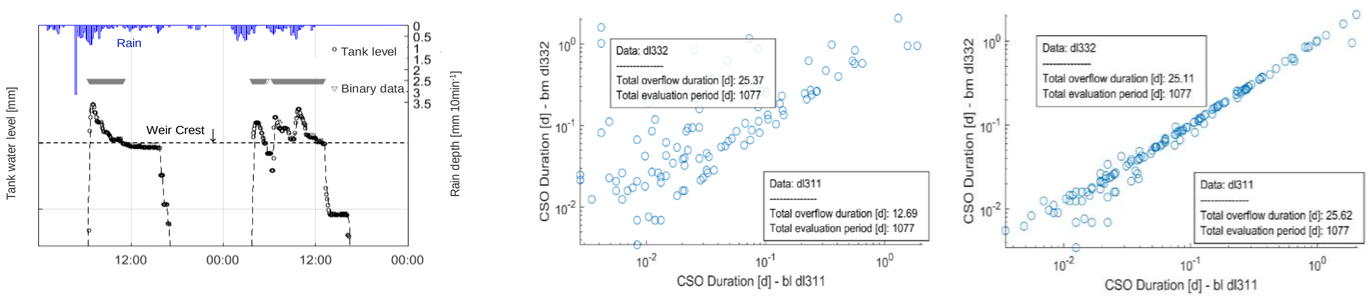

Figure 9, left: Continuous tank-level (dots) and binary (triangles) information, the latter derived from the capacitive sensor signal, reflecting the overflow activity for a period of two days for which two independent overflow events were recorded. Monitoring frequency are five minutes, for rainfall it is one minute. Middle and right chart: event-specific overflow durations derived from one capacitive sensor (bm_dl332_rub_morg) and one ultrasonic level sensor (bl_dl311_rubmorg_inflow). Middle chart: tank-level data (dl311) have been used with original (erroneous) information on sensor settings, i.e. the distance between sensor head and the tank invert. Right chart: information on sensor offset settings have been revised.

4.4. Assessing latent heat transfer in sewers – detecting condensation

Accurate predictions of sewer heat transfer processes are important for in-sewer processes (Huisman, 2001), and heat recovery (Figueroa et al., 2021; Abdel-Aal et al., 2019; Hadengue et al., 2021). Latent heat transfer in sewer systems has been often disregarded by urban drainage modellers (Elías-Maxil et al., 2017), although there is no quantitative evidence that latent heat transfer is not important. Using a subset of the provided temperature time series, we investigated the difference between headspace and bulk liquid temperatures to identify conditions for A) condensation and B) evaporation (SI of Figueroa et al., (2021)). We analyse the provided temperature data in three locations dl933, dl935 and dl931 (Figure 10). We find that condensation takes place in April when the typical temperature difference between the sewer headspace and the bulk liquid temperature is -2 K. In contrast, with warmer ambient air and headspace temperatures, evaporation occurs. Maximum differences can amount to 5 K, at location dl931. Based on these results, Figueroa et al. (2021) conclude that latent heat transfer should not be neglected, especially in areas with high relative velocity and lower bulk liquid depth in the sewers, i.e. in steep catchments and peripheral regions. Even in scenarios with high relative humidity values, latent heat processes play a crucial role and should be considered. This makes the provided temperature data an ideal source for further developing heat exchange models for sewer networks.

Figure 10: The difference between the sewer headspace temperature and the bulk liquid temperature observed at three specific locations in the main collector. Red dots indicate periods of evaporation (higher temperature in the headspace compared to the bulk liquid) and blue colours indicate periods of condensation (inverse temperatures). From the SI of Figueroa et al. (2022).

4.5 The performance of a LoRaWAN network for underground applications

Monitoring sewer systems faces challenges due to the restricted reception of diverse wireless data transmission technologies in underground settings. While LoRaWAN holds potential for subterranean monitoring, uncertainties persist regarding its performance in challenging conditions, particularly within metal-covered manholes, as highlighted by Ebi at el. (2019). We evaluated the quality of LoRaWAN network service by comparing packet error rates (PER) of underground and above-ground sewer nodes. PER represents the percentage of data packets that fail to reach their destination due to transmission errors Table 2. Generally, we find that data packets can be transmitted from sensor nodes located in sewers if the nominal distance between node and gateway is less than 500 meters. Analyses of the data on the LoRaWAN telemetry performance show a low average PER of about 5 % and a median of 3 % across practically all nodes. Interestingly, 1-minute nodes perform slightly better, likely due to the fact that 1-minute transmission frequency mostly applies to sensor nodes located above ground (SI Section 10).

