The analysis of influential factors for urban water supply system considering causality with time-series data

J W Ryu¹ and H K Park¹,²

¹Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, South Korea
²Email: hpark57@kaist.ac.kr

Abstract. Recently, the research on critical infrastructures have been extended its boundary and some previous research analysed interdependency between infrastructures tried to suggest operation and management policy of the infrastructure. This highlighted the importance of systemic approach such as System Dynamics. But there still exist the limit to select the modelling component in terms of causality. In this study, convergent cross mapping was applied to time series variables in Songpa-gu and Gangdong-gu in Seoul, South Korea. Selected variables include operating variable of water supply system, floating population and meteorological variables and 330 observations were collected. The result shows that daily average temperature and floating population that are equivalent to mobile population influences to the daily water supply, and vice versa. Other weather factors such as average precipitation and wind speed showed little causality with the daily water supply. The influential factors of water supply system can be investigated with convergent cross mapping and the result can be utilized for the pre-process of other methodology such as LSTM in short-term water demand prediction. It is expected to collect more available data to improve this study further.

1. Introduction

The urban system includes many infrastructures and becomes more complex. The infrastructures in urban system have been interdependent to each other. Understanding the interdependency between different infrastructures is essential to mitigate damage the cascading damage [1] and improve the operation and management policy. Along with the significantly increasing interdependency, it is essential to cross the boundary of each infrastructure and to perform integrated research.

Especially, urban water system has been analyzed with many other urban sectors such as energy, transportation, communication, socioeconomic sector. System Dynamics approach is one methodology to model the complex system and to suggest the management policy [2]. Sometimes the relationships between components in those models are based on regression [3].

Because the System Dynamics model is time dependent approach, the relationship between temporal components should be also considered. The Granger Causality test [4] and Convergent Cross Mapping [5] has been one of the most widely utilized approaches to analyze the causality between time series variables. These kind of approaches introduced to figure out the causality are important to distinguish the correlation and causality. It has been an issue that the causality of some hydrological and climate variables in time series are analyzed by the causality test such as Granger Causality test and Convergent Cross Mapping [6].
The causality test for urban water supply system can provide grounds to select influential factors and input variables for the modeling of urban water system. In this research, the Convergent Cross Mapping is introduced to analyze the causality between time series data in urban water supply system, population, and weather factors that are widely used for non-temporal analysis.

2. Research background

2.1. Convergent Cross Mapping

The convergent cross mapping is an algorithm developed for empirical dynamic modeling to figure out the causality between empirical time series data.

Time-series variables can be considered as the projection of a reconstruction along with dimensional axis. With given time-series variable \( X \) and \( Y \) which are projected from certain reconstruction \( M \), reconstructions of each variables, \( M_X \) and \( M_Y \), can be generated with time lags. This reconstruction is called shadow manifold. Then, nearby points in each shadow manifold connote the historical behavior of \( X \) and \( Y \). It is obvious that \( M_X \) and \( M_Y \) share \( M \) by one-to-one mapping. This implies that \( M_X \) and \( M_Y \) also map each other one-to-one.

When single time series variable is reconstructed like \( M_X \) and \( M_Y \), the embedding dimension of this multi-dimensional space is equivalent to the number of time lags that are used to construct the shadow manifold. The dimensionality that best describes this time series data best represents the information inherent in this data. As shown in figure 1, if the time series variable is reconstructed in a space with insufficient dimension, the singularity of data points is not guaranteed. It means that exist overlapped points when the reconstruction is projected. If the dimensionality of reconstruction is too high, the distances between points are exaggerated because of the Curse of Dimension. Therefore, it is crucial to find the optimal embedding dimension for convergent cross mapping.

![Figure 1. Behavior of data points along with different embedding dimension.](image)

With two shadow manifolds that map each other one-to-one, nearby point of a point on a shadow manifold at time \( t \) can be used to estimate a point on another shadow manifold. As shown in equation (1) to (3), the weight to estimate \( Y \) can be calculated by the Euclidean distance between points and estimate of \( Y \) is calculated based on the weight. Finally, the correlation between estimate of \( Y \) using the weights(=\( Y|M_X \)) and \( Y \) indicates the causality between two variables. It is important to note that the direction of causality is opposite to the direction of mapping.

\[
\hat{Y}(t) | M_X = \sum w_i Y(t_i) \quad (1)
\]

\[
w_i = u_i / \sum u_j \quad (2)
\]

\[
u_i = \exp\{-d[x(t),x(t_i)] / d[x(t),x(t_i)]^3\} \quad (3)
\]

3. Methodology

3.1. Data collection and handling process
The Songpa-gu and Gangdong-gu in Seoul, South Korea were selected for the research area. Daily water supply, daily floating population, and weather factors were collected. The Gangdong Water Supply Office services and control the water supply in research area. The floating population in research area was assumed equivalent to the mobile population because of the significant smartphone penetration rate, which was 95% in 2018. The real-time number of smartphones was counted hourly with the radio signal. The mobile population from 5th April 2018 to February 28th 2019 was collected from Seoul Open Data Plaza. Weather factors, daily average temperature, daily precipitation and daily average wind speed, were provided from the Korea Meteorological Administration. Because the mobile population was the shortest, all other data was collected for the same period of mobile population. The final data include 330 observations and each variable was scaled to assume that they are stationary time series variables.

