

# Predicting Urban Water Quality with Ubiquitous Data - A Data-driven Approach

Ye Liu, Yuxuan Liang, Kun Ouyang, Shuming Liu, David S. Rosenblum, *Fellow, IEEE*,  
Yu Zheng, *Senior Member, IEEE*

**Abstract**—Urban water quality is of great importance to our daily lives. Prediction of urban water quality help control water pollution and protect human health. However, predicting the urban water quality is a challenging task since the water quality varies in urban spaces non-linearly and depends on multiple factors, such as meteorology, water usage patterns, and land uses. In this work, we forecast the water quality of a station over the next few hours from a data-driven perspective, using the water quality data and water hydraulic data reported by existing monitor stations and a variety of data sources we observed in the city, such as meteorology, pipe networks, structure of road networks, and point of interests (POIs). First, we identify the influential factors that affect the urban water quality via extensive experiments. Second, we present a multi-task multi-view learning method to fuse those multiple datasets from different domains into an unified learning model. We evaluate our method with real-world datasets, and the extensive experiments verify the advantages of our method over other baselines and demonstrate the effectiveness of our approach.

**Index Terms**—Urban computing; data mining; urban water quality prediction; multi-view learning; multi-task learning; big data.

## 1 INTRODUCTION

Urban water is a vital resource that affects various aspects of human, health and urban lives. People living in major cities are increasingly concerned about the urban water quality, calling for technology that can monitor and predict the water quality in real time throughout the city. Urban water quality, which serves as “a powerful environmental determinant” and “a foundation for the prevention and control of waterborne diseases” [2], refers to the physical, chemical and biological characteristics of a water body, and several chemical indexes (such as residual chlorine, turbidity and pH) can be used as effective measurements for the water quality in current urban water distribution systems [3].

With the increasing demand for water quality information, several water quality monitoring stations have been deployed throughout the city’s water distribution system to provide the real-time water quality reports in a city. Figure 1 illustrates the water quality monitor stations that have been deployed in Shenzhen, China. Besides water quality monitoring, predicting the urban water quality plays an essential role in many urban aquatic projects, such as informing waterworks’ decision making (e.g., pre-adjustment of chlorine from the waterworks), affecting governments’ policy making (e.g., issuing pollution alerts or performing a pollution control), and providing maintenance suggestions (e.g., suggestions for replacements of certain pipelines).

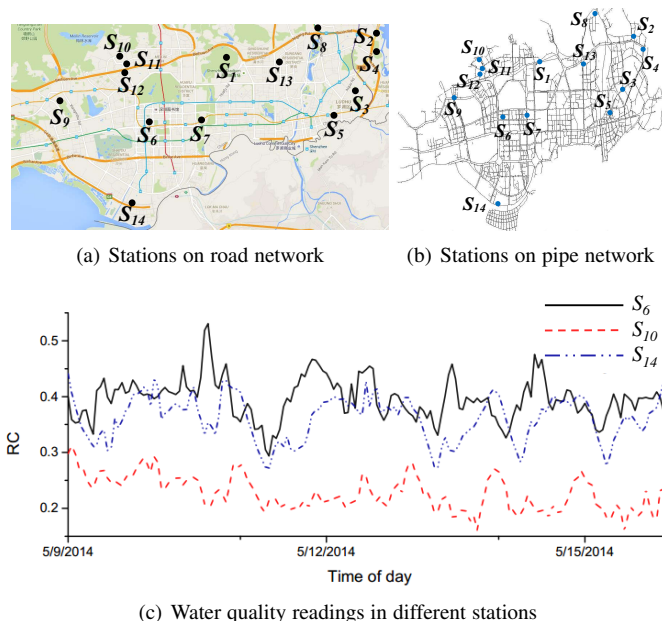


Fig. 1. Illustrations of water quality monitor stations and their readings.

- Ye Liu, Kun Ouyang and David S. Rosenblum are with the Department of Computer Science, School of Computing, National University of Singapore, Singapore.  
E-mail: {liuye, ouyangk, david}@comp.nus.edu.sg
- Yuxuan Liang is with the Department of Computer Science, School of Computer Science and Technology, Xidian University, Xi'an, China.  
E-mail: yuxliang@outlook.com
- Shuming Liu is with the Division of Drinking Water Safety, School of Environment, Tsinghua University, China.  
E-mail: shumingliu@tsinghua.edu.cn
- Yu Zheng is with JD Intelligent City Research, Beijing, China.  
E-mail: msyuzheng@outlook.com

This article is an expanded version of [1], which appeared in Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016).

Predicting urban water quality, however, is very challenging due to the following reasons. First, urban water quality varies by locations non-linearly and depends on multiple factors, such as meteorology, water usage patterns, land use, and urban structures. As depicted in Figure 1, the water quality indexes (RC) reported by the three stations demonstrate different patterns. Existing hydraulic model-based approaches try to model water quality from physical and chemical perspective, but such hydraulic model can hardly capture all of those complex factors. Moreover, the parameters in model are hard to get, which makes it difficult to extend to other water distribution systems. Second, as all the stations are connected

through the pipeline system, the water quality among different stations are mutually correlated by several complex factors, such as attributes in pipe networks and distribution of POIs. Traditional hydraulic model-based approaches build hydraulic model for each station and ignore their spatial correlations, and thus their performance is far from satisfactory. Hence, besides identifying the influential factors, how to efficiently characterize and incorporate such relatedness poses another challenge.

Fortunately, in the era of big data [4] [5], unprecedented data in urban areas (e.g., meteorology, POIs, and road networks) can provide complementary information to help predict the urban water quality. For example, temperature can be an indicator of water quality, with higher temperature indicating better water quality. The possible reason is that the water consumption tends to grow when temperature is high since most people may choose to take a shower, and the increased water consumption is one major cause that prevents the water quality's deterioration in the distribution systems.

To benefit from the unprecedented data in urban areas, in this paper, we predict the water quality of a station through a data-driven perspective using a variety of data sets, including water quality data, hydraulic data, meteorology data, pipe networks data, road networks data, and POIs. First, we perform extensive experiments and data analytics between the water quality and multiple potential factors, and identify the most influential ones that have an effect on the urban water quality. Second, we present a novel spatio-temporal multi-task multi-view learning (stMTMV) framework to fuse the heterogeneous data from multiple domains and jointly capture each station's local information as well as their global information into an unified learning model [1].

We summarize the contributions as follows:

- **Data-driven Perspective:** We present a novel data-driven approach to co-predict the future water quality among different stations with data from multiple domains. Additionally, the approach is not restricted to urban water quality prediction, but also can be applied to other multi-locations based co-prediction problem in many other urban applications.
- **Influential Factor Identification:** We identify spatially-related (such as POIs, pipe networks, and road networks) and temporally-related features (e.g., time of day, meteorology and water hydraulics), contributing to not only our application but also the general problem of water quality prediction.
- **Unified Learning Model:** We present a novel spatio-temporal multi-view multi-task learning framework (stMTMV) to integrate multiple sources of spatio-temporal urban data, which provides a general framework of combining heterogeneous spatio-temporal properties for prediction, and can also be applied to other spatio-temporal based applications.
- **Real evaluation:** We evaluate our method by extensive experiments that use real-world datasets in Shenzhen, China. The results demonstrate the advantages of our method beyond other baselines, such as ARMA, Kalman filter, and ANN, and reveal interesting discoveries that can bring social good to urban life.

The rest of this paper is organized as follows: Section 2 overviews the framework of our method. Section 3 and 4 analyze the correlations between multi-sources of urban data and the water quality. Section 5 introduces the multi-task multi-view learning method for urban water quality prediction, and Section 6 presents evaluations and visualizations. Section 7 summarizes the related work, followed by the conclusion in the last section.

