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 barriers, 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 cover~~s~~ 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 applications from detecting anomalies in the data to assessing groundwater infiltration and the capability of the low-power data transmission. We also suggest directions for future research using the UWO dataset from uncertainty analysis of rainfall runoff models to smart databases. 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 (Caradot et al., 2013; Deletic et al., 2011; Lepot et al., 2016; Ochoa-Rodriguez et al., 2015), it is not well established, neither in the research community nor among practitioners.

Understanding urban hydrological processes requires high-resolution data on both, the input - such as rainfall - and the output - such as wastewater flows and pollution. For rainfall ~~runoff~~, a high temporal resolution seems to be more important than spatial resolution, though both interact closely (Ochoa-Rodriguez et al., 2015), although ~~. However,~~ reliably operating rain gauges ~~collecting minute-by-minute rainfall data~~ over many years can be challenging (Bianchi et al., 2013). ~~, as it requires high frequency sensors that must operate reliably in harsh environments such as sewer pipes.~~ Similarly, ~~M~~monitoring stormwater runoff, wastewater flows and pollution processes at the~~a~~ minute scale is resource-consuming, as it requires high-frequency sensors that must operate reliably in harsh environments such as sewer pipes. Specialized equipment, training, and software (Dürrenmatt et al., 2013; Mourad and Bertrand-Krajewski, 2002) as well as considerable investments are required to collect and manage the data, as well as to maintain the sensors (Blumensaat et al., 2019; Hoppe et al., 2016). Arguably, sensor maintenance to ensure good data quality is one of the biggest challenges in urban monitoring, as sensors can drift, foul, or lose power without notice (Mourad and Bertrand-Krajewski, 2002; Nedergaard Pedersen et al., 2021). Data quality is often affected by limited familiarity of operational staff with data handling and interpretation as well as a 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, especially when trying to reuse them across different systems, tools, or studies.

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, several developments have contributed to a growing push for open datasets in urban hydrology. First, advancements in low-power electronics, data transmission (Ebi et al., 2019) and sensor application (Boebel et al., 2023; Mathis et al., 2022) have significantly reduced the effort of data collection. Second, standardized meta-data and exchange formats facilitate data sharing (Bustamante et al., 2021; Taylor et al., 2013). Third, scientific data collection efforts, such as the data set "Catchment Attributes and Meteorology for Large-sample Studies (CAMELS)" (Addor et al., 2017; Newman et al., 2015), 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 (Benyon, 2013; Giakoumis and Voulvoulis, 2023) and regulatory bodies, such as those in the UK (Environment Act, 2021) and in the EU (EC, 2022), are demanding more monitoring. At the same time industry initiatives, such as STREAM (Stream - Portal, 2025), 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 (NYS Combined Sewer Overflows (CSOs), 2022; Riveraction, 2025), 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 4b), and maintains a consistent quality. Thus, the UWO dataset offers extensive research opportunities, ranging 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.

The remainder of the article is structured as follows (Figure 1): we first describe the catchment of Fehraltorf and then the sensor data. We then describe the methods used to collect, process and explore the data and the accompanying data, including a hydraulic rainfall-runoff model implemented in SWMM. Finally, we highlight five exemplary research opportunities . We emphasize that the main novelty of this work is the dataset itself, not the described methods for data curation and cleaning or the provided examples, which motivate future research on the UWO dataset.

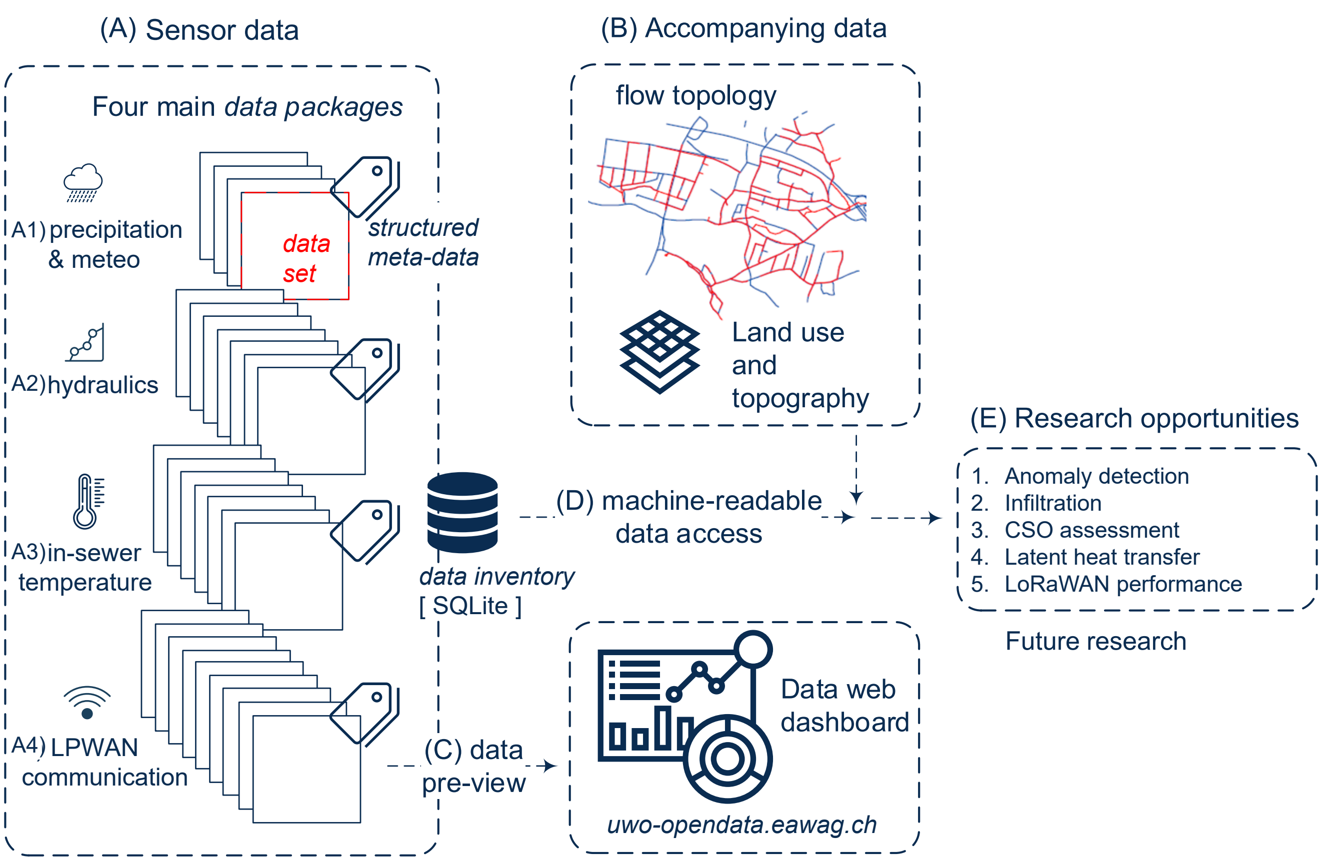


Figure 1: Graphical representation of data provided (A-B), the current data viewer (C), scripts to access the data (D), and examples to highlight future research opportunities in the prospect of the UWO dataset (E).