Future research can analyse the IoT radio network to explore LoRaWAN networks' potential for underground monitoring, advancing the maintenance of infrastructures such as heating, water, and electricity networks. Investigating environmental factors impact, such as heavy rainfall or extreme temperatures, on key radio performance indicators could provide further insights into LoRaWAN networks suitability for underground monitoring.

Table 2: Summary statistics of the Quality of Service of the LoRaWAN wireless network. The mean of sensor median packet error rates (PER) was computed from weekly values for two groups of gateways (1min, 5min). While the global average of data packet losses is very low, i.e. approximately 5 %, median values indicate an even lower packet loss, i.e. 3 %. Also, the 1-minute nodes perform slightly better, most likely since a fair number of these nodes do not transmit from underground locations. Values in brackets represent the number of nodes included in the evaluation in the corresponding evaluation period.

Median PER [-]	# of packets per week	2017 (# Sensors)	2018*	2019	2020	2021	2017 - 2021
1-min nodes	10,080	0.026 (2)	0.020 (13)	0.058 (34)	0.048 (36)	0.037 (31)	0.050 (39)
5-min nodes	2,016	0.056 (34)	0.043 (46)	0.055 (58)	0.061 (58)	0.032 (47)	0.056 (68)

6 Data availability

Data from the Urban Water Observatory are available at the Eawag Research Data Institutional Repository (ERIC/open). We provide the data under a CC0 license, granting users the freedom to use, modify, and distribute it without any restrictions, as the work is dedicated to the public domain. The geodata by swisstopo is supplied with conditions of use, which comply with the legal basis. The conditions of use enable free use for all purposes and oblige the user to indicate the source as "Source: Federal Office of Topography swisstopo" or "© swisstopo". The datasets are quality controlled by Eawag staff and provided as part of the UWO project[8]. The packages are organized as follows:

A) UWO - Field observations (2019 to 2021): Sensor data including corresponding meta-data (cf. A1-A4 in Fig. 1). The data are organized in three SQLite databases representing annual slices of UWO field data collected in 2019 until 2021 https://doi.org/10.25678/00091Y (Blumensaat et al., 2024d).

B) UWO - Accompanying information (cf. B in Fig. 1): a) a collection of prepared geodata, including a topographic basemap in raster format (National map 1:25'000) provided by swisstopo, a Digital Elevation Model (DEM, swissALTI3D) in 5m raster, an aerial photo "Swissimage", Cantonal information on groundwater resources, sewer network cadastre data provided by the municipality of Fehraltorf, and b) a hydraulic "base" model version (EPA SWMM v5.1) https://doi.org/10.25678/000991 (Blumensaat et al., 2024a).

C) UWO - Data viewer: a web-based dashboard that allows data users to view all UWO field observations, among them those provided in the sensor data package A https://doi.org/10.25678/00092Z (Blumensaat et al., 2024c).

D) UWO - Data access: script files to query the UWO sensor data provided in the abovementioned SQLite databases. Sample queries are provided to access observation data in Python, Octave/Matlab, and Julia respectively. An overview of which sensor (source) is associated which package (cf. A1-A4 in Fig. 1) is given in table format https://doi.org/10.25678/000980 (Blumensaat et al., 2024b).

7 Conclusions and Outlook

Monitoring of urban hydrological processes and making the data publicly available is of paramount importance, as sewer networks are under-monitored in relation to their significant monetary value and service. Therefore, we present this unique open set of urban hydrological data, which not only includes long-term observations on rainfall-runoff processes, but also information on in-sewer wastewater temperatures and the telemetry performance of our LoRa-based sensor network plus all relevant meta-data. Based on the results and experience of collecting and curating the UWO dataset, we draw the following conclusions:

