Supporting information to

Exploring the relationship between mobile phone data and wastewater flows: evidence from five Swiss catchments

Jörg Rieckermann[1]*, Andy Disch[1], Nicolas Neuenhofer[1], Stephan Baumgartner[1,2], Christoph Ort[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

Contents

A. Case study WWTP Oberengadin, 2018

In a MSc thesis at ETH Zürich, adaptive Kalman Filter models to improve the operation of a WWTP with sequencing batch reactors (SBR) were investigated (Troxler, 2018). The case study investigated the WWTP Oberengadin, which consists of five SBRs that would benefit from a precise inflow prediction, because the Oberengadin catchment is a popular tourist destination in a high Alpine valley region that is challenged with heavily varying pollution loads during touristic events. Troxler, 2018 investigated an inflow prediction tool to improve operation not only during variable dry weather flows, but also summer storm events, and spring snow melt conditions.

To address these challenges, an engineering framework was developed based on consultations with the practice partner Hunziker-Betatech, with a minimum prediction horizon of 6 hours, computational time budget of 10 minutes, and an estimation accuracy of 10 L/s.

The study evaluated different approaches (mechanistic and data-driven) to predict sewer flows and discussed various data sources, such as rain measurements, snow depth, groundwater level, river water level or rain predictions. Interestingly, mobile phone users in a catchment were also explored as a surrogate for the number of people present in the catchment. Two test datasets around Christmas (05.12.2016 - 08.01.2017) and around a FIS ski race (06.02.2017 - 12.02.2017) were related to flow data. These data were an earlier version of the Swiss Mobility Insight data, where the Telecom provider was not disclosing the detailed pre-processing.

For the first dataset (Figure SI 1, left), the mean daily inflow to WWTP Staz almost doubled, which corresponded to a similar increase in the daily number of mobile phone users ($R^2 = 0.95$). However, they mentioned that the Müsstexact interpretation of the mobile phone data was challenging, and further clarification is needed regarding whether the data includes roaming, phones in flight mode, or phones that were turned off. For the second period (Figure SI 1, right), an increase in mobile phone users is also reflected in increasing sewer flows. Due to the temporal mismatch between the daily flow data and the hourly mobile data, this correlation was not investigated in more detail.

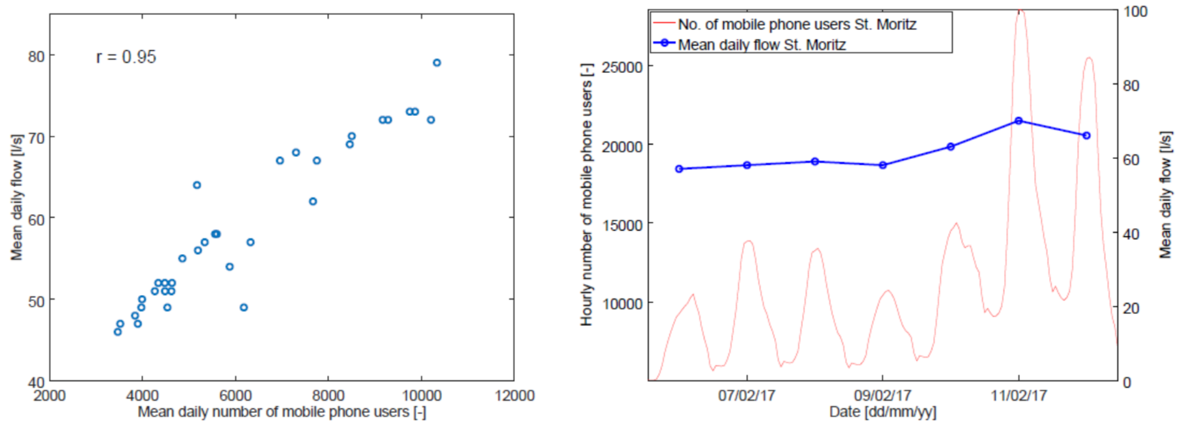

Figure SI 1: (Left) Mean daily number of mobile phone users in the municipality of St. Moritz vs. mean daily inflows at WWTP Staz in between 05.12.2016 - 08.01.2017 (including Christmas and New Year). (Right) The comparison of hourly number of mobile phone users to mean daily inflows between 06.02.2017 - 12.02.2017 shows an increasing inflow with increasing mobile phone users.