2. Description of the catchment2.1 Description of the study area

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

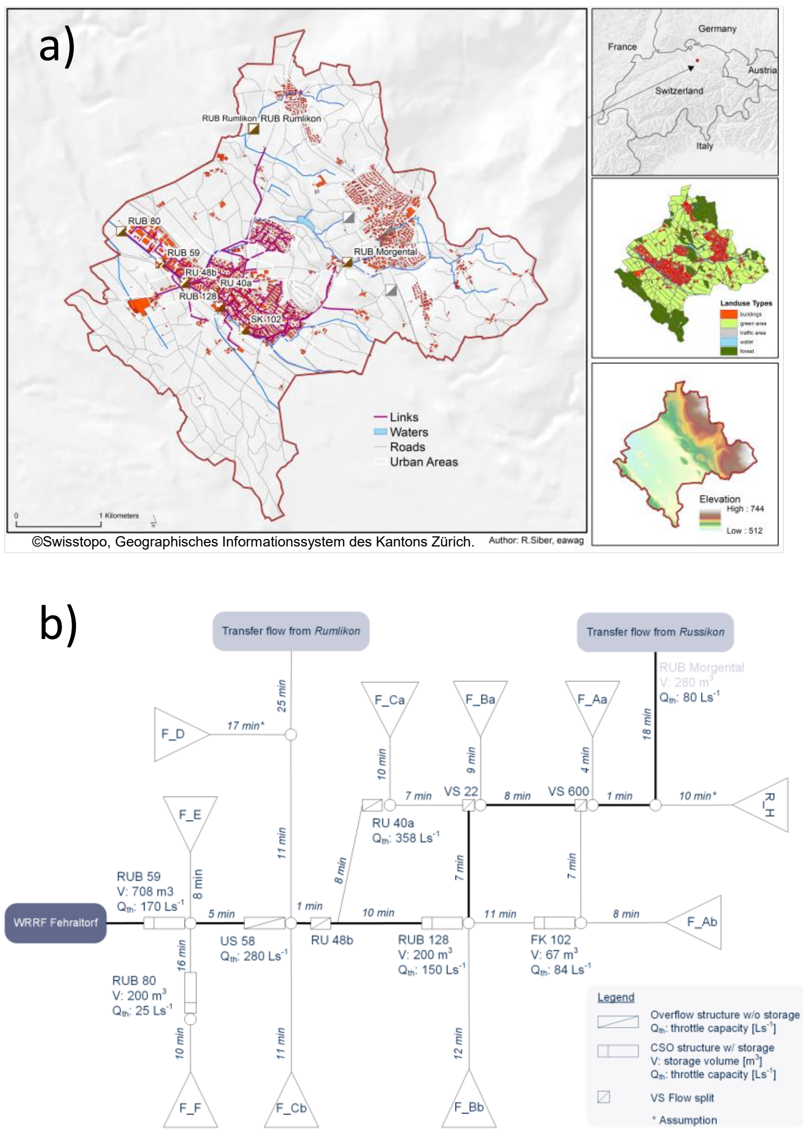


Figure 2a): Overview of the Fehraltorf catchment, including urban drainage network, geography and land use. b): Simplified flow scheme of the combined sewer system of Fehraltorf. The main flow path through the network is from top right to the left (bold line). The relevant characteristics of sub-catchments can be found in the supporting information. Indicated flow times are calculated assuming a constant 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 2a).

The climate in Fehraltorf is a typical humid continental climate and is classified *Dfb* according to the Köppen-Geiger climate classification (Speck-Fehraltorf Airport | SKYbrary Aviation Safety, 2025) 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 Section S1 in the Supporting Information (SI), for details). Our measurements show annual rainfall totals between 1096–1339 mm, compared to 833–1083 mm from the MeteoSwiss station KLO. The climate region has a high variability in weather patterns and the frequency of storms and extreme weather events. The intensity for a 5-minute rainfall with a 20-year return period is estimated at 134 mm/h (95% CI: 118–168 mm/h) according to new extreme value statistics from MeteoSwiss (see Section S1.6).

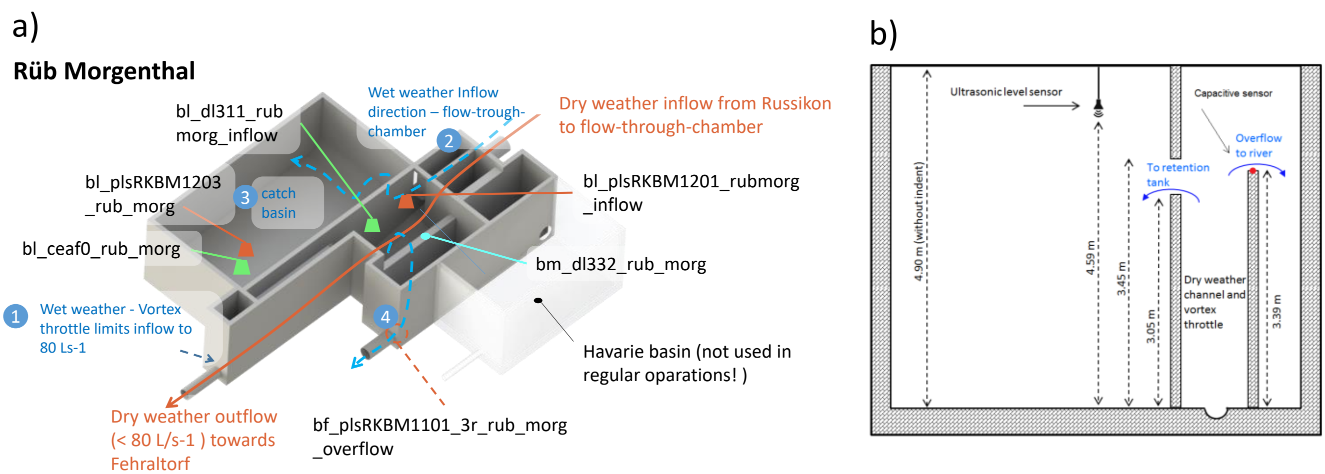


Figure 3a): CAD model of the overflow structure "RÜB Morgental" including inflow chamber, detention tank, overflow to receiving water, and installed sensors, b): Cross-section at the sensor bl_plsRKBM1201_rubmorg_inflow (not drawn to scale). Detailed information of all other detention basins and flow 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 2b, 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). Only a very small neighbourhood of Russikon (Madetswil) connects to a different WW catchment. The foul sewage contribution from this area is negligible; the contributing area is not included. Also included in the dataset is the RUB Morgental (Figure 3a), the catchment-concluding combined sewer overflow (CSO) structure in Russikon that limits the transfer flow to 80 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.

3 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 4b, A1), wastewater hydraulics (A2), temperature of wastewater and sewer atmosphere (A3), and metrics evaluating the performance of the LoRaWAN network (A4). LoRa® ("Long-Range") is a proprietary wireless modulation technique; LoRaWAN is its protocol, now standardized as ITU-T Y.4480 by the ITU. In our definition, a *sensor* is the actual device that measures a physical parameter (e.g. water level, temperature, flow, rainfall). A *sensor node* 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 4.1.

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.

In the following subsections we give an overview of the measurement types, sensors and data transmission.