As an extension of our previous work [1], this journal version claims following contributions: First, we focused on the data-driven perspective. Specifically, we included the insight of our methodology as well as the correlation analysis between different data with the urban water quality. The detailed correlation analysis is shown in Section 3 and 4. Second, we refined the task relationship computation in our stMTMV model by figuring out the best configuration over various pipe attributes, which is achieved through the data correlation analysis in Section 5.4.1. Third, we conducted more comprehensive experiments to validate our system. For instance, we added another two popular algorithms (Kalman, ANN) as the time series prediction baselines in Section 6.3. In addition, we compared the performance of our approach with other baselines over each individual station in Section 6.6.

## 2 OVERVIEW

### 2.1 Preliminary

**Definition 1 (Water Quality):** Urban water quality refers to the physical, chemical and biological characteristics of a water body [3]. In current urban water distribution systems, several monitoring stations are deployed in the distribution systems to report three important quality indexes, i.e., *residual chlorine*, *turbidity* and *pH*, in real time. The three indexes can be used as effective measurements for the water quality in current urban water distribution systems [2]. In this paper, we consider *Residual Chlorine (RC)* as the water quality index since it “inactivates the bacteria and some viruses that cause diarrheal disease”, and “can protect the water from recontamination during storage” [6], which is widely employed as the major water quality index in the environmental science [2] [3].

**Definition 2 (Water Hydraulics):** Water hydraulics describe the hydraulic characteristics of the water in urban water distribution systems. It consists of two major indexes: *flow* and *pressure*, and can be obtained through the deployed flow and pressure sensors.

**Definition 3 (Pipe Network):** Urban water is distributed through the pipe network, which is an underground network of interconnecting mains or pipes. A pipe network *PN* comprises of a set of pipeline segments *p* that connect between each other in the format of a graph. Each pipeline segment *p* an undirected edge having two terminal nodes, and has several attributes, including length *p.len*, diameter *p.d*, and age *p.age*.

**Definition 4 (Road Network):** A road network *RN* comprises of a set of road segments *r*, connecting each other in the format of a graph. Each road segment *r* is a directed edge having two terminal nodes, a series of intermediate points between the two terminals, a length of *r.len*.

### 2.2 Insight

In this section, we present the motivation of our method by analyzing the correlations of urban water quality with its influential factors, including hydraulic characteristics, road network, pipe network, meteorology, and point of interest (POI). Figure 2 depicts the interrelationship of various influential factors for the urban water quality, where pink nodes are temporal factors, yellow nodes denote spatial factors, and gray nodes represent unobserved factors. Generally, those factors can be categorized as direct and indirect contributing factors. Those that have intermediate impacts on the water quality are direct contributing factors. For example, hydraulic characteristics and pipe network have direct influence on the water

quality since water from the service reservoirs is distributed to users via extensive networks of water mains and pipe corrosion or flow velocities can cause water quality to deteriorate in a direct way. Hence, hydraulic characteristics and pipe network are main influencing factors for the water quality. Besides direct factors, those that do not have direct influences on the water quality are considered as indirect contributing factors. Take time as an example, time plays an important role in the prediction of water quality, even if it is not directly related to it. One reason is that time is a key factor that affect human behavior, which has an effect on the water quality by influencing the water usage patterns. Similarly, meteorology, road network, POIs are indirect contributing factors for the water quality.

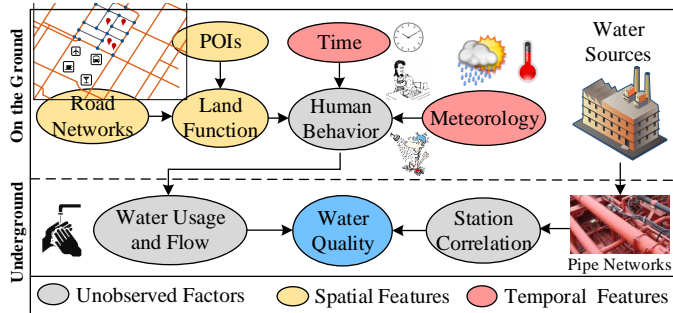


Fig. 2. Illustration of interrelationship among influential factors for the urban water quality prediction.

### 2.3 Framework

Figure 3 presents the framework of our approach, which consists of three major components. 1) Feature extraction. We extract meteorological features, hydraulic features, and time of day from the temporal data in each station, and we also extract road network features and POIs features from each station's spatial data. 2) Multi-view based prediction. After extracting spatial- and temporal-related features, spatial and temporal views are constructed by for each station, and multi-view based prediction is employed to capture the local information within a station (node) and combine its spatial and temporal features. 3) Multi-task based prediction. As the water quality among stations are mutually correlated through the complex water distribution system, we thus can co-predict the water quality over all stations by capturing their spatial correlations. Therefore, multi-task based prediction is used to jointly aligns the predictions among different stations (nodes) by incorporating the their similarities in the pipe networks. We will detail the three major components in the following sections respectively.

## 3 TEMPORAL FEATURE EXTRACTION

### 3.1 Hydraulic Features

It has been well studied in the area of environmental research that water hydraulics, such as water flow and pressure, are closely related to water quality, and the measurement of water hydraulics is an important component of most water quality monitoring projects [7] [8]. As the urban water travels through the water distribution system, the flow reflects its velocity as well as the water age in the system, where the water age is one major contributing factor in water quality deterioration within the distribution system [7]. The reason is that the residual chlorine in

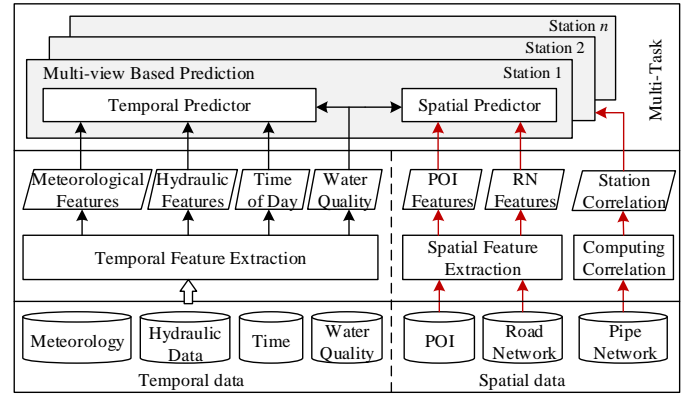


Fig. 3. The framework of our approach.

water normally reacts with organic material and pipe walls, and thus residual disinfectant concentration will diminish as water lingers in the distribution system. If treated water stays in the system a longer time before it reaches users, the disinfectant concentration may not be strong enough after undergoing various chemical reactions.

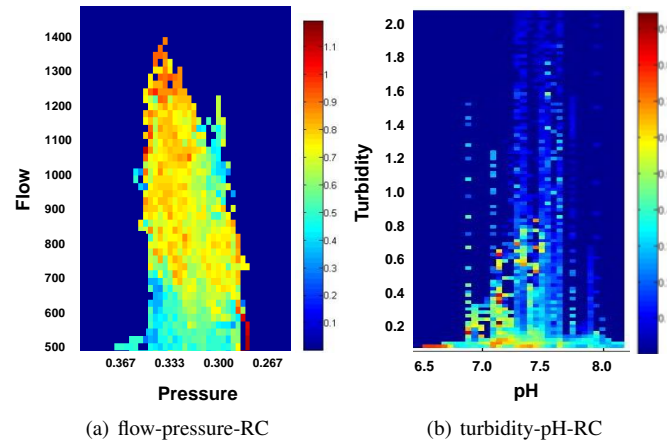


Fig. 4. Correlation between water quality and water hydraulics.