B. The potential of mobile phone population data in urban water management

The following section is a more detailed version of the discussion presented in the main body of the manuscript. It provides more references and interpretations, which we consider valuable for some readers.

B.1 Characteristics and processing of the SMI mobility phone data

The Swisscom data product provided valuable insights for our study, particularly through the "*Heatmap*" product. However, the "*Dwell Times"* or "*Origin Destination"* products[1] were found to be less interesting for our purposes. It is worth noting that the SMI data exhibited a characteristic 1 hour dip at the end of the day, which is similar to the dip in wastewater production during the night, when people sleep. As the majority of the phones stays on during the night, and do not log out of the network, this dip is likely due to technical issues and could not be fully explained by the Telecom provider. In our application, we obtained the best results by removing the night hours during data cleaning.

The adequate temporal resolution of both data sources may have an effect and is an open issue that requires further investigation, because the simple correlation analysis performed here cannot fully reflect the complex mapping from population size, and (hidden) wastewater generation in a subcatchment, to the wastewater flows at a monitoring location.

Privacy protection and the resulting cut-off of 20 SIM cards per tile can be a major drawback, especially for small subcatchments. Due to the nature of the "network events" of the mobility data, the use of this data would benefit greatly from custom processing for urban water applications. This is illustrated by the data from the sub catchment of Trogen, where data were always below the cutoff (Figure SI 2), with the exception of the first day of the school year. A probable explanation is that on this morning, unusually many parents and students gathered in a specific school to celebrate with their children, exceeding the threshold for a limited amount of time. Finally, it is important to consider that the topography of Switzerland can affect the accuracy of mobility data. In mountainous areas, mobile devices might connect to antennas that are far away from their actual location. This can occur when the signal from a distant antenna, e.g. located on top of a mountain, is stronger than the signal from a nearby antenna due to shadow fading.

[1] DWELL TIMES: the times people or vehicles spend in a particular geographical area in Switzerland on a particular day – from the previous day to two years ago. ORIGIN DESTINATION: the API provides information about the number of trips between two or more regions in Switzerland by weekday, weekend or calendar month.

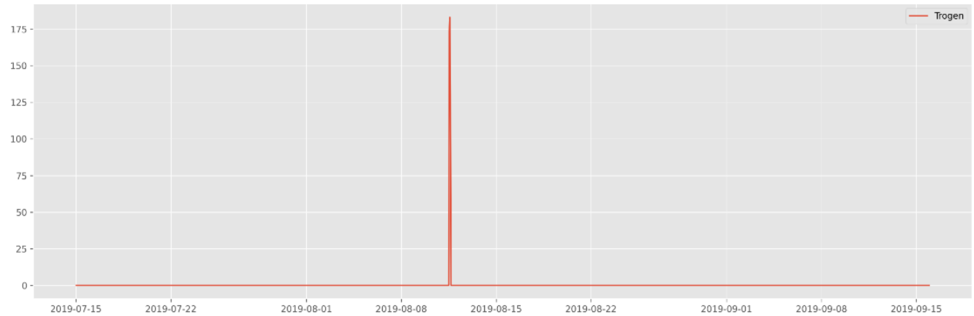

Figure SI 2: we see a PEAK on 11.08., the start of the new school semester. The minimum network usage of 20 people to release the data for privacy reasons is a problem in sparsely populated catchments.

B.2 Considering further alternative data sources to predict wastewater flows

In this feasibility study, the focus was on correlating the mobile phone data to wastewater flows, but the results show that simple linear relationships are not adequate, especially for large catchments. Thus, to enhance the prediction accuracy and gain a more comprehensive understanding of the relationship between population size and wastewater dynamics, other relevant data sources could be included. For instance, traffic data could be used to remove bias in population size, because people in traffic do not generate wastewater. In contrast, using drinking water consumption data is challenging because drinking water metering zones are usually different from the sewersheds. Data on electricity consumption are also tricky, because they also include industrial production, often during nighttime. Therefore, future studies should consider incorporating multiple data sources, addressing the limitations of different datasets to further improve the predictive models.

B.3 Improve the data-driven model, the statistical analysis and the approach by including prior engineering knowledge into the analysis

In this feasibility study, we merely analyzed linear correlations and to ensure accurate results, the analysis would benefit from a more advanced statistical analysis. A next step could be to include the effect of school holidays directly as an explanatory variable in the analysis. Also, the data could be pooled with information on catchment characteristics.