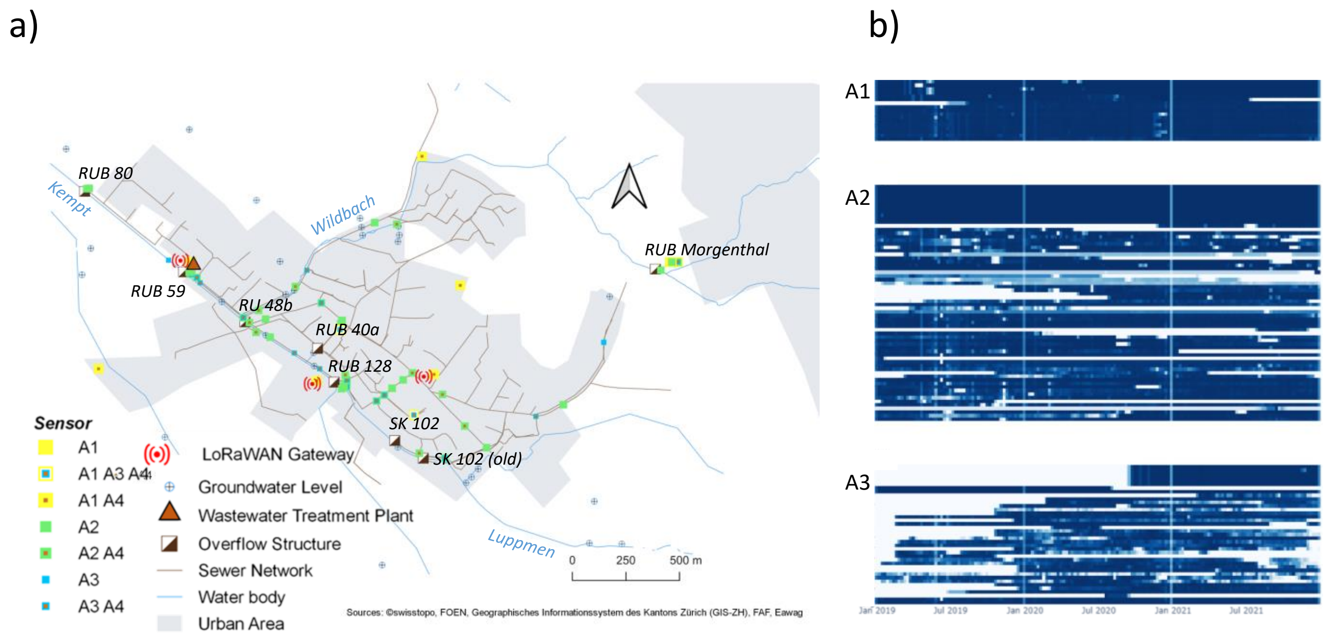


Figure 4a): Georeferenced locations of sensors, b): Example of a heat map representing the completeness of data in the monitoring period from 01-01-2019 to 31-12-2021. A1 = precipitation data; A2 = hydraulic data; A3 = temperature data. A4= LPWAN sensor data is not shown. The color saturation indicates the degree of data completeness (weekly granularity). Dark blue indicates periods with 100 % data completeness; white indicates periods with no data. Light blue indicates a reduced number of data points, either through sensor maintenance, sensor outage or incomplete data 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 S6 for details. Groundwater levels are not part of this dataset, but available at ERIC (Ramgraber, 2025).

Table 1: Overview of installed sensors including relevant characteristics on sensor type, temporal granularity of data, and the type of data transmission.

		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							

3.1 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 Bellinge rainfall data, which come from a national network (Nedergaard Pedersen et al., 2021) and are likely quality-controlled and processed using standardized tools, the UWO rain gauges were installed specifically for local monitoring, providing more targeted and detailed rainfall-runoff data although their measurements have not been quality controlled a priori. Nevertheless, the OTT Pluvios were checked twice per year and they operated reliably during the three-year period and measured the data in 1 min intervals. 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.

3.2 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 6.1.3) 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 4a) 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.

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 S4.4).

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 S4). 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.

3.3 Temperature measurements (A3)

Temperature can be used as a natural tracer to provide information on hydraulics (Dürrenmatt et al., 2013), groundwater infiltration (Panasiuk et al., 2022; Schilperoort et al., 2013) 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 4a). 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.

3.4. 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, 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 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 6.1.5. 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 (Section 4.3.1).

4 Data Pipeline and Quality Assurance

4.1 Data collection and transmission

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).

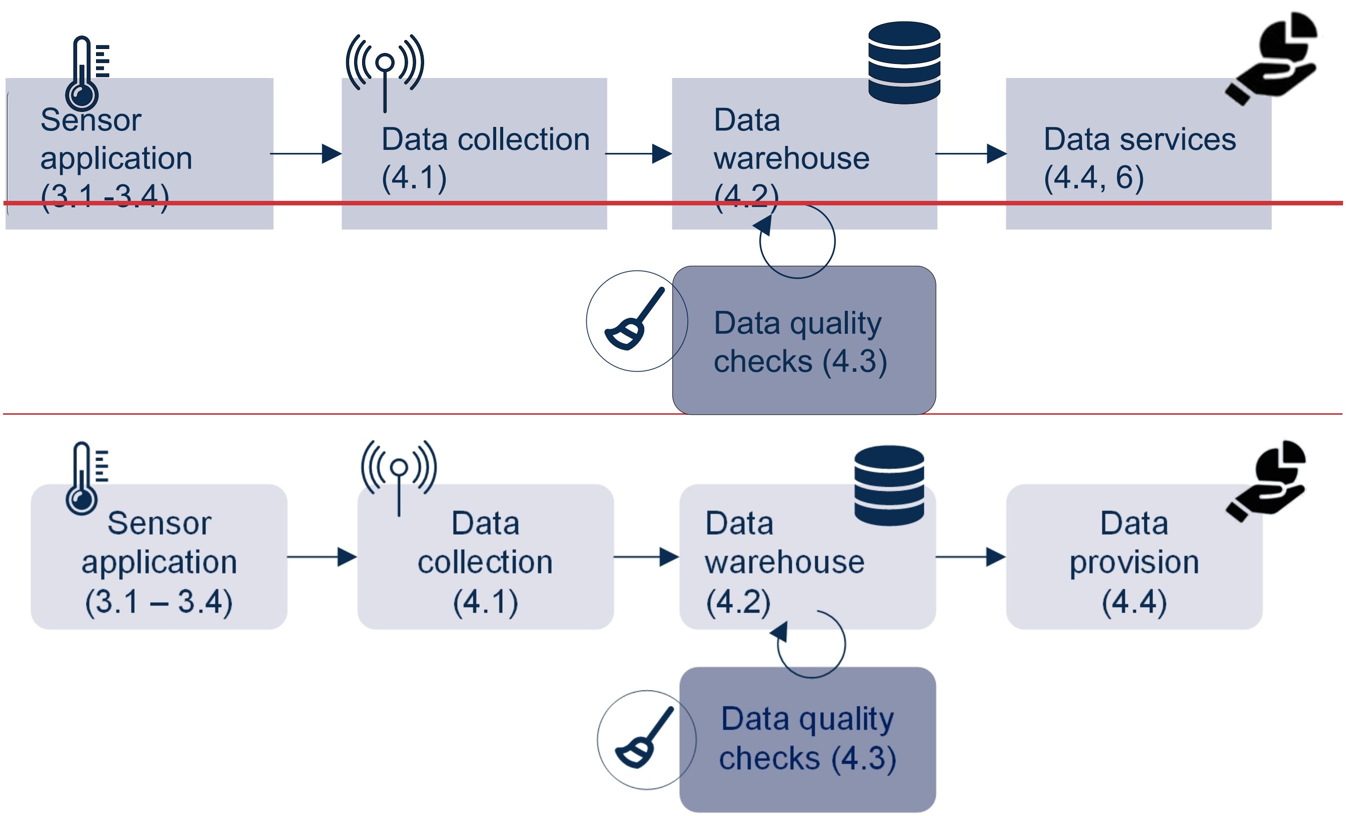


Figure 5: The individual components of the data pipeline for collecting, storing, and using UWO data from the sensor to the data service. The numbers refer to the corresponding sections.