This phenomenon can also be observed from data as illustrated in Figure 4(a), where color, x- and y-axis denote density of RC, pressure and flow, respectively. From Figure 4(a), it can be seen that when water travels with a higher speed in the distribution system, i.e.,  $flow > 700m^3/h$ , more instances of high density of RC occurred. On the contrary, the smaller the flow ( $flow < 700m^3/h$ ) is, the lower density of RC would be. To observe the correlation between flow and RC in Figure 4(a) in a more specific way, first, we fix the pressure value (e.g., set pressure value to 0.333, or other numbers are similar), then the color of grid (density of RC) has shown correlations with the flow readings. For example, when  $500 < flow < 700$ , the grid color is green, indicating  $0.4 < \text{density of RC} < 0.65$ ; when  $700 < flow < 1200$ , the grid color is yellow and orange, indicating  $0.7 < \text{density of RC} < 0.85$ ; when  $1200 < flow < 1400$ , the grid color is a bit more red, indicating  $0.9 < \text{density of RC} < 1.0$ . Hence, those evidences demonstrate the correlation exists between flow and density of RC.

Although the pressure does not have intermediate impact on water age, it has positive effects on the water flow, which, in turn, affects the water age as well as the water quality. An interesting observation is that a larger pressure indicates a higher density of

RC in the water while a smaller one has a very high probability of resulting in a lower density of RC, as depicted in Figure 4(a). In particular, when we keep flow reading to 1000, we can observe that the density of RC decreases as pressure reading decreasing from 0.367 to 0.267, this is because low water pressure in the water system always lead to slower water flow and the water will stay longer in the system. Therefore, the RC will undergo various chemical reactions with the pipe system and density of RC will decrease. Similar analysis can be applied to other hydraulic characteristics (e.g., turbidity, pH). The results well verify the assumption that pressure is another contributing factor of water quality. Thus, water hydraulic characteristics (flow and pressure) are closely correlated with the water quality, and can also be used as the general indicators of water quality.

As flow and pressure are both time series signals, we extract several time series features to capture the characteristics of the signal comprehensively. In particular, we use the latest 12 hours water hydraulic data and extract statistical features (mean, variance, maximum, minimum, skewness and kurtosis), and time series features (autocorrelation, PAA [9], PLA [10]) for each of the time series in the hydraulic data. Moreover, we also extract frequency related features (FFT [11] and DWT [12]) for each of the time series above, where we only use the top 3 coefficients and discard others.

### 3.2 Meteorological Features and Time of Day

The concentration of RC in urban water system is influenced by meteorology indirectly as shown in Figure 2, where meteorology denotes temperature, humidity, barometer pressure, wind speed, and weather. In particular, the meteorology has direct impacts on human behavior, and it thus affects the water usage patterns, which influence the water quality as a consequence. Take the temperature as an example, the temperature has positive correlation with RC and a high temperature is always accompanied with a high concentration of RC. This is partially due to the fact that the high water consumption is growing when the temperature goes high, which in turn has influence on the water quality. This pattern can be observed from the case example of “Area 2” in Figure 5. In particular, the second day shows a higher temperature as compared to the first day, and the concentration of RC on the second day is slightly higher as well.

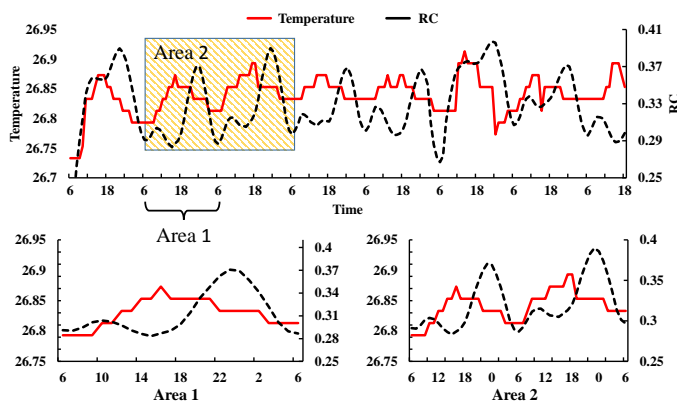


Fig. 5. Correlation between temperature and RC.

Time of day is another contributing cause for the urban water quality due to its strong correlation with human behavior. For instance, 6am to 10am could be the rush hour or morning peak

in most cities, and people in cities normally get up, wash or even take a shower at that time. The increased water consumption leads to a higher concentration of RC in the water system. From Figure 5, it can be seen that the concentration of RC has a slightly growing trend in that period of time. Similar patterns can also be observed from 6pm to 11pm since people usually go home after 6pm and do some cooking or washing up in the evening. On the contrary, the concentration of RC declines gradually after 11am. It is because people go to sleep after 11pm that the water consumption is reduced as well. In addition, some other features (e.g., date/month of year) may also has correlation with human behavior, which in turn influence RC. Since our prediction task is hourly-based, compared to other features, the ‘time of day’ feature has more direct impact on the prediction task. We admit that other features (e.g., date/month of year) are useful as well, but their affects might be small, so we only use the ‘time of day’ for our prediction task.

In summary, meteorology and time of day can provide complementary information of human behavior, helping predict the water quality. In particular, we extract temperature, humidity, barometer pressure, wind speed, and weather as the meteorological features, and time of day as the time feature.

### 3.3 Water Quality

A variety of water quality-related studies have been performed and several research efforts have demonstrated the interrelationship among the three water quality indexes (RC, turbidity, pH) [13] [2] [3]. As a measure of the cloudiness of the water, turbidity was found to be associated with the total organic carbon in the water, which has been shown to interfere with the free chlorine residual by chemical reactions [13]. In particular, turbidity normally exhibits opposite trend with RC since the chemical reactions of RC with pipe and bulk fluid will consume RC and increase the turbidity in water [7]. This negative correlation can also be observed from data as shown in Figure 4(b). Figure 4(b) shows the correlation between the RC and the the other indexes (turbidity and pH), using the data we collected from August to Dec. 2013 in Shenzhen, where color, x- and y-axis denote RC, pH and turbidity, respectively. Apparently, a high density of RC will lead to a low turbidity in the water, and the high turbidity in water occurs as a result of consumption of RC and causes low density of RC in water. Moreover, it can be observed that pH around 7.0 would result in a high density of RC. In short, turbidity and pH are correlated with RC, and can be employed as discriminative features for water quality (RC) predictions.

To fully capture the information from water quality data, we employ similar feature extraction procedure by using the latest 12 hours water quality data and extracting statistical features (mean, variance, maximum, minimum, skewness and kurtosis), frequency related features (FFT [11] and DWT [12]) and time series features (autocorrelation, PAA [9], PLA [10]) for each of the time series in the historical water quality data (RC, turbidity, and pH).

## 4 SPATIAL FEATURE EXTRACTION

### 4.1 Road Network Features

As most of the pipelines are constructed underneath the roads, the structure of a road network in an area, like the number of intersections and the total length of road segments, reflects the structure of the pipe network. Thus, it also has correlation with the region’s water quality, which can be used as complementary



information for water quality prediction. Figure 6 shows the correlation between water quality in a station and a few road network features (e.g. the total length of road segments, and the number of intersections). Each pillar represents one water quality monitoring station, and its height denotes the feature strength. Different colors of the circle around the station stand for different concentrations of RC in the station, e.g. orange denotes a high concentration of RC in a station. So each pillar in Figure 6 shows the concentration of RC in a station with respect to two road network features. For example, the height of purple pillar indicated the total number of road intersections around a water station. Take the second right and third right pillars as example, we can see that the second right has a higher purple height (larger # of road intersections) and its density of RC is higher than the third right. As demonstrated in the height of the pillar and color of the circle around it, we can clearly see that the more intersections are around a station, the more orange circles occur. We also find a similar pattern with respect to the total length of roads in Figure 6.