Choosing the appropriate model is also crucial, and a natural step would be to investigate non-linear models, such as random forest regression. In our view, the available data are a limiting factor to use more advanced methods, such as ANNs, because, essentially, we only have information from nine dry weather days.

One important consideration in data analysis is feature selection. Census data and traffic information may provide valuable insights into the relationship between mobile phone data and wastewater production. For example, census data can give an indication of the population density in the individual catchments, which can be informative for wastewater generation. Traffic data can also provide insight into non-wastewater generating traffic, which is to date included in the SMI population data and thus a potentially confounding factor.

Another challenge when analyzing the data is determining the appropriate time resolution for each sub-catchment and monitoring site. For small catchments, an hourly resolution may be too large, which can affect the accuracy of the analysis. However, including weekends and travel time can provide a more comprehensive understanding of the data. Travel time, in particular, is a challenging factor to account for due to its intricate nature. Nonetheless, the overall scatterplot already shows a smoothing effect, which can be used to adjust estimates of wastewater production.

To better estimate flow time, a detailed hydrodynamic model and topography could also be used. A hydrodynamic model can simulate the movement of the generated wastewater through the sewer system. Thus, it can not only help estimate flow times, but even take into account effects from important hydraulic structures, such as pumping stations. Catchment topography, e.g. from digital terrain models, can also provide information on the elevation and gradient of the sewer system, which can affect the flow of wastewater and thus the correlation between the SMI data and flows. However, it may be challenging to come up with an individual travel time for each sub-catchment. In this case, assuming that most people live close to the city center may be necessary.

C. The potential of mobile device data in urban water management applications

C.1 Urban drainage

Mobile data has numerous applications in urban drainage applications, including water pollution control, distributed wastewater pollution estimation, groundwater infiltration estimation, stormwater pollution estimation, pollution-based real-time control as well as quality control of monitoring data.

Water pollution control: As discussed here, one of the primary applications of mobile data in urban drainage applications would be estimating sewage flow and water pollution control. By analyzing mobile data, it is possible to estimate the amount of wastewater produced in different areas of the city, which can help identify areas with high pollution flows and develop effective pollution control strategies.

Distributed wastewater pollution: Mobile phone data can also be used to estimate distributed wastewater pollution across different spatio-temporal scales, which is not possible with traditional sensor-based methods. Specifically, this would provide more accurate data than the information which has been used by Pistocchi and Dorati (2018) to assess population densities and corresponding emissions from urban drainage systems on a European level. By analyzing mobile data, it is possible to determine the sources of pollution and develop targeted pollution control strategies.

Groundwater infiltration: The estimation of groundwater infiltration and rain-induced infiltration and inflow is another important application of more accurate data for wastewater production. If one could establish a stable relationship between mobile phone data and dry weather flow, which also holds during wet weather, one could better decompose the flow data into the wet weather and dry weather components. As this could be possible across different spatial scales, it would greatly help to identify areas where groundwater is entering the sewer system and develop strategies to reduce infiltration.

Stormwater pollution: Mobile phone data can also be used to estimate stormwater pollution, particularly from cars. Through a more detailed analysis of mobility patterns, specifically traffic patterns, in different areas of the city, it is possible to estimate the amount of pollution generated by tire wear and other sources of pollution. This information can help identify areas with high levels of stormwater pollution and develop targeted pollution control strategies.

Real-time control strategies: Prediction and optimization of pollution-based real-time control is another important application of mobile data in urban drainage applications. By analyzing mobile data, it is possible to predict the level of pollution in different areas of the city and develop real-time control strategies to minimize pollution levels. Rainfall data can also be integrated with mobile data to determine cut-off levels for pollution control measures.

C.2 Wastewater treatment

Population estimates: In the context of WWTPs, mobile data can also have several applications. For example, population estimates derived from mobile phone data can be used to design and optimize the performance of WWTPs. By understanding the fluctuations in population density and wastewater production, WWTPs can be designed to handle the expected flow rates and loads, which can ultimately improve the efficiency of treatment processes. This could also be relevant for billing purposes to share the cost of a joint WWTP among a group of municipalities.

Dynamic influent data generation: As many WWTPs are missing influent pollution data with a high temporal resolution, engineers use influent generators to create long time series for dynamic simulations and design studies (Bisinella de Faria et al., 2020). Mobile phone data can help to better parameterize these tools than currently possible.