4.2 Data storage and structure

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 S5). 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.

4.3 Data quality control

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.

4.3.1 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 (Disch and Blumensaat, 2019; Russo et al., 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. Upper and lower limits for range checks are individual for each sensor as they are defined through i) the device specification and configuration as well as ii) constraints given through actual installation in the field. 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. The values implemented in gradient tests vary depending on the observed system dynamics, i.e. they were manually adjusted to obtain meaningful results. 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 4.3.1).

4.3.2 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. The continuous verification of consistency and homogeneity helps to ensure that the UWO dataset reflects real-world processes adequately across time and space. In this context we define consistency and homogeneity as follows:

- **Data consistency** refers to the sufficient availability of plausible data over time. Specifically, data should not contain large gaps, inexplicable jumps or unrealistic values that conflict with the known behaviour or physical constraints of an urban drainage system, e.g. manhole water levels beyond terrain level.
- **Data homogeneity** refers to the uniformity of data characteristics over the observation period. For example, a sensor's signal should consistently represent the same physical quantity under the same conditions. Changes in installation, sensor type, or surrounding hydraulics that could affect the signal interpretation were tracked to ensure the dataset remains temporally coherent.

More specifically, we accomplish this by i) a visual inspection of sensor data time series every morning, reviewing new sensor data that has arrived overnight in the context with historic data. With the growing number of UWO sensors, i.e. data collected, this process was supported by ii) an automated flagging routine based on range and gradient checks (see Section 4.3.1). This daily inspection routine enabled us to identify anomalies early (e.g. due to sensor malfunction, clogging, configuration changes, or physical alterations to the sewer system), and to timely address these phenomena by on-demand sensor maintenance in the field. While basic automated data validation routine (step ii) help to filter out "questionable data", expert knowledge is used in i) to further differentiate observed phenomena (is it a sensor failure or just typical, normal system behaviour?) in "questionable" data and identify errors and the need for sensor maintenance. We differentiate between "doubtful" and "undoubtedly" data points. The so as "doubtful" identified data were kept in the dataset but annotated, enabling users to exclude or inspect them as needed. ~~Clemens et al., 2021 propose more classes, but we found that this was difficult to implement in our case.~~

In addition to regular data consistency and homogeneity checks, advanced data validation methods, such as anomaly detection, have been proposed in literature (Clemens et al., 2021; Russo et al., 2019, 2021). However, these are a highly challenging topic of current research (Deheer, 2022; Disch and Blumensaat, 2019; Rieckermann and Disch, 2024) 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 6.

4.4 Data access and tools

To support open and reproducible science, data from the Urban Water Observatory (UWO) are shared according to the FAIR principles—Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). First, the data are published via the Eawag Research Data Institutional Repository (ERIC/open) with persistent DOIs and descriptive metadata. Each dataset is labelled using standardized formats and appropriate keywords to facilitate discovery. Metadata on sensor sources, locations, and signal types are documented in a structured format and include quality annotations (see SI Section S5, Table "source_meta_data"). Second, the validated dataset covering 2019–2021 is provided in annual slices and can be downloaded as self-contained SQLite databases. Public access is granted via ERIC/open, which currently offers only limited data exploration. Therefore, we provide an R-Shiny web dashboard for interactive preview and filtering by signal type, source, or location (Figure 6) (Eawag-SWW, 2025). The downloadable SQLite databases contain a validated and frozen subset (2019–2021) with broad sensor coverage, while the online viewer shows real-time data that may include artefacts, as automated quality control is still evolving. To prevent misuse of unchecked data, we restrict downloads to the validated subset and plan to release further curated data over time.

Third, data are provided in open, machine-readable formats (SQLite, CSV, ASCII) to support integration into various workflows. Sensor meta-data describe the methods and devices used for data collection and processing. This structure supports downstream use in simulation models, statistical tools, or real-time dashboards. The data are reusable, because they are released under the CC0 license, enabling unrestricted use, modification, and redistribution. 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 (https://doi.org/10.25678/000C5K). 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).

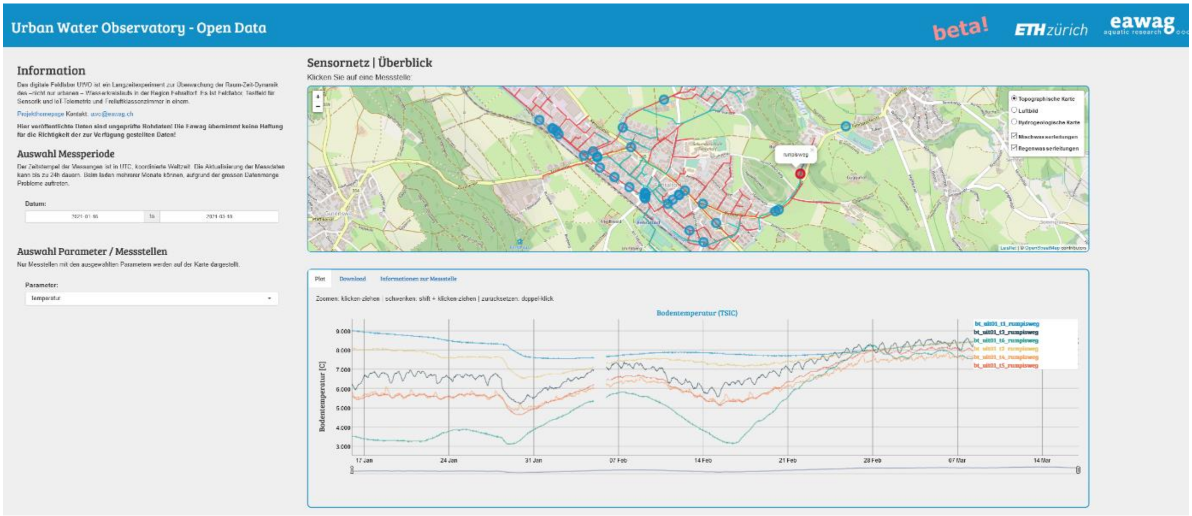


Figure 6: Online pre-viewer to the UWO dataset, which includes information on the location of the data source, as well as time series and meta-data on the characteristics of the sensor and the monitoring site. Web access: https://uwo-opendata.eawag.ch/, Source: © OpenStreetMap contributors 2024. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

5 Simulation model

5.1 Hydraulic rainfall-runoff model implementation in SWMM

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. 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.