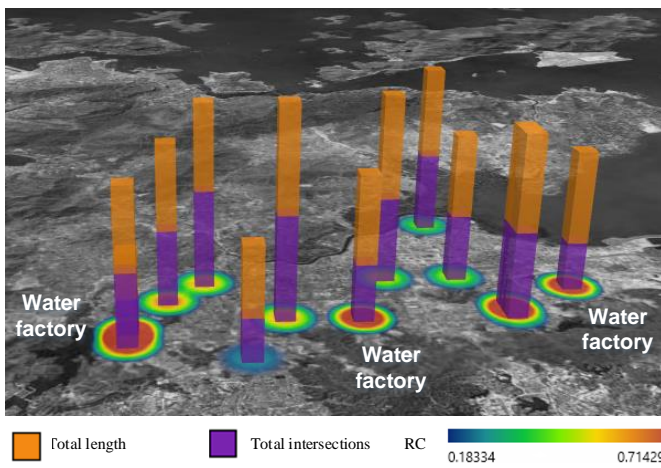


Fig. 6. Correlation between road network features and RC.

Thus, we identify two features for each station based on the road network structure: 1) the total length  $f_{rl}$ , 2) the number of intersections  $f_{rs}$  in the station's affecting region.

## 4.2 POIs Features

The information on POIs in a region, such as the number of POIs in different categories and the density of POIs, indicates the function of the region as well as the water usage patterns in the region, which has an impact on the region's hydraulic conditions (e.g., flow, pressure) and greatly affect water quality as a consequence. For example, if many factories or car repair workshops appears in a region, the consumption of water in that area tends to be high. This leads to a high concentration of RC in the water since the water stays much shorter in the pipeline system and the short time stay can hardly cause chlorine residual to fade. On the contrary, the demand for water in places like parks is usually low, and thus the long detention times in the distribution system result in degraded water quality.

Figure 7 illustrates the correlation between the concentration of RC and the density of POIs among different monitoring stations. Each pillar denotes a water quality monitoring station, and its height represents the density of POIs around the station. The concentration of RC for each station is expressed by the color of the circle around

it, where the orange one denotes a high concentration of RC and the blue one is the opposite. Among 14 monitoring stations, three of them are waterworks. Hence, the concentration of RC in the treated water is high.

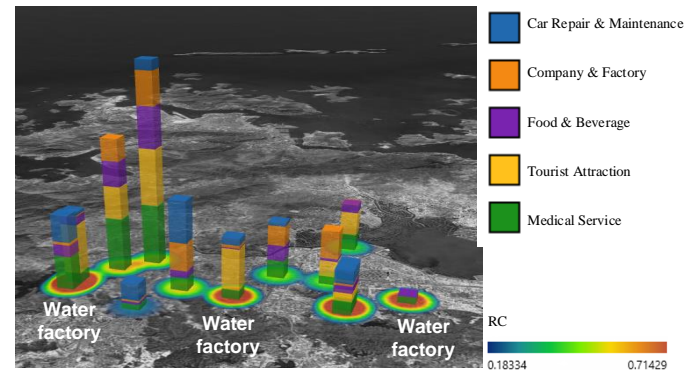


Fig. 7. Correlation between the density of POIs and RC.

As depicted in Figure 7, the density of POIs demonstrates positive correlation with the concentration of RC. In particular, it can be observed that a high density of POIs can cause the concentration of RC to be high in that region. An exceptional case is the station that near the waterworks, which does not demonstrate this correlation strongly. However, those exceptional cases provide evidences of the geographical similarity among stations squarely. So, POI data can also be treated as complementary information to help predict the water quality of a station. As compared to the distributions in road network, there are still some differences between them, and each piece of data may only tell us a part of the panoramic view of urban water quality. Thus, that is the reason why we need to incorporate multiple data sources.

## 5 MULTI-TASK MULTI-VIEW LEARNING MODEL

### 5.1 Notation

We first define some notations. In particular, we use bold capital letters (e.g.,  $\mathbf{X}$ ) and bold lowercase letters (e.g.,  $\mathbf{x}$ ) to denote matrices and vectors, respectively. We employ non-bold letters (e.g.,  $x$ ) to represent scalars, and Greek letters (e.g.,  $\lambda$ ) as parameters. Unless stated, otherwise, all vectors are in column form. Let us assume that we have  $M$  nodes for the water quality prediction, and each node  $l$  is described by its spatial view  $\mathbf{X}_l^s \in \mathbb{R}^{N_l \times D_s} = [\mathbf{x}_{l,1}^s, \mathbf{x}_{l,2}^s, \dots, \mathbf{x}_{l,N_l}^s]^T$  and temporal view  $\mathbf{X}_l^t \in \mathbb{R}^{N_l \times D_t} = [\mathbf{x}_{l,1}^t, \mathbf{x}_{l,2}^t, \dots, \mathbf{x}_{l,N_l}^t]^T$ , where  $\mathbf{x}_{l,i}^s \in \mathbb{R}^{D_s}$  and  $\mathbf{x}_{l,i}^t \in \mathbb{R}^{D_t}$  denote the spatial feature and temporal feature extracted from the node  $l$  at time point  $i$ .  $N_l$  is the number of samples at node  $l$ , and  $D_s$  and  $D_t$  is the feature dimension of the spatial view and temporal view, respectively. The whole feature matrix at node  $l$  can be written as  $\mathbf{X}_l = [\mathbf{X}_l^s, \mathbf{X}_l^t] \in \mathbb{R}^{N_l \times D}$ , where  $D = D_s + D_t$ . The target vector at node  $l$  is denoted as  $\mathbf{y}_l = \{\mathbf{y}_{l,1}, \mathbf{y}_{l,2}, \dots, \mathbf{y}_{l,N_l}\} \in \mathbb{R}^{N_l}$ , which represents the water quality of node  $l$  observed at the discrete time points  $1, 2, \dots, N_l$ .  $N = \sum_{l=1}^M N_l$  is the total number of samples over all nodes.

### 5.2 Technical Formulation

Technically, our multi-task multi-view learning model consists of two components: multi-view learning and multi-task learning, and each component plays a different role in our model. In

particular, multi-view learning aims to combine the two different views within a station, which helps boost the local predictive performance. The multi-task learning captures the pipe network correlations among different stations and aligns their predictions, and this alignment can improve the co-predictive performance over all stations. By combining the multi-view learning and multi-task learning components, we can get an unified spatio-temporal multi-task multi-view learning model (stMTMV) to perform the water quality co-prediction problem. Moreover, the dimension of features for prediction is usually very high, but not all features are sufficiently discriminative for water quality prediction. To select a common set of discriminative features among all tasks, we employ a group Lasso penalty [14], which can identify the top sharing features automatically. The overall objective function for the spatio-temporal multi-task multi-view learning model (stMTMV) can be formulated as

$$\min_{\mathbf{W}} \frac{1}{2} \sum_{l=1}^M \|\mathbf{y}_l - \frac{1}{2} \mathbf{X}_l \mathbf{w}_l\|_2^2 + \lambda \sum_{l=1}^M \|\mathbf{X}_l^s \mathbf{w}_l^s - \mathbf{X}_l^t \mathbf{w}_l^t\|_2^2 + \gamma \sum_{l,m=1}^M C_{l,m} \|\mathbf{w}_l - \mathbf{w}_m\|_2^2 + \theta \|\mathbf{W}\|_{2,1}, \quad (1)$$

where the first term is the loss function, the second term is the multi-view learning component, the third term is multi-task learning component and  $\lambda, \gamma, \theta$  are regularization parameters. The  $\ell_{2,1}$ -norm of a matrix  $\mathbf{W}$  is defined as  $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^D \sqrt{\sum_{j=1}^M W_{ij}^2}$ . In particular,  $\ell_{2,1}$ -norm applies an  $\ell_2$ -norm to each row of  $\mathbf{W}$  and these  $\ell_2$ -norms are combined through an  $\ell_1$ -norm. As we assume that only a small set of features are predictive for prediction task, the  $\ell_{2,1}$ -norm encourages all tasks to select a common set of features and thereby plays the role of group feature selection.