[8] https://doi.org/10.25678/000C5K

- IoT-prone transmission technologies and advances in sensor development offer the opportunity to revolutionize sewer process monitoring by providing spatially differentiated information on key system parameters in real time. But, they do not make traditional monitoring techniques obsolete. The promise of ubiquitous monitoring is not solved by low-power data transmission. Even if battery life can be extended to 5 or 10 years, possibly through adapting monitoring intervals to expected dry and wet weather dynamcis, managing a swarm of hundreds or thousands of sensors will require substantial resources and manpower until energy self-sufficient systems for monitoring and data transmission are available.
- In terms of sensors, we have learned that it pays to establish a monitoring backbone consisting of a few full-featured but (often) maintenance-intensive sensors, such as rainfall and pipe flows. Due to their high measurement accuracy, their data can be used as reference observations, i.e. to establish volume balances. In addition to that, low-tech sensors may reveal relative deviations from normal system states, allowing operators to detect operational problems such as blockages or leaks earlier. To correctly interpret spatially resolved runoff patterns, adequate spatial rainfall measurements are necessary and, generally speaking, more level and flow sensors also require more rain gauges. Rainfall and urban climate monitoring provides further insight the impact of weather on the sewer system and helps operators to anticipate critical phenomena such as overflows or flooding. In comparison to spatially distributed rainfall, the monitoring of wastewater thermal energy, e.g. for heat recovery, is comparatively cheap and can help authorities to achieve an integrated operation of the wastewater system.
- To fully benefit from monitoring data, it is crucial to ensure that streaming the sensor data in real-time is sufficiently reliable. We found that *automated* state-of-the-art quality checks, such as range and gradient checks, are adequate to identify anomalies in the data. Advanced analysis is required to further differentiate whether anomalies are due to sensor malfunction, triggering on-demand sensor maintenance, or abnormal system behaviour, requiring operational troubleshooting. Regular *manual* consistency checks, e.g. with simple physical models as Manning-Strickler, and homogeneity checks are still essential to ensure the accuracy and reliability of sensor data.
- Collecting field observations without validating the data in a timely manner does not make sense. The significance of automated data validation becomes paramount, particularly when handling a considerable number of sensor signals, ranging, for instance, from 30 to 40 devices or more. Nevertheless, generally applicable solutions have not yet been developed, probably because sewer flows, both in terms of quality and quantity, display substantial variability and apparent randomness, which makes it challenging to develop universal solutions. We believe that further research is needed to ensure data quality, possibly using advanced methods and novel strategies. In our experience, the use of long time series (months to years) and the inclusion of data from nearby, i.e. adjacent sensors or alternate (surrogate) signals collected at the same site have been particularly promising in quality control. The UWO dataset and its meta data can serve as a nucleus for testing modern data-driven approaches such as generative adversarial networks or recurrent neural network models, as discussed in Russo *et al.* (2021). However, creating universally applicable

methods demands a more extensive collection of annotated monitoring datasets, akin to the scale seen in image and text databases used for machine learning. To facilitate seamless reuse, it is essential that these datasets be released into the public domain without imposing any license restrictions.

- For in-sewer processes, such as sewer infiltration or overflow behaviour, the data provides a very comprehensive picture of the sewer network performance. In addition, it can be useful to detect causes of bias between the model and environmental factors, such as weather or temperature, the built environment, e.g. land use, or sensor failures. Data are useful to test hybrid models and data assimilation approaches.
- Meta-data is crucial for accurately interpreting sensor signals. Standardising meta-data descriptions is essential for uniform interpretation. Coordinated efforts are needed to introduce open data and open science concepts in urban drainage. This includes standardized exchange formats for monitoring data (e.g. OGC-standards such as WaterML or InfraML) and a uniform meta-data description for machine-to-machine communication. To effectively improve the reusability of wastewater monitoring data, following FAIR data principles, it is crucial to improve existing ontologies, such as the urban water dictionary GSWS[9], to ensure clear and precise descriptions of the provided information.