5.2 Model calibration

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). The final calibration of the provided SWMM model was based on measured inflow to the WWTP and flow observations at four additional locations during several non-extreme rain events (March–May 2016). The model 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. Details are described in Section S7.3 and Figure S41.

Wani et al. (2022) investigated the benefit of using multiple data sources in model calibration and performed a spatially differentiated calibration using a Bayesian approach using 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 commissioning of new flow-limiting hardware in August 2020, modifications were made to the network that are not yet considered in the currently provided model structure (*cf.* SI Section S3.5). Further information on the rainfall-runoff model, as well as on the accompanying data can be found in the SI.

6. 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 first demonstrate the validity and applicability of the UWO dataset by briefly highlighting five specific research opportunities using the UWO dataset from automated data quality checks to improved event-duration monitoring of combined sewer overflows and wireless sensor network performance (seeSI Section S9 for details). Second, we suggest five directions of future research from optimal sensor placement and adaptive sensing to uncertainty analysis and semantic developments that enable natural language queries to urban drainage datasets.

6.1 Applications of the datasets

6.1.1 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 (Kerkez et al., 2016; Ruggaber et al., 2007). 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 S9.1). 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~~Figure 7~~).

The F1 scores for monitoring data with synthetic errors during wet weather (Figure 7~~Figure 7~~b): 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.

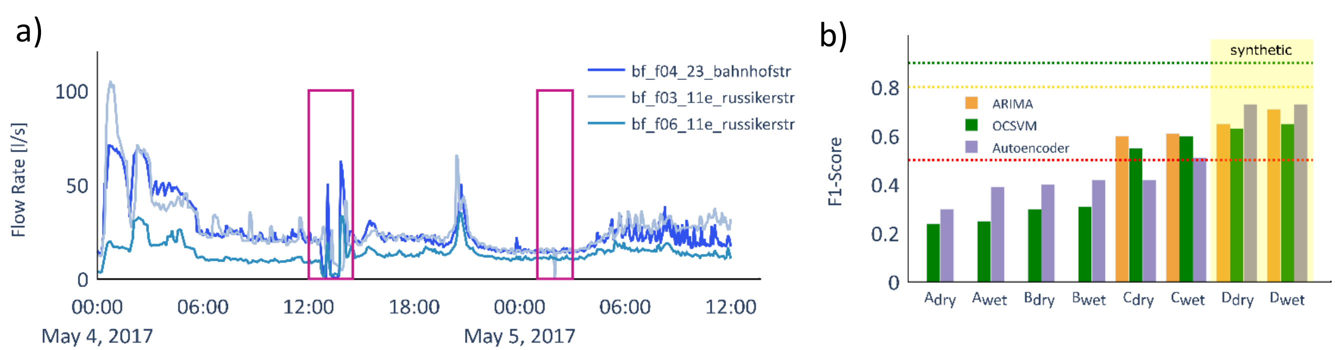
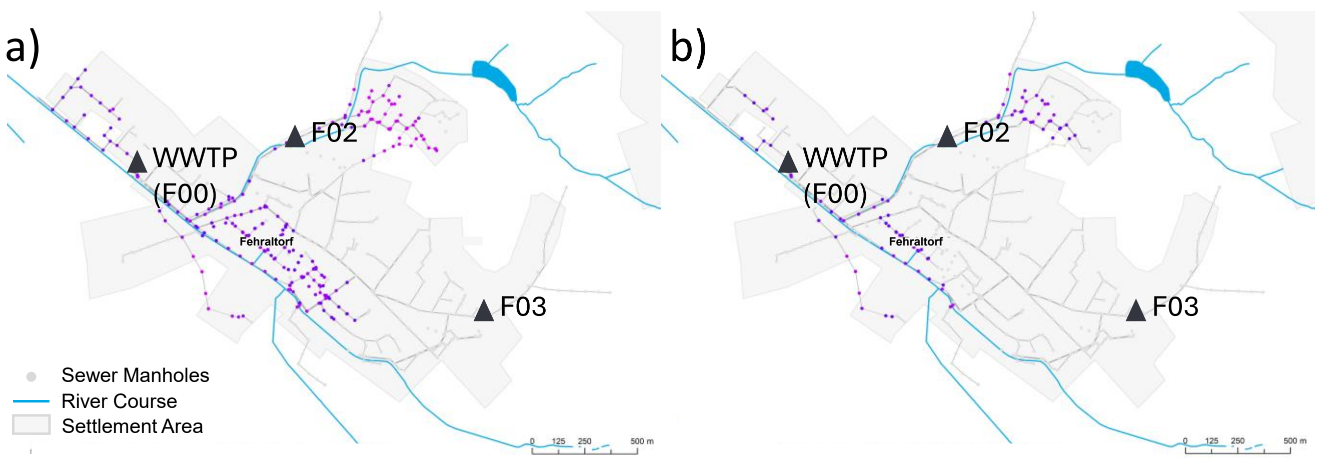


Figure 7a): Observations with potential anomalies. Highlighted periods show the benefit of multiple sensors in detecting anomalies. While the high variability on the left is reflected by all sensors, on the right only a single signal appears abnormal, b): F1 scores for different methods for pre-processing (A-D). An F1 score of 0.5-0.8 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).

6.1.2 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 the contribution of groundwater infiltration (GWI) rates from the Fehraltorf catchment, 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 as the difference between GWI at the outlet (F00) and the inflow, i.e. two upstream contributions from Rumlikon and Russikon (F02 and F03).

GWI rates ranged from 10 to 15 Ls^{-1}, depending on the season (see SI Section S9.2). The detailed monitoring data revealed spatial variations in GWI with changing seasons. For example, in April 2018, using the data from Ramgraber (2025), a high groundwater table affected 256 out of 459 manhole inverts, while in October 2018, only 100 manholes were impacted (Figure 8). Future work could develop spatially detailed GWI rates to better capture the sewer-groundwater interactions. This could build on recent work, which added a groundwater module to the SWMM model and simulated dynamic infiltration in Fehraltorf (Rodriguez Bennadji, 2022; Rodriguez et al., 2024).



Source: swisstopo (2020), TLM, swissbuildings, BA für Landestopographie (Art. 30, Geo IV), 5704 000 000, Reproduced with permission/ JA 100119

Figure 8a): 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, b): In total 100 sewer manholes in the Fehraltorf network were affected by groundwater in October 2018.

6.1.3 The value of redundant sensors in event-duration monitoring

Assessing CSO activity is crucial for quantifying pollution and optimizing sewer networks. Typically, level sensors are used to monitor water levels in the tank and next to overflow weir crests. This inside-tank level data is then used to derive overflow volumes or even spill duration. The latter process can be a very difficult topic, because it essentially means splitting a continuous process into “events”, which are not precisely defined (hysteresis). In addition, a single (level) sensor can drift or change over time due to issues such as temperature influence and the presence of spider webs. In our example, a second, capacitive sensor serves as a rather robust indicator for spill duration, i.e. the start and stop of a spill event. This becomes even more relevant as 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 (EC, 2022; Rieckermann et al., 2021). However, ensuring data quality is a real challenge and research explores various monitoring techniques to make data-based compliance assessment more reliable.

