Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 683-690
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

Analysis of the Main Causes of Water Leakage in Universities Based on Big Data Analysis

Limin Wang[1], Xuexin Wang[2], Shuoshuo Li[2], Quancai Deng[1], Yachao Si[2], Yongqiang Wu[3*]

[1] Department of Mathematics, Hebei University of Architecture, Zhangjiakou, Hebei, China
[2] College of Information Engineering, Hebei University of Architecture, Zhangjiakou, Hebei, China
[3] Hebei Key Laboratory of Water Quality Engineering and Comprehensive Utilization of Water Resources, Zhangjiakou, Hebei, China
*Corresponding Author.

Abstract:

An important step in controlling the leakage of reclaimed water supply systems is finding the main cause of reclaimed water leakage through data analysis in the water supply system. This can not only save water resources, but also provide a theoretical basis for the treatment of water leakage. Three parts of the work have been done in this paper. Firstly, the reclaimed water supply data at a university was collected and averaged by hours. Then, principal component analysis was used to reduce the dimension of the 24-hour reclaimed water supply data. Finally, the dimension-reduced datasets were clustered, which classified them into four categories. By analyzing the graphs of the four types of data, it was found that the period when water supply differences were relatively large was mainly concentrated at 0–5 o'clock in the evening. The main cause of water leakage was found after an investigation with logistics managers, which was the breakdown of the toilet in the student dormitory at night. This study revealed the main cause of reclaimed water leakage in a university. With the joint efforts of reclaimed water management personnel and logistics managers, the leakage of the reclaimed water pipe network has been greatly improved. This study has a certain reference significance for revealing the actual situation of water supply by data analysis, looking for reclaimed water reuse, and finding the cause of water leakage in urban water supply pipe network systems.

Keywords*: Cluster analysis, Leaking reason, Principal component analysis, Reclaimed water.*

I. INTRODUCTION

The water supply system is the lifeline of the city. It's the indispensable material foundation for guaranteeing life quality and developing production and construction [1]. It has become an important factor in Chinese urban sustainable development and construction. And it's directly related to social stability and economic development. Our country has abundant water resources, but its per capita available water is that of a water-stressed country. The problem of water scarcity in cities is becoming increasingly prominent as the social economy develops, urbanization levels improve, the supply-demand gap for water resources worsens, and the cost of water production continues to rise [2].

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 683-690
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

According to the Statistical Yearbook of China's Urban and Rural Construction [3], the total annual water supply leakage in China reached 7.855 billion cubic meters in 2017, and the average difference between production and sales in cities and towns was as high as 17.25%, which meant that for every 100 tons of water produced, 17.25 tons of water was lost, so water resources were seriously wasted [4]. Therefore, many measures should be taken to control the water supply network leakage, which can not only relieve the huge pressure brought by water resource shortage, but also improve the economic benefits of water supply enterprises, guarantee people's living standards, and play a positive role in the healthy development of the city [4, 5].

Studies have shown that water leakage in the water supply system is mainly reflected in water pipeline leakage, which can be divided into valve leakage, pipe fittings leakage, pipe interface leakage, fire hydrant leakage, pipe body leakage, etc.

The water that between tap water and sewage water is reclaimed water, also known as recycled water, which can be applied in industry and daily life [3]. Reclaimed water reuse systems are gradually being promoted due to the characteristics of high water consumption, low pollution degree of wastewater, and single component of water use in colleges and universities, with reclaimed water serving as a link for miscellaneous water reuse on campus [4].

The operation process of the reclaimed water supply system in a university is to collect the sewage generated in each area, such as teaching buildings, dorm buildings, and libraries. And then discharge it into the sewage treatment system of the campus water station to finally get the secondary water after purification. The secondary water is supplied to dormitory buildings, teaching buildings, and libraries for flushing and campus greening irrigation, so as to achieve the purpose of water conservation and secondary utilization of resources. Fig 1 shows the water reuse route on a campus.