### 5.3 Multi-view Learning

Multi-view learning has been proposed to leverage the information from diverse domains or from various feature extractors, and combining the heterogeneous properties from different views can better characterize objects and achieve promising performance [4] [15]. In this work, as there are various spatial- and temporal data around a water quality station, the information in a station can be represented by two separate views, i.e., spatial view and temporal view, thus, combining the two different views within a station can achieve better performance.

#### 5.3.1 Spatio-temporal Views Construction

First, we construct the spatio-temporal views by extracting spatial- and temporal-related features for each station from spatial and temporal datasets, respectively. In particular, the temporal view of each station is constructed by incorporating its temporal information, such as historical water quality data (RC, turbidity, pH), historical hydraulic data (flow, pressure), and meteorological data, and extracting temporal features from above data. The spatial view of each station is constructed by integrating its spatial inferential factors, such as pipe network, road network, and POIs, and extracting spatial features as well as the aggregated surrounding temporal features, which encodes its spatial information as well.

#### 5.3.2 Technical Formulation

Next, we will get into the technical formulation of multi-view learning. The prediction at each node  $l$  consists of spatial prediction and temporal prediction, i.e.,  $f_l^s(\mathbf{X}_l^s) = \mathbf{X}_l^s \mathbf{w}_l^s$  for spatial prediction

and  $f_l^t(\mathbf{X}_l^t) = \mathbf{X}_l^t \mathbf{w}_l^t$  for temporal prediction, where  $\mathbf{w}_l^s \in \mathbb{R}^{D_s}$  and  $\mathbf{w}_l^t \in \mathbb{R}^{D_t}$  denote the linear mapping function for the node  $l$  with spatial view and temporal view, respectively. In this paper, we use the linear function for simplicity. However, our model can be extended to other convex and smooth non-linear prediction functions. Without prior knowledge on the contributions of spatial and temporal view, we assume that both contribute equally. Thus, the final prediction model of both spatial and temporal view for node  $l$  is obtained by the following late fusion [16] [15]:

$$f_l(\mathbf{X}_l) = \frac{1}{2} (f_l^s(\mathbf{X}_l^s) + f_l^t(\mathbf{X}_l^t)) = \frac{1}{2} (\mathbf{X}_l^s \mathbf{w}_l^s + \mathbf{X}_l^t \mathbf{w}_l^t) = \frac{1}{2} \mathbf{X}_l \mathbf{w}_l, \quad (2)$$

where  $\mathbf{w}^l \in \mathbb{R}^D$  is the weight vector for node  $l$ . The weight matrix over  $M$  nodes is denoted as  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M] \in \mathbb{R}^{D \times M}$ .

Information distributed in spatial and temporal views in fact describes the inherent characteristics of the same node from various aspects, we thus can reinforce the learning performance of individual views by enforcing the agreement on the their prediction results. Considering the least-squares loss function, we can define the following objective function [16] [15]:

$$\min_{\mathbf{W}} \frac{1}{2} \sum_{l=1}^M \|\mathbf{y}_l - \frac{1}{2} \mathbf{X}_l \mathbf{w}_l\|_2^2 + \lambda \sum_{l=1}^M \|\mathbf{X}_l^s \mathbf{w}_l^s - \mathbf{X}_l^t \mathbf{w}_l^t\|_2^2, \quad (3)$$

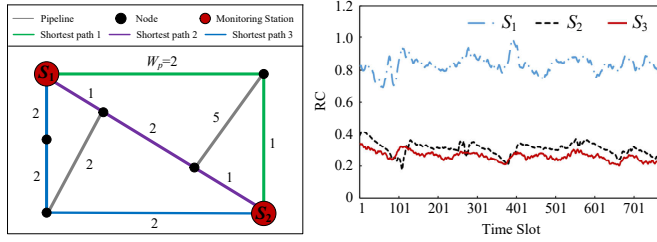
where  $\lambda$  is a regularization parameter,  $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$  is the  $\ell_2$ -norm. In a real pipeline system, each node is not only affected by its local information, but also affected by the information from its neighbors or other nodes.

### 5.4 Multi-task Learning

Multi-task learning is a learning paradigm that jointly learns multiple related tasks and has demonstrated its advantages in many urban applications [17] [5]. In particular, it is more effective in handling those with insufficient training samples [18] [19] [17]. In this work, the water quality prediction in each node is aligned with a task. As the urban water is distributed across the water distribution systems, and hence the water quality among stations are mutually correlated. Therefore, we can co-predict the water quality over all stations by capturing the spatial correlations among them, and these correlations are pre-computed by the structure of pipe network.

#### 5.4.1 Pipe Network Correlation

The water quality among stations are mutually correlated through the complex water distribution system and thus there exist spatial correlations among them. Therefore, how to capturing those spatial correlations becomes a quite important question. As water is distributed through the water distribution system, the pipe network becomes the major contributing factor for the spatial correlations. The effects of various attributes of water pipes on the water quality have also been explored in the environmental science [20], and the impact of water pipes on the water quality varies from system to system, depending on the attributes of each pipe as well as the whole pipe network configuration [20]. Based on existing urban water system, among various pipe attributes, only three attributes of a pipe are available, i.e., *length*, *diameter*, and *age*. Those three attributes of a pipe are potential water quality deterioration factors that play an important role in affecting the water age, which consequently influences water quality. In this paper, we computed the spatial correlations by taking account of the three pipe attributes and the pipe network configurations.



(a) Pipe network correlation of stations (b) RC readings of stations

Fig. 8. Illustration of pipe network correlation among stations.

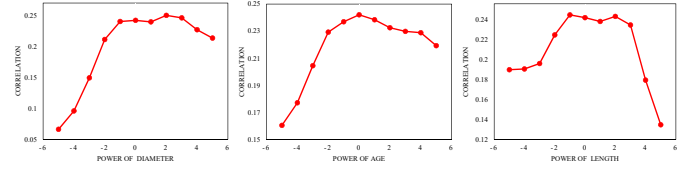
Take the correlation between station  $S_1$  and  $S_2$  as an example, there are multiple paths between  $S_1$  and  $S_2$  in the pipe network. Given all of these paths, the correlation between  $S_1$  and  $S_2$  is affected by the pipe attributes along those paths. Specifically, this correlation is not only affected by the length of a pipe, but also influenced by other factors, including the diameter and age of a pipe. For each pipe  $p$  in the network, we employ the three most important attributes (*length*, *diameter* and *age*) and define its weight  $w_p$  in this network as a function of the three attributes, i.e.,  $w_p = f(p.d, p.len, p.age)$  (we will give the definition of  $w_p$  in later section). Figure 8(a) illustrates the computation of correlation between  $S_1$  and  $S_2$ , where the number along each edge denotes its weight  $w_p$ . As the interconnections between  $S_1$  and  $S_2$  are quite complicated, therefore a single path can hardly characterize their correlation; Consequently, we employ the top- $k$  shortest paths as the characterization of those interconnections and calculate  $C_{i,j}$  by the sum of top- $k$  shortest paths with normalization. Specifically,  $C_{i,j}$  is computed as follows:

$$C_{i,j} = \frac{1}{k} \sum_{t=1}^k \sum_{p \in \text{path}(i,j,t)} w_p, \quad (4)$$

where  $\text{path}(i,j,t)$  represents the  $t$ -th shortest path between  $S_i$  and  $S_j$ . As shown in Figure 8(a), there are multiple paths between  $S_1$  and  $S_2$ . When  $k = 1$ , we can get  $C_{1,2}$  by the shortest path (the green path) between  $S_1$  and  $S_2$ , where  $C_{1,2} = \frac{1}{1} \sum_{t=1}^1 \sum_{p \in \text{path}(1,2,t)} w_p = 2 + 1 = 3$ . If we set  $k = 3$ , the correlation between  $S_1$  and  $S_2$  is computed by  $C_{1,2} = \frac{1}{3} \sum_{t=1}^3 \sum_{p \in \text{path}(1,2,t)} w_p = \frac{(2+1)+(1+2+1)+(2+2+2)}{3} = \frac{13}{3}$ . The choice of  $k$  is based on experimental results. In particular, in the experiments, we varied  $k$  from 3 to 10 for the top- $k$  shortest path computation, and for each  $k$ , we compute a correlation matrix  $C^{(k)}$  and evaluate the predictive performance of our stMTMV model with the correlation matrix  $C^{(k)}$ , where  $k = 5$  gives the best overall model predictive performance, so we keep  $k = 5$  for all settings in our experiments.

Similarly, we also compute the correlation between  $S_2$  and  $S_3$  via Eqn. (4). In Figure 8(b), as compared to the correlation between  $S_1$  and  $S_2$ ,  $S_2$  and  $S_3$  has a larger correlation in the pipe network. From Figure 8(b), it can be seen that the concentrations of RC in  $S_2$  and  $S_3$  demonstrate the similar patterns, while the patterns and the readings in  $S_1$  are quite different from  $S_2$  and  $S_3$ , which verifies the existence of pipe network correlations among stations. Hence, capturing these spatial correlations can help predict the water quality accurately. One possible way of getting these spatial correlations is to use the Pearson correlation of the stations' historical data. However, this method has its own limitations if the historical data in some of stations is not available. Therefore,

we take steps to another direction, which analyzes the spatial correlations by using the pipe network attributes.



(a) Power of diameter (b) Power of age (c) Power of length

Fig. 9. Illustration of the impacts of various pipe attributes (*diameter*, *age*, *length*) on the station correlations.

To get a comprehensive understanding of the effects, we investigate how different configurations of *length*, *diameter*, *age* of a pipe can affect the correlations among different stations. Figure 9 comparatively illustrates the correlation results of all stations with respect to various pipe attributes by controlling other factors. In particular, we control the age and length by setting them to zero, and observe the effects of the power of diameter on the station correlations. From Figure 9(a), it can be observed that the correlation shows an increasing trend from  $-5$  to  $-1$  and stabilize from  $-1$  to  $1$ , then arrives its peak at  $2$  followed by a decline. This demonstrate that, when keeping the other two factors (*age*, *length*) as zero, the square of diameter affects station correlations the most. Similarly, we can get the observations that age to the zero power and length to the negative one power have the greatest influence on the correlations while keeping the power of the other two as zeros.

Moreover, the optimal value of power for the *diameter*, *age* and *length* of a pipe is determined by experiments. Table 1 illustrates the top power triplets  $\langle \text{pow}_d, \text{pow}_{len}, \text{pow}_{age} \rangle$  which give the highest station correlations. In particular, the steps for getting the station correlations are as follows:

- 1) Suppose there are  $M$  stations in total, we can thus get a  $M \times M$  correlation matrix *CorrMat*, which encodes the pairwise Pearson correlations of one year's RC readings from the  $M$  stations.
- 2) The pipe network can be seen as a weighted graph, where the weight  $w$  for a pipe  $p$  is computed from its attributes via  $w_p = p.d^{\text{pow}_d} \times p.len^{\text{pow}_{len}} \times p.age^{\text{pow}_{age}}$
- 3) For each power triplet  $\langle \text{pow}_d, \text{pow}_{len}, \text{pow}_{age} \rangle$ , where each entry in the triplet is iterated from  $-5$  to  $5$ , we can get a  $M \times M$  weight matrix *WeightMat* by computing the top- $k$  shortest paths between any two stations.
- 4) Thereafter, the correlation value in Table 1 is calculated by the correlation between *CorrMat* and *WeightMat*, which indicates the degree of approximation of *WeightMat* for the *CorrMat*.

Interestingly, it can be observed that the power triplet  $\langle 2, -1, -1 \rangle$ , which gives the second largest correlation result, is closest to the previous observations, i.e., the optimal value with the others fixed. Also, its correlation result is close to the top correlation result. Therefore, we choose the power triplet  $\langle 2, -1, -1 \rangle$  as the power of *diameter*, *length* and *age* for computing the weight  $w_p$  of a pipe  $p$ . Specifically,  $w_p$  is calculated by Eqn. 5,

$$w_p = p.d^2 \times p.len^{-1} \times p.age^{-1}. \quad (5)$$

To utilize these spatial correlations for co-predictions, we pre-compute those correlations through the structure of pipe network. In particular, the pipe network can be seen as a weighted graph,

TABLE 1  
Illustration of correlations from the top power triplets.

Power of Diameter	Power of Length	Power of Age	Correlations
1	2	2	0.370383
2	-1	-1	<b>0.370336</b>
0	2	2	0.367362
2	-2	2	0.366843
2	-1	0	0.366589
2	2	2	0.365698
0	2	1	0.364624

where the weight  $w$  for a pipe  $p$  is computed from the attributes of the pipe by  $\frac{p.d^2}{p.len \cdot p.age}$ . Given station  $S_i$  and  $S_j$ , as there are multiple different paths between  $S_i$  and  $S_j$ , their pipe network correlation  $C_{i,j}$  is computed by the sum of top- $k$  shortest path between them.

#### 5.4.2 Technical Formulation

Next, we will get into the technical details of the multi-task learning. Given the  $M$  nodes for water quality prediction, the whole feature matrix  $\mathbf{X}_l = [\mathbf{X}_l^s, \mathbf{X}_l^t] \in \mathbb{R}^{N_l \times D}$ , and the target vector at  $\mathbf{y}_l = \{\mathbf{y}_{l,1}, \mathbf{y}_{l,2}, \dots, \mathbf{y}_{l,N_l}\} \in \mathbb{R}^{N_l}$ , for each node  $l$ , we can formulate the water quality prediction over all stations as a multi-task learning problem by considering the global impact on node  $l$ . In particular, we expand the model in Eqn. (3) to incorporate a graph Laplacian penalty among node  $l$  and the other nodes. This penalty ensures a small deviation between two nodes that are near in the pipeline system, and incorporates the domain knowledge about the spatial correlations of the water quality among different nodes in the pipeline systems. We formulate the multi-task learning problem as [21] [22] [23]

$$\min_{\mathbf{W}} \frac{1}{2} \sum_{l=1}^M \|\mathbf{y}_l - \frac{1}{2} \mathbf{X}_l \mathbf{w}_l\|_2^2 + \gamma \sum_{l,m=1}^M C_{l,m} \|\mathbf{w}_l - \mathbf{w}_m\|_2^2, \quad (6)$$

where  $\gamma$  is a regularization parameter,  $C_{l,m}$  is the pipe network correlation between task (station)  $l$  and task (station)  $m$ , and measures the spatial autocorrelation between task  $l$  and  $m$ . Intuitively, if  $C_{l,m}$  is large, the graph Laplacian regularizer term will force  $\mathbf{w}_l$  to be as similar as  $\mathbf{w}_m$ . Thus, this graph Laplacian penalty automatically encodes Tobler's first law of geography [24]. In implementation, we can pre-compute the correlation  $C_{l,m}$  by top- $k$  shortest path in the pipe networks.