In the UWO, we equipped all CSO tanks in the Fehraltorf system with multiple ultrasonic level and capacitive sensors to independently monitor overflow duration and to investigate how redundant signals would reduce the uncertainty of CSO event-duration monitoring (SI, Section S9.3). The results shown in Figure 9b) and c) illustrate that the CSO duration derived from an erroneously calibrated level sensor (Fig. 9b - dl311, x-axis) can significantly differ from the CSO duration derived from a correctly configured level sensor (Fig. 9c - dl311, x-axis). Only by comparing the level-sensor prone information with the CSO duration derived from the capacitive sensor signal (reference signal), the systematic deviation due to the wrongly configured level sensor becomes obvious. In the given example, with the incorrect level sensor configuration the CSO activity is severely underestimated - about 50 % less cumulative overflow duration is documented in the overall monitoring period of 1'077 days. Generally, the redundant information from a robust capacitive sensor (only dry/wet status; consecutive wet intervals are counted to give an overflow duration) helps to verify data from typically implemented level sensors. Once successfully verified, level data can then be used to quantify not only overflow duration but also overflow volume, i.e. spill rates. Capacitive sensors alone provide reliable information on overflow duration and overflow frequency.

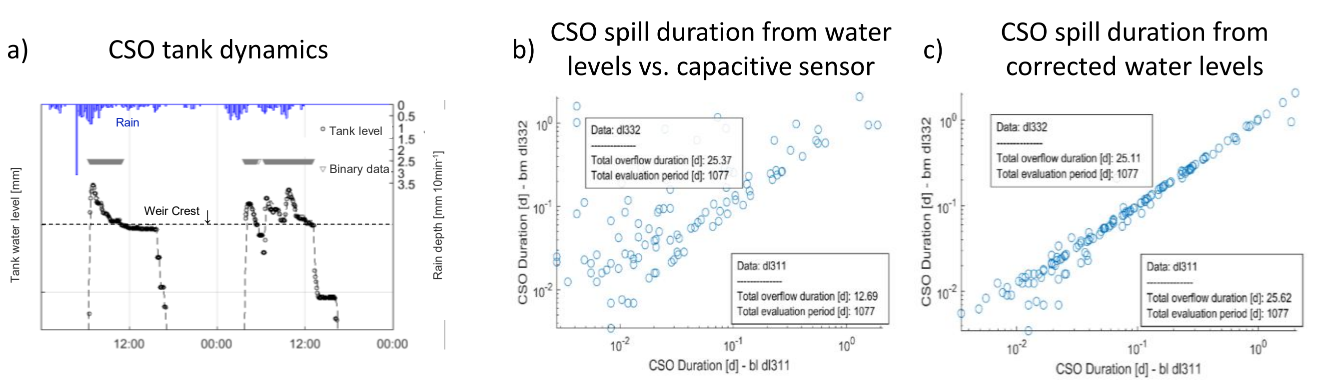


Figure 9a): Continuous tank level measurements (dots) and binary data (triangles). The latter are derived from the capacitive sensor signal, reflecting an overflow activity during a period of two days for which two independent overflow events were recorded, b) and c): Event-specific overflow durations derived from one capacitive sensor (bm_dl332_rub_morg) and one ultrasonic level sensor (bl_dl311_rubmorg_inflow). The overflow duration derived from the erroneously calibrated level sensor (dl311) with an incorrect offset in b) is corrected to a correct offset in c).

6.1.4. 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 (Abdel-Aal et al., 2019; Figueroa et al., 2021; 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)). Headspace temperature refers to the air above the wastewater surface in the sewer pipe, bulk liquid temperature to the wastewater itself. To do this, we analysed the provided temperature data in three locations dl933, dl935 and dl931 (Figure 10), assuming a fixed relative humidity of 91.4%, which is the mean value observed across several sewer locations (see SI of Figueroa et al. (2021), Figure 2). This simplification enabled a didactical classification of temperature observations into evaporation (red) and condensation (blue) periods, based solely on the temperature gradient between the headspace and the wastewater. We found that condensation takes place in April when the typical temperature difference between the sewer headspace and the bulk liquid temperature was -2 °C. In contrast, with warmer ambient air and headspace temperatures, evaporation occured. Maximum differences amounted to +5 °C, at location dl931. Based on these results, Figueroa et al. (2021) concluded 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.

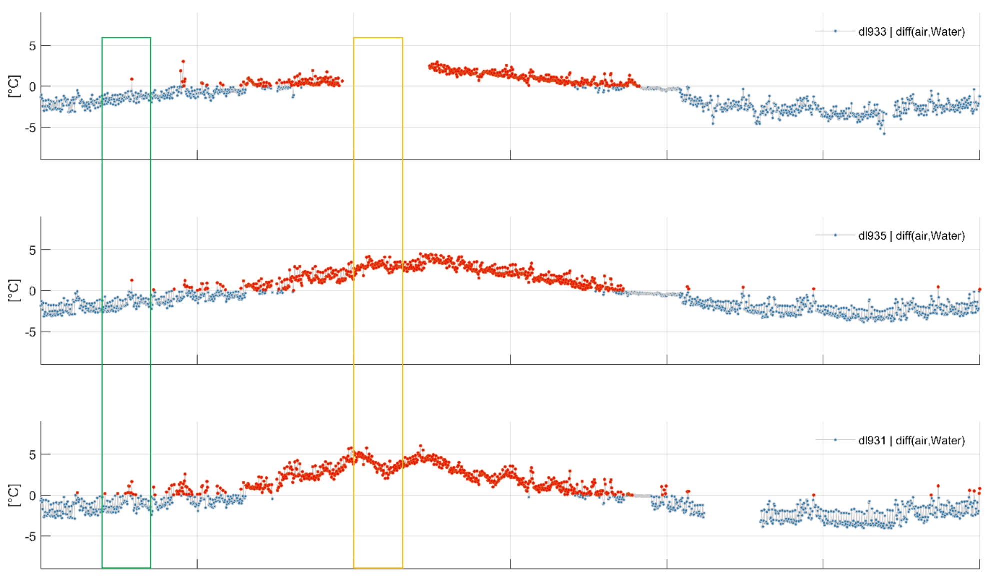


Figure 10: The difference between the sewer headspace temperature and the bulk liquid temperature observed at three specific locations in the main collector. Red dots indicate periods of evaporation (higher temperature in the headspace compared to the bulk liquid) and blue colours indicate periods of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

6.1.5 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 S9.4).

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)

6.2 Future research directions

In the future, our drainage systems should be more adaptive and resilient, and we should have a good understanding of their performance. Future research ~~should~~ can therefore focus on optimizing sensor placement for both understanding and control, developing predictive models that account for uncertainty, and adopting adaptive monitoring strategies. Advancing FAIR data infrastructure will further support broad, intuitive access and reuse of high-resolution observational data, e.g. for machine learning and data-driven modelling.