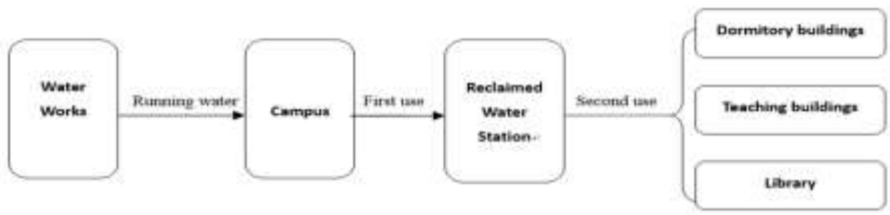


Fig 1: Schematic diagram of reclaimed water reuse on campus

II. DATA ACQUISITION

For the time being, the school has not measured the water flow, so it cannot directly obtain the data of the reclaimed water supply flow. This study used the power consumption of the variable frequency pump to simulate the water supply flow. In this study, there are four pumps to realize a constant pressure water supply (specific information is shown in TABLE I). In the actual constant pressure water supply system, the power consumption of the variable frequency pump is proportional to the first power of its flow [6].

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 683-690
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

The power consumption of the variable frequency pump was obtained from the ABB ACS510 fan water pump inverter, which provided power for the variable frequency pump. One computer was used to collect the electric current, frequency, power, and torque of the inverter through the Modbus protocol.

TABLE I. Performance parameters of variable frequency pump

TYPE	QUANTITY	RATE OF FLOW	WATER PUMP LIFTS	REV	MAX-REV	FREQUENCY	POWER
A	3 sets	$63m^3/h$	63m	2900rpm	2930rpm	50Hz	15KW
B	1 sets	$8m^3/h$	63m	2900rpm	2930rpm	50Hz	2.2KW

m^3/h = Cubic metres per hour, m = meter, rpm = round per minute, Hz = hertz, KW = kilowatt.

The water supply data was sourced from the water pump power consumption, which was obtained in this way. The data was collected from March 6, 2019 to September 9, 2019. And 827504 pieces of actual data were collected. After deducting holidays, failure of collection equipment, incomplete collection data, and other reasons, the complete collection time was 61 days, as shown in TABLE II.

TABLE II. Data collection time

Segment	Start time	End time	Days
1	2019-3-6	2019-3-9	4
2	2019-4-18	2019-5-24	36
3	2019-6-27	2019-7-3	7
4	2019-7-9	2019-7-10	2
5	2019-8-29	2019-9-9	12

III. DATA ANALYSIS

Because of the large amount of data, the time-power diagram was obtained by averaging all the data by hour, as shown in Figure 2.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 683-690
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

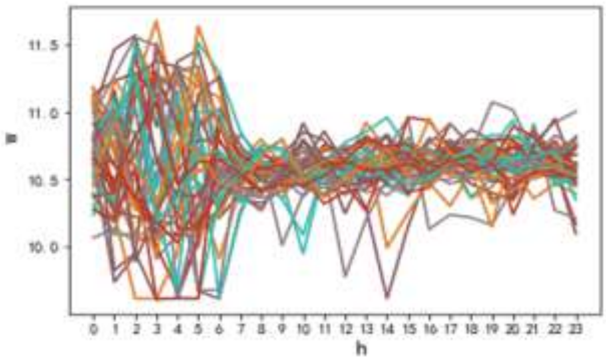


Fig 2: Time-power diagram. The horizontal axis represents time, and the symbolic unit is h =hour. The vertical axis represents power, and the symbolic unit is w = watt.

Based on practical experience, the reclaimed water is mainly used for green irrigation and flushing, so the graph data should be low in the evening and high in the specific period of the day (7:00-8:00 in the morning, 12:00-13:00 in the afternoon). However, the graph obtained from the collected data does not reflect the envisaged regularity.

3.1 Principal Component Analysis

The reclaimed water supply data was redundant and overlapped to some extent, so the complexity of the problem would be affected in the process of solving the problem. The amount of computation would also increase, and more resources would be consumed. Therefore, redundant information must be removed.

Principal component analysis (PCA) is one of the important statistical methods to transform multi-variable problems into fewer comprehensive variables. And it transforms the problem of high-dimensional pace into low-dimensional space, so it makes the problems relatively simple and intuitive. While reducing the dimension of a multi-variable data system, it can also simplify the statistical characteristics of the variable system. When optimal simplification of multivariable data systems is achieved, a lot of important system information can also be provided, such as the average level of data points, the direction of the greatest changes in data, etc. [7].

In this study, the principal component analysis method was adopted to analyze the reclaimed water supply data, and the 24-hour data was simplified and four principal components were obtained. The vector contribution rates are, respectively, [0.33326918 0.19118237 0.14841815 0.06752895], and the cumulative contribution rate is 74%.

