EGU23-7355, updated on 10 Feb 2024
https://doi.org/10.5194/egusphere-egu23-7355
EGU General Assembly 2023

From Correlation to Causation: Discovering the Drivers of Urban Water Demands in the Contiguous United States

Wenjin Hao[1], Andrea Cominola[2,3], and Andrea Castelletti[1]
[1]Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy
[2]Chair of Smart Water Networks, Technische Universität Berlin, Berlin, germany
[3]Einstein Center Digital Future, Berlin, Germany

Urban water demands vary across spatio-temporal scales, driven by multiple socio-demographic, climatic, and urban form factors. Identifying influential drivers, along with their individual and compound effects on urban water consumption, is essential to forecasting future water demand, addressing urban water security, and informing water governance. Model-free and model-based Input Variable Selection (IVS) has been extensively applied to investigate important predictors of urban water demands. However, most IVS methods identify correlations and mutual information between variables, which do not imply causation. More recently, causal discovery has developed as an active area of research in many research fields, including, e.g., neuroscience, finance, and climate science. Causal discovery improves IVS by identifying causally meaningful relationships between variables, distinguishing indirect from direct dependencies, and recognising relevant drivers among multiple variables.

In this work, we investigate predictive and causal factors of urban water use across the Contiguous United States (CONUS). We rely on open data of monthly municipal water consumption from 126 cities in the US for the period 2010-2017 and data on candidate socio-demographic, climatic, and built environment predictors from multiple sources, including the U.S. Census Bureau, the American Community Survey, and the PRISM climate data set. We first test the state-of-the-art W-QEISS wrapper method to identify equally-informative subsets of predictive factors for urban water demands. These subsets are the solutions of a four-objectives optimisation problem that maximises the predictive accuracy of a data-driven model and feature relevance while minimising the number of selected predictors and their redundancy. Results show that historical water consumption is the most relevant factor to predict future demands, followed by some socio-demographics, climatic factors, and building characteristics, including the median number of rooms in housing units, unemployment rate, Palmer Drought Severity Index (PDSI), and building construction years. Preliminary results for individual climate regions also highlight local effects, with PDSI becoming more relevant for arid regions than the continental-scale results. Second, we extend our analysis to causal discovery by applying a neural Granger model to interpret non-linear Granger causality and temporal structures within time series. Granger causality describes whether past values of a time series x_t could predict future values of another series $y_{t,}$ assuming causal effects are ordered in time (i.e., cause before effect). This allows for finding the specific causes of

urban water demands in our case (in a Granger’s sense). We finally compare the causality results with the results of IVS to illustrate the different interpretations of urban water demand drivers.