6.2.1 Towards robust and automated data quality checks

Although the UWO dataset is unique and carefully curated, it still contains sensor failures and inconsistencies. This reflects a broader challenge: robust and automated data quality control methods are not yet well established. Across metrological (Bertrand-Krajewski et al., 2025), scientific (Disch and Blumensaat, 2019), and industry (Deheer, 2022) perspectives, the key bottleneck to date is not algorithm complexity but data readiness. The UDMT approach emphasizes that validation should be physically informed, traceable, uncertainty-aware, transparent, and reproducible. Automated checks should therefore combine calibration information, measurement uncertainty, and hydraulic plausibility (e.g., pipe capacity), rather than relying solely on statistical anomaly detection.

Machine-learning trials further show that structured preprocessing (range and gradient checks, normalization, gap handling, dry/wet separation) has a stronger influence on performance than the choice of the machine-learning model (see Section S9.1, Deheer, 2022). They also highlight the need for clearer anomaly definitions, multi-sensor integration, cautious use of complex models, and synthetic data to address limited labelled events. Key questions for future research using the UWO dataset could therefore explore: “How can automatic checks better use simple hydraulic rules and known sensor errors?” “How can clear and shared definitions of “anomalies” be established?” “How can sensors, physical models, and simple data methods be combined for reliable operational quality control?”

6.2.2~~1~~ Optimizing sensor placement and sampling for system understanding and control

Monitoring networks in urban drainage systems are important to understand the behaviour of the system and to enable real-time control, but they are also challenging to maintain and manage. Future research ~~should~~ can therefore focus on methods for sensor placement that maximize the value of the data collected. Key research questions include: “How many locations should we monitor with which measurement type to capture critical flow dynamics?” “Which monitoring points provide little (added) value?”. Addressing these questions requires a systematic assessment of the information contribution of each sensor location, possibly using information theory or graph-based methods (Crowley et al., 2025; Villez et al., 2016).

Also, future research can~~should~~ investigate adaptive spatio-temporal sampling strategies. Sewer observation using uniform high-frequency monitoring is rarely sustainable, as it consumes power and continuously changing batteries puts a practical limit to the number of sensors a utility can maintain. Future monitoring strategies ~~should~~ could therefore be adaptive, adjusting logging intervals based on system needs and external triggers, e.g. weather radar. Key questions include “When should sensors switch to high-frequency mode?” and “How can targeted sampling schemes for sewer capacity or groundwater infiltration or routine monitoring look like?” The UWO dataset is dense, well-documented and, with the accompanying SWMM model, provides an excellent foundation for sensor placement research.

6.2.3~~2~~. ~~Adapting Urban Drainage to Climate Change with Grey-Green Infrastructure~~Temperature-Related Processes in Urban Drainage Systems

Urban drainage systems are thermodynamic systems in which temperature strongly influences physical, chemical, and biological processes. Wastewater temperature governs heat recovery potential, seasonal heat losses, and temperature-dependent reactions such as H_2S formation. Understanding in-sewer temperature dynamics is therefore essential for advancing water–energy nexus research and improving sewer operation. Key research questions in this domain that can be investigated with the UWO dataset include: “How does wastewater temperature propagate through the network under varying hydraulic conditions?” , “How do seasonal and diurnal patterns affect recoverable heat potential?”, “How does temperature influence in-sewer biochemical processes?” The UWO dataset provides high-resolution, spatially distributed wastewater and sewer headspace temperature data combined with hydraulic and meteorological observations. Although no dedicated thermal infrastructure measurements are included, e.g. on Blue-Green Infrastructure, the dataset uniquely enables empirical analysis of in-sewer temperature dynamics and supports the development and testing of temperature-based process models at network scale (Abdel-Aal et al., 2019; Duque et al., 2025; Figueroa et al., 2021).

~~Urban drainage systems must be adapted to climate change using smart combinations of grey infrastructure and Blue-Green Infrastructure (BGI), yet the interaction between these systems remains poorly understood. Future research should therefore focus on quantifying the contribution of BGI to overall system performance, especially under extreme weather conditions and in the context of climate resilience. Key research questions include: “How effectively does BGI reduce peak runoff under extreme rainfall?” and “Can BGI also support thermal energy recovery or mitigate urban heat stress?” To address these questions, detailed hydraulic and thermal simulations are needed that account for both conventional and nature-based infrastructure. Combining the available rainfall, flow, and temperature measurements with a model such as SWMM-HEAT — offers a unique opportunity to simulate such future climate scenarios and assess the performance of integrated grey-green solutions.~~

6.2.4~~3~~ Developing predictive models and Real-Time Control strategies

Real-time control (RTC) of urban drainage systems depends on the ability to predict system states reliably under a wide range of rainfall and runoff conditions. Future research ~~could~~ can use the UWO data to develop predictive models that support effective control strategies. Central research questions include: "Which types of rainfall events are most critical for triggering control actions?", ~~"How can we use monitoring data, which are never 100% accurate, to quantify the performance of RTC systems?",~~ "Can observed flow patterns be used to predict system states and guide real-time decisions?", 2"How can we use monitoring data, which are never 100% accurate, to quantify the performance of RTC systems?". A key aspect for future research is not only to evaluate the performance of RTC under faulty or missing data conditions, but also to develop control strategies that remain robust despite imperfect information. The goal should be RTC systems that maintain performance close to that achieved under perfect system knowledge or at least exhibit only minor degradations when data are incomplete, delayed, or erroneous.

One very interesting direction, which is not widely discussed, is "Can we use the SCADA system to identify important characteristics of the full-scale system?" where controlled interventions in the full-scale system can actively test sensor performance and system behaviour. For example, operators can trigger actuators like pumps or gates and directly observe the sensor response to identify delays, outliers, or faults (Sant'Anna et al., 2024).

6.2.~~4~~5 Accounting for uncertainty and improving statistical inference

Reliable model predictions depend on properly accounting for uncertainty especially given the fact that urban drainage models contain structural bias (Del Giudice et al., 2013, 2015). Also, flow and rainfall measurements are also often systematically wrong, e.g. through offset, calibration shifts or drift (Del Giudice et al., 2016). Future research can ~~could~~ therefore use the provided data packages to explore what methods are promising to describe prediction uncertainty in a statistically correct way, e.g. conformal predictions (Vovk et al., 2025) and how autocorrelation, multiple data sources and input errors from rain gauges can be accounted for (Auer et al., 2024; Sun and Yu, 2022).

6.2.~~5~~6 Enabling smart data infrastructures

Despite the increasing volumes of monitoring data, urban drainage datasets remain underused due to insufficient metadata, inconsistent formats, and limited user access. To unlock their full potential, future research should invest in FAIR data infrastructure that enables intuitive, domain-aware interaction with complex environmental data. One promising avenue is the development of interfaces that translate natural language queries into structured database queries (text-to-SQL), allowing users to retrieve relevant data without deep technical knowledge of the schema (Allemang and Sequeda, 2024). Ideally, one could directly ask the database "Which pipe is at capacity?", "What was the annuity of the thunderstorm last Friday?" and "Which city district has the highest infiltration rate?” To support such functionality, data models must be semantically rich and ideally rely on ontologies and standardized knowledge graphs. A starting point could be the Dutch Urban Drainage ontology (GWSW) (The Dutch Urban Drainage Ontology (GWSW), 2025), which defines classes and relationships within urban drainage systems and is already integrated with national tools for asset management and modelling. As it is mostly focused on asset representation rather than on sensor time series, further semantic developments are needed to support such queries.