8 Author contributions

MM conceived the original idea of making water infrastructure “transparent”, through advanced monitoring. FB conceptualized the idea, designed the data collection campaign, identified the study area, initiated, and maintained the cooperation with the municipality/communal utility, and continuously coordinated field work and research studies. Since 2016, all authors contributed to adapting the concept and design of the study, discussing the methods, and research opportunities and to the writing or revising the manuscript. Specifically, SB and CE developed, customized, tested and prepared monitoring and data communication hardware for the field. SB and CE implemented and maintained devices in the field, with support from AD and FB. SB, FB, CE, CF and AD collected and curated the observation data. Andreas Scheidegger (Eawag), Uwe Schmitt (SIS) and FB conceptualized the data pipeline for data collection, data quality control and data storage. Uwe Schmitt, CF, AD, SB, CE and FB implemented the data pipeline in software. CF improved the preliminary version of the webapp to view the data in RShiny. AD implemented the automated data quality control. CF, with the support of AD maintained the webserver and continuously ensured the interoperability of all data pipeline components. FB, MBR and JR prepared the SWMM model, with support from Joshi Prabhat and Claudia Keller. SB, FB and AD prepared the Geodatabase and accompanying information, such as metadata on sensor functioning and maintenance. JR conceptualized the manuscript, with the help of FB, SB and AD and prepared the submitted paper, which was approved by all authors. FB, MM and partly JR supervised the project.

[9] https://data.gwsw.nl/

9 Competing interests

The contact author has declared that none of the authors has any competing interests.

10 Acknowledgements

We acknowledge ETHZ, Eawag and the Urban Water Management Department for providing the main funding for this work. The authors are grateful to several organisations for agreeing to freely share data for the Urban Water Observatory project, especially the municipality of Fehraltorf and the wastewater utility Zweckverband ARA Fehraltorf-Russikon, who provided their own routine in-sewer operation data. Also, we thank the municipality to provide us permission to install our own sensors and run field campaigns and perform scientific experiments. This project would not have been feasible without their practical day-to-day support and their openness to support our research ideas. We therefore thank the team of Roman Kern, Beat Appenzeller, Michael Rüegg and especially Stefan Mathys for their invaluable help. We also thank other local stakeholders for their support, especially for the permission to place and operate our gateways at Electrosuisse and Schütz.

Furthermore, we greatly acknowledge the important work of Uwe Schmitt (Scientific IT services of ETH Zurich, ETH Zurich) for his great contribution in designing and coding the Python wrapper for the data warehouse application "Datapool". We also thank Andreas Scheidegger and Tobias Doppler from Eawag, who made important contributions for the conceptual foundations of the data pool and the adequate level of detail for the metadata. We also thank Rosi Siber from Eawag for help with the Geodatabase. Last, but not least, we would like to thank the many civil servants and master students who contributed to this work with their dedication and creativity.

11 References

Abdel-Aal, M., Villa, R., Jawiarczyk, N., Alibardi, L., Jensen, H., Schellart, A., Jefferson, B., Shepley, P., and Tait, S.: Potential influence of sewer heat recovery on in-sewer processes, Water Sci. Technol. J. Int. Assoc. Water Pollut. Res., 80, 2344–2351, https://doi.org/10.2166/wst.2020.061, 2019.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

AWEL: Mikroverunreinigungen - Messkampagne zu Belastungen aus Industrie und Gewerbe, AWEL, Zürich, 2021.

Benyon, R.: Letter from Richard Benyon MP to CEOs of water companies regarding monitoring of combined sewer overflows, 2013.

Bertrand-Krajewski, J.-L., Laplace, D., Joannis, C., and Chebbo, G.: Mesures en hydrologie urbaine et assainissement, Éditions Technique & Documentation, 793 pp., 2008.

Bianchi, B., Rieckermann, J., and Berne, A.: Quality control of rain gauge measurements using telecommunication microwave links, J. Hydrol., 492, 15–23, https://doi.org/10.1016/j.jhydrol.2013.03.042, 2013.

Blumensaat, F., Ebi, C., Dicht, S., Rieckermann, J., and Maurer, M.: Langzeitüberwachung der Raum-Zeit-Dynamik in Entwässerungssystemen mittels Niedrigenergiefunk - Ein Feldexperiment im Großmaßstab, Korresp. Abwasser, 64, 2017.

Blumensaat, F., Leitão, J. P., Ort, C., Rieckermann, J., Scheidegger, A., Vanrolleghem, P. A., and Villez, K.: How urban storm- and wastewater management prepares for emerging opportunities and threats: digital transformation, ubiquitous sensing, new data sources, and beyond – a horizon scan, Environ. Sci. Technol., 53, 8488–8498, https://doi.org/10.1021/acs.est.8b06481, 2019.