#### 5.5 Optimization

The optimization of Eqn.(1) is convex with respect to  $\mathbf{W}$ . First, we can rewrite the graph Laplacian term in Eqn.(1) as

$$\sum_{l,m=1}^M C_{l,m} \|\mathbf{w}_l - \mathbf{w}_m\|_2^2 = \text{tr}(\mathbf{W}(\mathbf{D} - \mathbf{C})\mathbf{W}^T) = \text{tr}(\mathbf{W}\mathbf{L}\mathbf{W}^T) \quad (7)$$

where  $\mathbf{D}$  is a diagonal matrix with  $D_{l,l} = \sum_m C_{l,m}$ ,  $\mathbf{C}$  is the similarity matrix, and  $\mathbf{L} = \mathbf{D} - \mathbf{C}$  is known as the Laplacian matrix [21] [22]. We define

$$h(\mathbf{W}) = \frac{1}{2} \sum_{l=1}^M \|\mathbf{y}_l - \frac{1}{2} \mathbf{X}_l \mathbf{w}_l\|_2^2 + \lambda \sum_{l=1}^M \|\mathbf{X}_l^s \mathbf{w}_l^s - \mathbf{X}_l^t \mathbf{w}_l^t\|_2^2 + \gamma \text{tr}(\mathbf{W}\mathbf{L}\mathbf{W}^T), \quad (8)$$

$$g(\mathbf{W}) = \theta \|\mathbf{W}\|_{2,1}. \quad (9)$$

The optimization in Eqn.(1) can be rewritten as  $\min_{\mathbf{W}} h(\mathbf{W}) + g(\mathbf{W})$ , where  $h(\mathbf{W})$  is smooth and  $g(\mathbf{W})$  is non-smooth. We can thus use the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [25] or Accelerated Gradient Descent (AGD) [26] to solve it. One of the key steps in using FISTA algorithm is to solve the proximal step [25]:

$$\mathbf{W}^{(k)} = \underset{\mathbf{W}}{\text{argmin}} \left\{ g(\mathbf{W}) + \frac{L_k}{2} \|\mathbf{W} - (\mathbf{V}^{(k)} - \frac{1}{L_k} \nabla h(\mathbf{V}^{(k)}))\|_F^2 \right\}, \quad (10)$$

where  $\mathbf{V}^{(k)}$  is the search point and is defined by the affine combination of  $\mathbf{W}^{(k-1)}$  and  $\mathbf{W}^{(k-2)}$ ; and  $L_k$  is a scalar that can be determined by the line search method [25]. The gradient  $\nabla h(\mathbf{W})$  for each column  $\mathbf{w}_l$  in Eqn. (10) can be computed as

$$\nabla h(\mathbf{w}_l) = \frac{1}{2} \mathbf{X}_l^T (\frac{1}{2} \mathbf{X}_l \mathbf{w}_l - \mathbf{y}_l) + \lambda \mathbf{P} \mathbf{w}_l + 2\gamma \mathbf{W} \mathbf{L}_l, \quad (11)$$

where  $\mathbf{L}_l$  is the  $l$ -th column of  $\mathbf{L}$ ,  $\mathbf{P} \in \mathbb{R}^{D \times D}$  is a block matrix with  $2 \times 2$  blocks, and  $\mathbf{P}_{11} = 2(\mathbf{X}_l^s)^T \mathbf{X}_l^s$ ,  $\mathbf{P}_{12} = -2(\mathbf{X}_l^s)^T \mathbf{X}_l^t$ ,  $\mathbf{P}_{21} = -2(\mathbf{X}_l^t)^T \mathbf{X}_l^s$ ,  $\mathbf{P}_{22} = 2(\mathbf{X}_l^t)^T \mathbf{X}_l^t$ .

As Eqn.(10) is computed in every FISTA iteration, it needs to be solved efficiently. Specifically, each row  $\mathbf{w}^i$  of  $\mathbf{W}$  can be decoupled in Eqn.(10). Thus, to obtain the  $i$ -th row  $\mathbf{w}^{i(k)}$  in  $\mathbf{W}^{(k)}$ , we only need to solve the following problem:

$$\mathbf{w}^{i(k)} = \underset{\mathbf{w}}{\text{argmin}} \left\{ \theta \|\mathbf{w}\|_2 + \frac{L_k}{2} \|\mathbf{w} - (\mathbf{v}^{i(k)} - \frac{1}{L_k} \nabla f(\mathbf{v}^{i(k)}))\|_2^2 \right\}, \quad (12)$$

where  $\mathbf{v}^{i(k)}$  is the  $i$ -th row of  $\mathbf{V}^{(k)}$ . It can be reformulated as

$$\mathbf{w}^{i(k)} = \underset{\mathbf{w}}{\text{argmin}} \left\{ \frac{1}{2} \|\mathbf{w} - \mathbf{b}\|_2^2 + \beta \|\mathbf{w}\|_2 \right\}, \quad (13)$$

where  $\mathbf{b} = \mathbf{v}^{i(k)} - \frac{1}{L_k} \nabla f(\mathbf{v}^{i(k)})$  and  $\beta = \frac{\theta}{L_k}$ . It has been proved by Wright *et al.* [27] that Eqn.(13) have the closed-form solution,

$$\mathbf{w}^{i(k)} = \max(0, 1 - \frac{\beta}{\|\mathbf{b}\|_2}) \mathbf{b}. \quad (14)$$

We thus can solve Eqn.(10) quite efficiently. The FISTA algorithm is depicted in Algorithm 1, where  $\|\mathbf{W}\|_F$  is the Frobenius norm and  $L_k$  is a scalar that can be determined by the line search method [25].

#### Algorithm 1 FISTA algorithm for solving Eqn.(1)

- 1: **Input:** Iterations  $K$
- 2: **Initialize:**  $\mathbf{V}^{(1)} = \mathbf{W}^{(0)} \in \mathbb{R}^{D \times M}$ ,  $t_1 = 1$
- 3: **for**  $k = 1, 2, 3, \dots, K$  **do**
- 4:  $\mathbf{W}^{(k)} \leftarrow \underset{\mathbf{W}}{\text{argmin}} \left\{ g(\mathbf{W}) + \frac{L_k}{2} \|\mathbf{W} - (\mathbf{V}^{(k)} - \frac{1}{L_k} \nabla f(\mathbf{V}^{(k)}))\|_F^2 \right\}$
- 5:  $t_{k+1} \leftarrow \frac{1 + \sqrt{1 + 4t_k^2}}{2}$
- 6:  $\mathbf{V}^{(k+1)} \leftarrow \mathbf{W}^{(k)} + (\frac{t_k - 1}{t_{k+1}})(\mathbf{W}^{(k)} - \mathbf{W}^{(k-1)})$
- 7: **end for**
- 8: **Output:**  $\mathbf{W}^{(K)}$

#### 5.6 Computational Complexity Analysis

In this section, we discuss the computational complexity for solving the stMTMV model. For the optimization of  $\mathbf{W}$ , the main computational cost comes from calculating the gradient  $\nabla h(\mathbf{W})$  in each FISTA iteration. In particular, as  $\mathbf{X}_l^T \mathbf{X}_l$ ,  $\mathbf{X}_l^T \mathbf{y}_l$  and  $\mathbf{P}$  remain unchanged in every FISTA iteration, we can pre-compute and store them with time complexity of  $O(ND^2)$ . In each FISTA iteration, the time complexity for calculating the gradient  $\nabla h(\mathbf{W})$  in Eqn.