7 Conclusions

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 provide the unique UWO open dataset which includes urban hydrological observations as well as telemetry data. Based on our experience of collecting and curating the UWO dataset, we draw the following conclusions:

- Urban drainage monitoring is evolving, but despite progress in data acquisition and processing, widespread open data sharing is still the exception rather than the rule.
- IoT-prone transmission technologies and advances in sensor development offer the opportunity to revolutionize sewer process monitoring by enabling real-time, spatially differentiated information. However, they do not replace traditional methods and are not yet a plug-and-play solution for ubiquitous sensing. Power supply of many sensors is still a major bottlenecks and limits scalability, especially for small utilities with few resources.
- In terms of sensors, we conclude that a dual-sensor strategy pays off: using a few high-quality sensors for key measurements (e.g., rainfall, pipe flow) ensures trustworthy reference data, while low-cost sensors, such as ultrasonic water level sensors support dense deployment and gathering spatial information, which then provides a deeper understanding how the urban drainage system functions. It is important that sensor deployment must be supported by adequate rainfall and climate monitoring, especially if spatially resolved runoff patterns are to be interpreted meaningfully. Simply put: More flow sensors generally necessitate more rain gauges.
- To fully benefit from monitoring data, it is crucial to ensure that streaming the sensor data in real time is sufficiently reliable. Collecting field observations without timely validation does not make sense, and regular *manual* checks for consistency and homogeneity remain essential to ensure data accuracy. We found that *automated* quality checks, such as range and gradient tests, are adequate to detect anomalies and become particularly important when handling more than 30 to 40 sensor signals. Differentiating whether anomalies are caused by sensor malfunction or actual system behaviour requires more advanced analysis which still has to be standardized.
- For important urban drainage processes, such as sewer infiltration, overflow behaviour and thermal energy exchange, the suggested applications of the UWO dataset demonstrate that high-resolution, long-term monitoring is essential to

understand the relevant dynamics. In addition, the provided telemetry data offer a real-world benchmark for assessing IoT performance from underground environments, enabling further research on optimizing network design, power use, and data reliability in smart city applications.

- Future research should use the UWO dataset to develop and benchmark ML-based anomaly detection and data recovery methods using the long time series. It also seems very promising to investigate in how far BGI can help us to make our urban drainage systems more resilient and adaptable and especially to test modern methods of uncertainty quantification, such as conformal predictions. Future work should also focus on smart databases by combining semantic web technologies with data-driven modelling, e.g. Large Language Models. Urban drainage ontologies and standardized descriptions of monitoring data will probably improve integration, interoperability and even reasoning. Combined with natural language interfaces, these systems could allow users to query sewer data more intuitively, e.g. “Which pipe is at capacity?”, “What was the annuity of the heavy storm last Friday?”

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. 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 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.

Allemang, D. and Sequeda, J.: Increasing the LLM Accuracy for Question Answering: Ontologies to the Rescue!, https://doi.org/10.48550/arXiv.2405.11706, 20 May 2024.

Auer, A., Gauch, M., Kratzert, F., Nearing, G., Hochreiter, S., and Klotz, D.: A data-centric perspective on the information needed for hydrological uncertainty predictions, Hydrol. Earth Syst. Sci., 28, 4099–4126, https://doi.org/10.5194/hess-28-4099-2024, 2024.

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.

Bertrand-Krajewski, J.-L., Lepot, M., and Invernon, N.: UDMT – Urban drainage metrology toolbox: a software tool facilitating the adoption of better metrology practices, Water Pract. Technol., 20, 1784–1796, https://doi.org/10.2166/wpt.2025.102, 2025.

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/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/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/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/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.

Crowley, G., Tait, S., Panoutsos, G., Speight, V., and Esnaola, I.: Information-theoretic sensor placement for large sewer networks, Water Res., 268, 122718, https://doi.org/10.1016/j.watres.2024.122718, 2025.

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

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., and Rieckermann, J.: Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, Hydrol. Earth Syst. Sci., 17, 4209–4225, https://doi.org/10.5194/hess-17-4209-2013, 2013.

Del Giudice, D., Löwe, R., Madsen, H., Mikkelsen, P. S., and Rieckermann, J.: Comparison of two stochastic techniques for reliable urban runoff prediction by modeling systematic errors, Water Resour. Res., 51, 5004–5022, https://doi.org/10.1002/2014WR016678, 2015.

Del Giudice, D., Albert, C., Rieckermann, J., and Reichert, P.: Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation, Water Resour. Res., 52, 3162–3186, https://doi.org/10.1002/2015WR017871, 2016.

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

Duque, H., Diao, K., Villa, R., Leitao, J. P., Djordjević, S., and Abdel-Aal, M.: Context-aware data driven sensor data analysis: With application to H2S concentration prediction in urban drainage networks, Water Res. X, 28, 100346, https://doi.org/10.1016/j.wroa.2025.100346, 2025.

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.

Eawag-SWW: UWO Open Data, https://uwo-opendata.eawag.ch/, 2025.

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/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.

NYS Combined Sewer Overflows (CSOs): https://kaggle.com/datasets/new-york-state/nys-combined-sewer-overflows-csos, last access: 24 March 2022.

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.

Ramgraber, M.: Data for: Non-Gaussian parameter inference for hydrogeological models using Stein Variational Gradient Descent - ERIC, 2025.

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. and Disch, A.: Challenges and Prospects in Anomaly Detection of Sewer Monitoring Data: Annotating Synthetic Sewer Data with Known Sensor Failures, https://doi.org/10.31224/3520, 5 February 2024.

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.

The Dutch Urban Drainage Ontology (GWSW): https://data.gwsw.nl/, last access: 29 June 2025.

Riveraction: Sewage Map, 2025.

Rodriguez Bennadji, M.: The influence of Green Infrastructure on the resilience of urban drainage systems, 2022.

Rodriguez, M., Fu, G., Butler, D., Yuan, Z., and Cook, L.: The effect of green infrastructure on resilience performance in combined sewer systems under climate change, J. Environ. Manage., 353, 120229, https://doi.org/10.1016/j.jenvman.2024.120229, 2024.

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, https://doi.org/https://doi.org/10.1080/15730620902934801, 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.

Sant'Anna, M., Souza, R. G., Prudente, G., Brentan, B., and Meirelles, G.: Analysis of the impact of leaks and valve maneuvers on pressure transient data in water distribution networks, RBRH, 29, e45, https://doi.org/https://doi.org/10.1590/2318-0331.292420240085, 2024.

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.

Speck-Fehraltorf Airport | SKYbrary Aviation Safety: https://skybrary.aero/airports/lszk, last access: 23 May 2025.

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.

Stream - Portal: https://www.streamwaterdata.co.uk/, last access: 29 June 2025.

Sun, S. H. and Yu, R.: Copula Conformal Prediction for Multi-step Time Series Forecasting, 2022.

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.

Villez, K., Vanrolleghem, P. A., and Corominas, L.: Optimal flow sensor placement on wastewater treatment plants, Water Res., 101, 75–83, https://doi.org/10.1016/j.watres.2016.05.068, 2016.

Vovk, V., Gammerman, A., and Shafer, G.: Algorithmic Learning in a Random World | SpringerLink, 2025.

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), 2022.