3.2. Application of convergent cross mapping

The empirical study was carried with open-source programming language R and package ‘rEDM’ developed by Sugihara laboratory. As described in section 2, the optimal embedding dimension was investigated before applying convergent cross mapping algorithm between time series variables. 231 observations were sampled as the library for prediction and 99 observations were sampled for the target of prediction. Each variable compared the predictability with seven time lags and the embedding dimension with the best predictability was selected for the dimension for convergent cross mapping. Then, the convergent cross mapping was applied between the water supply and rest other variables, which the correlation for each variable and estimate from coupled variable was calculated. Fine tuning for the hyperparameters for convergent cross mapping was performed by trial and error.

4. Result and discussion

Table 1 shows the optimal embedding dimension of each variable. The meaning of some optimal embedding dimensions corresponded to behavior of urban population. The optimal embedding dimension of water supply reflects the weekly repeating pattern of water usage. The water supply during weekdays and weekends are different so that the water supply shows a clear weekly pattern. The low embedding dimension of precipitation results from the characteristics of data itself. It is not necessary to look back much because most of observations in precipitation have zero value.

| Water supply | Average temperature | Mobile population | Precipitation | Average wind speed |
|--------------|---------------------|-------------------|---------------|-------------------|
| E 7          | 3                   | 4                 | 2             | 1                 |

From figures 2 and 3, it is clear that water supply has bidirectional causal relationship with average temperature and mobile population, respectively. Especially, daily average temperature showed the most significant causal relationship to water supply in the view of convergent cross mapping algorithm. The direction of causality indicates that average temperature results in water supply more that water results in average temperature. The range of correlation (\(=p\)) is very moderate so that their exist many unknown causal components between them.

Water supply and mobile population also has causal relationship but the direction is different to the case of average temperature. It is figured out that the floating population in daytime is not a significantly influential factor to water supply. The mobile population should be split to residential population in research area in future study.

Figures 4 and 5 shows little causal relationship between water supply and rest weather factors. In the past researches that are not time series analysis, the water supply was sometimes regressed or modeled with weather factors such as precipitation, humidity and wind speed. With the result of convergent cross mapping, it should be carefully considered to take input from those weather factors to model water supply.
Figure 2. Convergent cross mapping between water supply and average temperature.

Figure 3. Convergent cross mapping between water supply and mobile population.
5. Conclusion

Water supply has been one of major urban infrastructure. The improvement of operation and management policy has been significantly important issue so that related researches such as short-term water demand prediction has been performed. In this study, the convergent cross mapping test was...
applied to time series water supply data, floating population assumed to be equivalent to the mobile population and weather factors. It was concluded that average temperature and mobile population are in causally influential to the water supply.

Daily average temperature and daily floating population were identified as the influential factors of water supply by the convergent cross mapping algorithm. It is important not to conclude them as a direct cause and effect. The precipitation and average wind speed showed little causality to water supply. This result can be utilized to improve the reliability of time series modeling such as ARIMA and LSTM in future study.

The energy consumption such as natural gas and electricity in research area was also collected but those data were gathered monthly. The generation of daily data from monthly data should be performed very strictly unless it highly increases the uncertainty, so that it was considered as out of the scope of this study. The demonstration of that availability in future study will provide a chance to compare relationship between many data from different urban infrastructures.

References
[1] Little R G 2003 Toward more robust infrastructure: observations on improving the resilience and reliability of critical systems Proc. 36th Annual Hawaii Int. Conf. on System Science (Big Island, HI, USA) p 9
[2] House–Peters L A and Chang H 2011 Urban water demand modeling: Review of concepts, methods, and organizing principles Water Resour. Res. 47 15
[3] Qi C and Chang N B 2011 System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts J. Environ. Manage. 92 1628-41
[4] Granger C W 1969 Investigating causal relations by econometric models and cross-spectral methods Econometrica 37 424-38
[5] Sugihara G, May R, Ye H, Hsieh C H, Deyle E, Fogarty M and Munch S 2012 Detecting causality in complex ecosystems Science 338 496-500
[6] Huang X, Maçaira P M, Hassani H, Oliveira F L C, and Dhesi G 2019 Hydrological natural inflow and climate variables: Time and frequency causality analysis Physica A: Statistical Mechanics and its Applications 516 480-95