The four principal components are, respectively (where XI represents the average power between time I and time I+1):

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 683-690
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

$$Y1=0.096x_{1}+0.30x_{2}+0.41x_{3}+0.46x_{4}+0.56x_{5}+0.44x_{6}+0.047x_{7}+0.0008x_{8}+0.037x_{9}+0.015x_{10}-0.018x_{11}-0.028x_{12}-0.0024x_{13}-0.0006x_{14}-0.068x_{15}-0.063x_{16}+0.00_{19}x17+0.015x_{18}-0.017x_{19}+0.018x_{20}+0.016x_{21}+0.016x_{22}+0.022x_{23}+0.036x_{24}$$

$$Y2=-0.13x_{1}-0.24x_{2}+0.3x_{3}+0.57x_{4}-0.047x_{5}-0.59x_{6}-0.38x_{7}-0.053x_{8}-0.027x_{9}-0.046x_{10}+0.029x_{11}+0.01x_{12}-0.018x_{13}-0.037x_{14}+0.01x_{15}+0.036x_{16}+0.032x_{17}+0.046x_{18}+0.019x_{19}+0.073x_{20}-0.029x_{21}+0.0017x_{22}+0.0056x_{23}-0.0054x_{24}$$

$$Y3=0.25x_{1}+0.39x_{2}+0.52x_{3}-0.012x_{4}-0.59x_{5}-0.1x_{6}+0.32x_{7}+0.026x_{8}-0.025x_{9}-0.033x_{10}-0.09x_{11}-0.021x_{12}-0.013x_{13}-0.031x_{14}+0.0051x_{15}0.027x_{16}+0.025x_{17}+0.052x_{18}+0.038x_{19}+0.046x_{20}+0.014x_{21}+0.0054x_{22}+0.11x_{23}+0.14x_{24}$$

$$Y4=-0.53x_{1}-0.46x_{2}+0.51x_{3}-0.19x_{4}-0.13x_{5}+0.33x_{6}+0.11x_{7}+0.015x_{8}-0.0049x_{9}-0.047x_{10}-0.063x_{11}+0.016x_{12}-0.024x_{13}+0.0013x_{14}-0.066x_{15}+0.061x_{16}-0.024x_{17}-0.12x_{18+}0.022x_{19}+0.072x_{20}-0.04x_{21}-0.044x_{22}-0.16x_{23}-0.12x_{24}$$

Due to the excessive coefficients of four principal components, the researcher visualized the principal component coefficients in order to get the main coefficients of the four principal components more clearly. As can be seen from the principal component factor graph, the high loads in the four principal components are distributed between X1 and X5, as shown in Figure 3, and 0-5 power data change is the main ingredient in the overall power data changes.

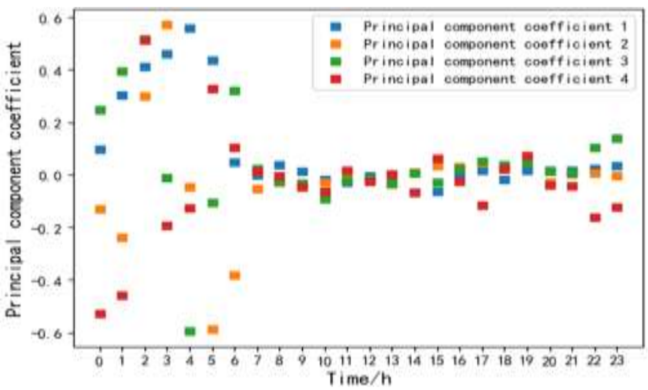


Fig 3: A diagram of principal component coefficient

Based on practical experience, the reclaimed water is mainly used for green irrigation and flushing, so the graph data should be low in the evening and high in the specific period of the day (7:00-8:00 in the morning, 12:00-13:00 in the afternoon). However, the graph obtained from the collected data does not reflect the envisaged regularity.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 683-690
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

3.2 Cluster Analysis

Cluster analysis is a method to classify data into several clusters according to data similarity or dissimilarity. As a result, the similarity degree of data or individuals within the same group is large, while the similarity degree between groups is small [8]. The commonly used clustering algorithms include five clustering algorithms based on hierarchical, density, grids, and models. The K-means clustering algorithm belongs to the partition clustering algorithm, which is the most commonly used algorithm in clustering analysis. It adopts the distance between samples as a clustering index to carry out relevant clustering analysis activities. As a typical unsupervised clustering method, it does not need to be marked by manual experts, so it reduces the clustering cost to a large extent.