Surveying important wastewater producers in real-time: Another important application of mobile data in WWTPs is the monitoring of individual wastewater producers. By drawing a polygon around specific sites, such as ski resorts or an industrial site, it is possible to estimate the volume and composition of non-residential wastewater being discharged into the sewer system. This information can be used to improve the operation of WWTPs, to trigger monitoring campaigns, to identify potential sources of pollution and to develop strategies for mitigating their impact on the environment.

C.3 Drinking water

Drinking water applications of mobile data have been gaining traction in recent years. One important use case is the detection of drinking water leaks in distribution systems. Mobile data can be used to monitor the pressure and flow rates within the system and detect any irregularities that may indicate a leak. By detecting leaks early, water loss and associated costs can be minimized.

Real-time control of drinking water systems is another application of mobile data. With the help of sensors and mobile data, water utilities can monitor the water quality in real-time and control the treatment processes accordingly. For example, if there is a sudden increase in turbidity, the system can automatically adjust the coagulant dosing to maintain the required water quality standards.

Mobile data can also be used for water demand forecasting, allowing utilities to optimize their water supply and treatment processes. By analyzing the historical data on water consumption patterns, utilities can forecast the demand for water in different locations and at different times of the day, and ensure that the water supply meets the demand.

In addition to municipal water supply, mobile data can also be used to monitor and optimize industrial water use. By analyzing mobile phone data from employees or visitors to industrial sites, it is possible to estimate the water usage patterns of individual companies or facilities. This information can be used to develop strategies for reducing water consumption and improving the overall sustainability of industrial operations.

C.4 Wastewater-based epidemiology

Since the onset of sanitary engineering, wastewater-based epidemiology (WBE) has proved very useful in the early detection of diseases and the tracking of the prevalence and spread of infectious diseases within a community or region (Wilson, 1928). By analyzing wastewater samples over time, researchers can identify trends in disease prevalence, and monitor the effectiveness of public health interventions. As it is cost-effective and provides data in near real-time, WBE is gaining popularity, especially after the COVID-19 pandemic, and has been suggested as a new standard in the proposal of the revised UWWTD (EC, 2022).

To compare WBE results accurately across time, different locations, and with other epidemiological data, it is crucial to normalize WBE results to the actual population size. Today, WBE data is commonly normalized with de jure (residential) population numbers. This is not optimal since catchment populations are known to fluctuate (Thomas et al., 2017). The diurnal (e.g. commuters) and intra-annual (e.g. holidays) population fluctuations within WWTP catchment boundaries are often considered to follow a regular pattern (except for special events, e.g. festivals, pandemic measures). This allows to compare 24-h loads of respective days over time and to interpret variations as changes in consumption/excretion. The evaluation of the validity of this assumption is often difficult and it generally limits the use of the WBE approach to the comparison of “similar” days. The pandemic measures, such as local or regional lockdowns, or working from home, changed the population mobility behavior. This fundamentally challenges normalization with de jure data and emphasizes the need for novel approaches to normalize the de facto population, e.g. with human biomarkers (O’Brien et al., 2023) or mobile phone data. In addition, mobile phone data could also be accompanied by relevant socio-demographic information, such as proportion of males and females or the age distribution of the population.

C.5 Potential of wastewater monitoring data to improve mobile phone data

As cell phone providers do not have perfect market penetration, it seems reasonable to assume that wastewater flow measurements can also help to remove bias in population size estimates derived from mobile phone data. This can be permanent adjustments, e.g. offset corrections in catchments with an insufficient number of clients, or temporarily, e.g. in times of network outages.

Based on our experience from this case study, wastewater flow data can also improve mobile phone data, because the latter can be affected by bias from phone ownership, market penetration, cellular network coverage, different usage patterns in urban or rural areas, and connectivity, which can introduce biases into estimated population sizes. By comparing these estimates with data on the dynamics of foul sewage, which reflect the actual number of people in an area, telecom providers can better understand the uncertainties and limitations of their mobile phone data and derived products. Consequently, they can adjust their internal data pipelines accordingly.

Additionally, wastewater flows can provide insights into the temporal dynamics of population movement, such as changes in population size during holidays or special events, which can be difficult to capture with mobile phone data alone. Such reference data might be especially valuable in catchments which are small or have complex topographies, where cut-off effect and privacy issues might compromise the usefulness of the data. In such cases, falling back to wastewater flow derived estimates might be a possibility to improve the reliability of the data.