Blumensaat, F., Bloem, S., Ebi, C., Disch, A., Förster, C., Rodriguez, M., Maurer, M., and Rieckermann, J.: UWO - Accompanying data (2019 to 2021) [dataset], https://doi.org/10.25678/000991, 2024a.

Blumensaat, F., Bloem, S., Ebi, C., Disch, A., Förster, C., Rodriguez, M., Maurer, M., and Rieckermann, J.: UWO - Data access (2019 to 2021) [dataset], https://doi.org/10.25678/000980, 2024b.

Blumensaat, F., Bloem, S., Ebi, C., Disch, A., Förster, C., Rodriguez, M., Maurer, M., and Rieckermann, J.: UWO - Data viewer (2019 to 2021) [dataset], https://doi.org/10.25678/00092Z, 2024c.

Blumensaat, F., Bloem, S., Ebi, C., Disch, A., Förster, C., Rodriguez, M., Maurer, M., and Rieckermann, J.: UWO - Field observations (2019 to 2021) [dataset], https://doi.org/10.25678/00091Y, 2024d.

Boebel, M., Frei, F., Blumensaat, F., Ebi, C., Meli, M. L., and Rüst, A.: Batteryless Sensor Devices for Underground Infrastructure—A Long-Term Experiment on Urban Water Pipes, J. Low Power Electron. Appl., 13, 31, https://doi.org/10.3390/jlpea13020031, 2023.

Bustamante, G. R., Nelson, E. J., Ames, D. P., Williams, G. P., Jones, N. L., Boldrini, E., Chernov, I., and Sanchez Lozano, J. L.: Water Data Explorer: An Open-Source Web Application and Python Library for Water Resources Data Discovery, Water, 13, 1850, https://doi.org/10.3390/w13131850, 2021.

Caradot, N., Sonnenberg, H., Riechel, M., Matzinger, A., and Rouault, P.: The influence of local calibration on the quality of UV-VIS spectrometer measurements in urban stormwater monitoring, Water Pract. Technol., 8, 417–424, https://doi.org/10.2166/wpt.2013.042, 2013.

Clemens, F., Lepot, M., Blumensaat, F., Leutnant, D., and GRUBER, G.: Data validation and data quality assessment, in: Metrology in Urban Drainage and Stormwater Management: Plug and Pray, IWA Publishing, 327–390, 2021.

Deheer, K.: A Look Behind Using Machine Learning for Anomaly Detection, 2022.

Deletic, A., Dotto, C. B. S., McCarthy, D. T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T. D., Rauch, W., Bertrand-Krajewski, J. L., and Tait, S.: Assessing uncertainties in urban drainage models, Phys. Chem. Earth Parts ABC, 2011.

Disch, A.: Annotation Challenge, 2022.

Disch, A. and Blumensaat, F.: Messfehler oder Prozessanomalie? – Echtzeit-Datenvalidierung für eine zuverlässige Prozessüberwachung in Kanalnetzen, Aqua Urbanica 2019: Regenwasser weiterdenken - Bemessen trifft Gestalten, Rigi Kaltbad, 2019.

Disch, A., Rieckermann, J., and Blumensaat, F.: Challenges and Prospects in Anomaly Detection of Sewer Monitoring Data: Addressing Data Quality, Training Data Scarcity, and Randomness of Sewer Systems through Preprocessing and Multi-Sensor Integration, Water Sci Technol, in prep.

Dürrenmatt, D. J., Del Giudice, D., and Rieckermann, J.: Dynamic time warping improves sewer flow monitoring, Water Res., 47, 3803–3816, https://doi.org/10.1016/j.watres.2013.03.051, 2013.

Ebi, C., Schaltegger, F., Rust, A., and Blumensaat, F.: Synchronous LoRa mesh network to monitor processes in underground infrastructure, IEEE Access, 7, 57663–57677, https://doi.org/10.1109/ACCESS.2019.2913985, 2019.

EC: Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL concerning urban wastewater treatment (recast), 2022.

Eggimann, S., Mutzner, L., Wani, O., Schneider, M. Y., Spuhler, D., Moy de Vitry, M., Beutler, P., and Maurer, M.: The Potential of Knowing More: A Review of Data-Driven Urban Water Management, Environ. Sci. Technol., 51, 2538–2553, https://doi.org/10.1021/acs.est.6b04267, 2017.