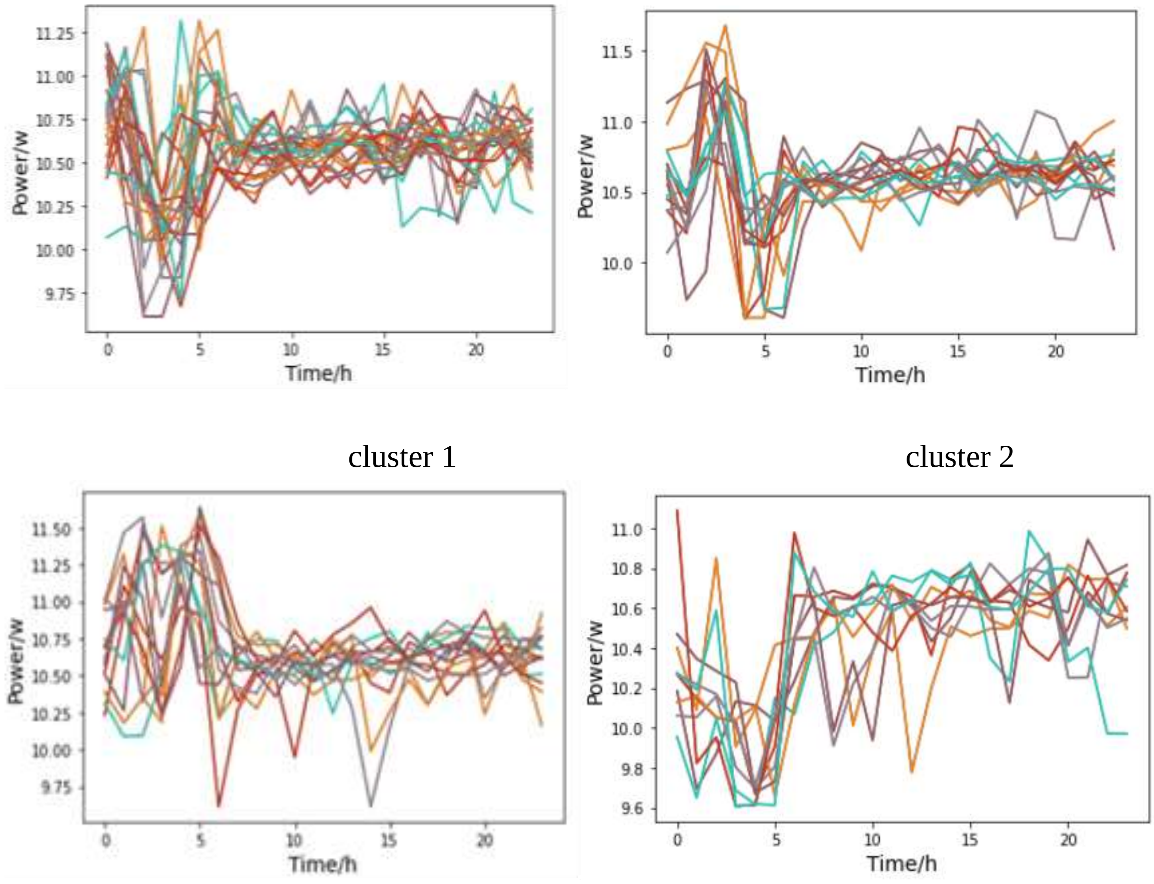
In this study, the k-means clustering method was used. K-means clustering is characterized by a small amount of computation, less memory occupied and fast processing speed, which makes it suitable for clustering analysis with large amounts of data and many variables. The k value of this study was set at 4. The principal component value transfer method was used to directly return the data after dimensionality reduction and that data was clustered. As shown in TABLE III, four types of results were obtained.