Elías-Maxil, J. A., Hofman, J., Wols, B., Clemens, F., van der Hoek, J. P., and Rietveld, L.: Development and performance of a parsimonious model to estimate temperature in sewer networks, Urban Water J., 14, 829–838, https://doi.org/10.1080/1573062X.2016.1276811, 2017.

Environment Act: Environment Act 2021, 2021.

Ferriman, A.: BMJ readers choose the “sanitary revolution” as greatest medical advance since 1840, BMJ, 334, 111, https://doi.org/10.1136/bmj.39097.611806.DB, 2007.

Figueroa, A., Hadengue, B., Leitão, J., Rieckermann, J., and Blumensaat, F.: A distributed heat transfer model for thermal-hydraulic analyses in sewer networks, Water Res., 117649 (11 pp.), https://doi.org/10.1016/j.watres.2021.117649, 2021.

Giakoumis, T. and Voulvoulis, N.: Combined sewer overflows: relating event duration monitoring data to wastewater systems’ capacity in England, Environ. Sci. Water Res. Technol., 9, 707–722, https://doi.org/10.1039/D2EW00637E, 2023.

Hadengue, B., Joshi, P., Figueroa, A., Larsen, T. A., and Blumensaat, F.: In-building heat recovery mitigates adverse temperature effects on biological wastewater treatment: A network-scale analysis of thermal-hydraulics in sewers, Water Res., 204, 117552, https://doi.org/10.1016/j.watres.2021.117552, 2021.

HBT: GEP Fehraltorf - Gewässer, Hunizker Betatech Engineering, Winterthur, 2016.

Hoppe, H., Fricke, K., Kutsch, S., Massing, C., and Gruber, G.: Von Daten zu Werten – Messungen in Entwässerungssystemen, Aqua Gas, 96, 26–31, 2016.

Huisman, J. L.: Transport and transformation processes in combined sewers, ETH Zürich, 2001.

IEC 81346-2: IEC 81346-2:2019, Industrial systems, installations and equipment and industrial products — Structuring principles and reference designations — Part 2: Classification of objects and codes for classes, 2019.

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

van Kranenburg, R. and Bassi, A.: IoT Challenges, Commun. Mob. Comput., 1, 9, https://doi.org/10.1186/2192-1121-1-9, 2012.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Krejci, V., Fankhauser, R., Gammeter, S., Grottker, M., Harmuth, B., Merz, P., and Schilling, W.: Integrierte Siedlungsentwässerung Fallstudie Fehraltorf, Dübendorf Eawag 1994 303 P Schriftenreihe Eawag Vol 8 ISBN 3-906484-09-2, 1994.

Lepot, M., Torres, A., Hofer, T., Caradot, N., Gruber, G., Aubin, J.-B., and Bertrand-Krajewski, J.-L.: Calibration of UV/Vis spectrophotometers: A review and comparison of different methods to estimate TSS and total and dissolved COD concentrations in sewers, WWTPs and rivers, Water Res., 101, 519–534, https://doi.org/10.1016/j.watres.2016.05.070, 2016.

Manny, L., Duygan, M., Fischer, M., and Rieckermann, J.: Barriers to the digital transformation of infrastructure sectors, Policy Sci., https://doi.org/10.1007/s11077-021-09438-y, 2021.

Mathis, S., Gruber, J.-M., Ebi, C., Bloem, S., Rieckermann, J., and Blumensaat, F.: Energy self-sufficient systems for monitoring sewer networks, in: Sensors and Measuring Systems; 21th ITG/GMA-Symposium, Sensors and Measuring Systems; 21th ITG/GMA-Symposium, 1–8, 2022.

Mourad, M. and Bertrand-Krajewski, J. L.: A method for automatic validation of long time series of data in urban hydrology, Water Sci. Technol., 45, 263–270, 2002.

Nedergaard Pedersen, A., Wied Pedersen, J., Vigueras-Rodriguez, A., Brink-Kjær, A., Borup, M., and Steen Mikkelsen, P.: The Bellinge data set: open data and models for community-wide urban drainage systems research, Earth Syst. Sci. Data, 13, 4779–4798, https://doi.org/10.5194/essd-13-4779-2021, 2021.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., and ten Veldhuis, M.-C.: Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation, J. Hydrol., 531, Part 2, 389–407, https://doi.org/10.1016/j.jhydrol.2015.05.035, 2015.