TABLE III. Clustering results

Clusters	Days	Dates
1	20	2019-3-8,2019-3-9,2019-4-27,2019-5-2,2019-5-7, 2019-5-15, 2019-5-17, 2019-5-18, 2019-6-28, 2019-6-29, 2019-7-1, 2019-7-2, 2019-7-9, 2019-8-29, 2019-8-31, 2019-9-1, 2019-9-3, 2019-9-5, 2019-9-7, 2019-9-8
2	13	2019-3-7, 2019-4-18, 2019-4-19, 2019-4-22, 2019-4-24, 2019-4-25, 2019-5-4, 2019-5-6, 2019-5-20, 2019-5-21, 2019-5-23, 2019-6-27, 2019-8-30
3	17	2019-3-6, 2019-4-20, 2019-4-29, 2019-5-1, 2019-5-12, 2019-5-13, 2019-5-14, 2019-5-16, 2019-5-19, 2019-5-22, 2019-5-24, 2019-6-30, 2019-7-3, 2019-9-2, 2019-9-4, 2019-9-6,2019-9-9
4	11	2019-4-23, 2019-4-26, 2019-4-28, 2019-4-30, 2019-5-3, 2019-5-5, 2019-5-8,2019-5-9, 2019-5-10, 2019-5-11, 2019-7-10

The graphs after clustering are shown in Figure 4. As can be seen from the figures, between 0 o'clock and 5 o 'clock, the regional power of cluster 1 and cluster 4 was decreased, while the regional power of cluster 2 and cluster 3 was increased. It is obvious that cluster 2 and cluster 3 do not accord with the actual experience of using reclaimed water, so it can be inferred that there are abnormal outlet points where water leakage occurs.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 683-690
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022



cluster 1 cluster 2

cluster 3 cluster 4

Fig 4: The graphs after clustering

After communicating with the water management personnel, it was found that the cause of cluster 2 and cluster 3 was the common running water caused by foreign body adhesion on the sealing flap cover of the flush toilet in the student dormitory at night. From the data analysis, it can be found that the days of abnormal power increase are up to 30, which indicates that this kind of situation occurs frequently, and the highest power is about 11.5, compared with the average power of 10.50, so the leakage is serious, which is the main cause of reclaimed water leakage.

IV. CONCLUSIONS

In this study, the reclaimed water supply data at a university was collected and analyzed. The water supply data was found abnormal. And further research found the main reason for reclaimed water supply network leakage was the constant water flow caused by the failure of the toilet in the dormitory at night. After the overhauled of the water supply pipe network by the reclaimed water management personnel, the leakage of the reclaimed water pipe network in the school has been greatly improved. However, there are some problems in this study, such as the experimental data is only limited to one university and the analysis of the causes of water supply network leakage is not comprehensive enough. Therefore, in the follow-up research, we should collect the reclaimed water supply data of many colleges and universities, and use more innovative methods to analyze the causes of water supply network leakage more comprehensively.

ACKNOWLEDGEMENTS

This work was supported in part by the Basic Scientific Research Funds Research Project of Hebei Province (No. 2021QNJS11), and in part by the Visual Case Library Construction Based of Algorithm Design Course of Hebei Province (No. KCJSZ2022099), and in part by the Research and Implementation of Network Teaching Platform under the Situation of Micro Class & MOOC of Hebei Province (No. 2017GJJG173)

REFERENCES

[1] Yingzi Lin, Jinliang Zhou (2014) Control and research status of leakage phenomenon of urban water supply network. Journal of Jilin institute of civil engineering and architecture 31:35-37+70.
[2] Xiaoting Deng. Analysis and control of leakage and Loss factors of urban water supply network. p. 92. Dissertation. Master's degree in Municipal Engineering. Taiyuan Technology University, Taiyuan, China, 2012.
[3] Ministry of Housing and Urban-Rural Development, PRC. (2017) Annals of China Urban and rural Construction Statistical. Available at: https://www.mohurd.gov.cn/gongkai/fdzdgknr/sjfb/tjxx/jstjnj/index.html
[4] Xinzhe Zhang. Research on leakage loss evaluation index of water supply network based on Modeling. p. 82. Dissertation. Master's degree in Architecture and Civil Engineering. Beijing Civil Engineering and Architecture University, Beijing, China, 2020.
[5] Wenze Li, Xinlu Lu (2019) The superiority and application value of reclaimed water. Regional Management 50:170-172.
[6] Qiongyao Ruan, Baohui Men (2019) Research on the Status quo of Reclaimed Water Reuse in Universities in China. Water Resources Development and Management 12:39-43.
[7] Xi Jin (2011) Linearization solution method for probabilistic hydraulic model of water supply network. Journal of Hehai University (natural science edition) 39:458-463.
[8] Qi Cheng. Research on improvement of K-means clustering algorithm based on differential privacy. p. 62. Dissertation. Master's degree in Computer Science and Technology. Guangxi University, Guangxi, China.