Panasiuk, O., Hedström, A., Langeveld, J., and Viklander, M.: Identifying sources of infiltration and inflow in sanitary sewers in a northern community: comparative assessment of selected methods, Water Sci. Technol. J. Int. Assoc. Water Pollut. Res., 86, 1–16, https://doi.org/10.2166/wst.2022.151, 2022.

Regueiro-Picallo, M., Anta, J., Naves, A., Figueroa, A., and Rieckermann, J.: Towards urban drainage sediment accumulation monitoring using temperature sensors, Environ. Sci. Water Res. Technol., https://doi.org/10.1039/D2EW00820C, 2023.

Rieckermann, J., Bertrand-Krajewski, J.-L., Blumensaat, F., Ort, C., Pistocchi, A., and Schellart, A.: Assessing Combined Sewer Overflows (CSOs) - A growing need for evidence base, compliance assessment, and future regulation, 15th ICUD - International Conference on Urban Drainage, 2021.

Rossi, L., Chevre, N., Fankhauser, R., and Krejci, V.: Probabilistic environmental risk assessment of urban wet-weather discharges: an approach developed for Switzerland, Urban Water J., 6, 355–367, 2009.

Ruggaber, T. P., Talley, J. W., and Montestruque, L. A.: Using Embedded Sensor Networks to Monitor, Control, and Reduce CSO Events: A Pilot Study, Environ. Eng. Sci., 24, 172–182, https://doi.org/10.1089/ees.2006.0041, 2007.

Russo, S., Disch, A., Blumensaat, F., and Villez, K.: Anomaly Detection using Deep Autoencoders for in-situ Wastewater Systems Monitoring Data, ETHZ, Copenhagen, 2019.

Russo, S., Besmer, M. D., Blumensaat, F., Bouffard, D., Disch, A., Hammes, F., Hess, A., Lürig, M., Matthews, B., Minaudo, C., Morgenroth, E., Tran-Khac, V., and Villez, K.: The value of human data annotation for machine learning based anomaly detection in environmental systems, Water Res., 206, 117695, https://doi.org/10.1016/j.watres.2021.117695, 2021.

Schilperoort, R., Hoppe, H., de Haan, C., and Langeveld, J.: Searching for storm water inflows in foul sewers using fibre-optic distributed temperature sensing, Water Sci. Technol. J. Int. Assoc. Water Pollut. Res., 68, 1723–1730, https://doi.org/10.2166/wst.2013.419, 2013.

Semtech Corporation: AN1200.22: LoRa™ Modulation Basics, 2015.

Špačková, A., Bareš, V., Fencl, M., Schleiss, M., Jaffrain, J., Berne, A., and Rieckermann, J.: A year of attenuation data from a commercial dual-polarized duplex microwave link with concurrent disdrometer, rain gauge, and weather observations, Earth Syst. Sci. Data, 13, 4219–4240, https://doi.org/10.5194/essd-13-4219-2021, 2021.

Spraakman, S.: IAHR/IWA JOINT SPECIALIST GROUP ON URBAN DRAINAGE – MARCH 2023 NEWSLETTER 36, 2023.

Staufer, P., Scheidegger, A., and Rieckermann, J.: Assessing the performance of sewer rehabilitation on the reduction of infiltration and inflow, Water Res., 46, 5185–5196, https://doi.org/10.1016/j.watres.2012.07.001, 2012.

Taylor, P., Cox, S., Walker, G., Valentine, D., and Sheahan, P.: WaterML2.0: development of an open standard for hydrological time-series data exchange, J. Hydroinformatics, 16, 425–446, https://doi.org/10.2166/hydro.2013.174, 2013.

Wani, O., Maurer, M., Rieckermann, J., and Blumensaat, F.: Does distributed monitoring improve the calibration of urban drainage models?, in: 12th Urban Drainage Modeling Conference, Costa Mesa, California, January 2022, Urban Drainage Modelling conference (UDM), Costa Mesa, USA, 2022.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., ’t Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The

FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.