(11) is  $O((D+M)DM)$ , and computing the closed-form solution in Eqn.(14) takes  $O(DM)$  time for  $\mathbf{W}^{(k)}$ . Hence, the complexity for each iteration in the FISTA algorithm is  $O((D+M)DM)$ . Moreover, the FISTA algorithm converges within  $O(1/\varepsilon^2)$  iterations, and the total time cost of FISTA for solving stMTMV is  $O(\frac{(D+M)DM}{\varepsilon^2})$ , where  $\varepsilon$  is the desired accuracy. Thus, the stMTMV model can be solved efficiently. Since the per-iteration complexity of FISTA for solving stMTMV is independent of  $N$ , which shows that our model can potentially scale to large-scale urban data.

## 6 EVALUATION

### 6.1 Datasets

In the evaluation, we use the following six real datasets detailed in Table 2, where the data are collected from August 2011 to August 2014 in Shenzhen City, China.

TABLE 2  
Details of the datasets

Data Sources		Shenzhen
Water Quality	Stations	14
	Time Spans	8/1/2011-7/31/2014
	Hours	17,544
POIs	POIs	185,841
	Categories	20
Road Network	Segments	149,559
	Total length	15,132km
Pipe Network	Nodes	8,707
	Segments	6,152
	Max Connected Component	3,511
Meteorology	Hours	55,633
Hydraulic Data	Flow Stations	10
	Pressure Stations	13
	Time Spans	8/1/2011-7/31/2014

- 1) Water quality data: We collect water quality data every five minutes from 15 water quality monitoring stations in Shenzhen City. It comprises residual chlorine (RC), turbidity (TU) and pH. In this paper, we only use RC as the index for water quality, since RC is the most important and effective measurement for water quality in current urban water distribution system [3], and the range of RC density is normally within 0.0-1.0 mg/L (refer to Figure 4(a) and 4(b)).
- 2) Hydraulic data: Hydraulic data consists of flow and pressure, which are collected every five minutes from 13 flow sites and 14 pressure sites, respectively.
- 3) Road networks data: Each road segment is associated with two terminal points and some properties, such as level, capacity and speed limit.
- 4) Pipe attributes data: It describes the pipe attributes in water distribution system, and the attributes of a pipe consist of diameter, length, age, material, etc.
- 5) Meteorology data: Meteorological data consists of weather, temperature, humidity, barometer pressure, wind strength, which is collected every hour.
- 6) POIs: There are 185841 POIs of 20 categories. Each POI has a name, category, address and geo-coordinates.

### 6.2 Ground Truth and Metrics

We can predict the water quality of a site from its historical data, and the ground truth is obtained from its later readings. In particular, we evaluate the predictive performance with respect to its readings in next 1, 2, 3, 4 hours, and the performance is evaluated in terms of their root-mean-square-error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{l=1}^M (y_l - \hat{y}_l)^2}. \quad (15)$$

### 6.3 Learning Model Comparison

To validate our stMTMV model, we compared it with the following baselines:

- RC Decay Model (Classical): Residual Chlorine (RC) decay model is a simple (also well-known) mathematical model that is typically applied in environmental science to model and predict chlorine residual in water supply systems [8] [3]. This model describes both bulk and wall chlorine consumption via first order decay kinetics  $\frac{dC}{dt} = -kC$ , where  $k$  is the first order chlorine decay constant that depends on the distribution systems.
- ARMA: Auto-Regression-Moving-Average (ARMA) is a well-known model for predicting time series data, which makes predictions solely based on historical data [28].
- Kalman: Kalman filter is a traditional predictive model that uses a series of measurements observed data over time and predicts the future value in a probabilistic manner [29].
- ANN: Artificial Neural Network (ANN) is feeded with all features from different views into a single model without treating different views differently. Employing ANN as a baseline is to justify the advantages of using a combination of multiple views. As ANN has also been used in environmental science to predict water quality [30], surpassing ANN also justifies our contribution over traditional approaches.
- LR: Linear Regression (LR) [31] is applied for each node individually, which is a single-task learning method.
- LASSO: Lasso [32] tries to minimize the objective function  $\frac{1}{2} \sum_{l=1}^M \|\mathbf{y}_l - \mathbf{X}_l \mathbf{w}_l\|_2^2 + \alpha \|\mathbf{W}\|_1$  and encodes the sparsity over all weights in  $\mathbf{W}$ . It keeps task-specific features but ignores the task-sharing features.
- MRMTL: As a typical example of traditional multi-task learning, Mean-Regularized Multi-Task Learning (MRMTL) [33] [26] assumes all tasks are related and penalizes the deviation of each task from their mean by optimizing  $\frac{1}{2} \sum_{l=1}^M \|\mathbf{y}_l - \mathbf{X}_l \mathbf{w}_l\|_2^2 + \lambda \sum_{l=1}^M \|\mathbf{w}_l - \frac{1}{M} \sum_{m=1}^M \mathbf{w}_m\|_2^2 + \theta \|\mathbf{W}\|_F^2$ .
- regMVMT: The regularized multi-view multi-task learning model (regMVMT) [16] jointly regularizes view consistency and uniform task relatedness.
- stMTMV-IJCAI: Our previous results in IJCAI-16 [1].

The experimental results are demonstrated in Table 3. From this table, we have the following observations: 1) The prediction accuracy of all models shows a decrease trend for the next 1-4 hours. This is consistent with the intuition that the distant future tends to be more difficult to forecast than the near future. 2) The last six machine learning approaches demonstrate superior performance over first three methods, which justifies the effectiveness of extracted features as well as the benefits of data fusion from multiple sources. Moreover, it is not unexpected that RC Decay Model achieves the worst performance since it may fail to capture the real dynamics of the RC in the system.

3) The last four multi-task learning methods stably outperform the single-task learning methods (ANN, LR), which verifies that the tasks are not independent and that jointly learning them can boost learning performance. 3) The accuracies of MRMTL are slightly lower than other multi-task learning methods. This may be caused by the inappropriate assumption of penalizing the deviation of each task from their mean, since these tasks tend to be spatially autocorrelated. 4) As compared to MTL, our model and regMVMT achieve higher performance due to the fact that stMTMV and regMVMT can incorporate heterogeneous information from spatial and temporal views, which may help to improve overall performance. 5) The stMTMV model shows superiority over regMVMT, which underscores the importance of incorporating structure of the water distribution system and this structure can further improve performance. 6) The outperformance of stMTMV over stMTMV-IJCAI demonstrates that the refined  $C_{i,j}$  can better characterize the geographical correlation among stations and hence boost the overall performance as a consequence. In overall, since the range of RC density is normally within 0.0-1.0 mg/L (refer to Figure 4(a) and 4(b)), our stMTMV model can achieve 0.089 in RMSE, which demonstrates the predictive accuracy of our model.

TABLE 3

Predictive performance comparison among various approaches over next 1-4 hours.

Model Comparison	1 hour	2 hour	3 hour	4 hour
RC Decay Model	3.51e-1	3.53e-1	3.59e-1	3.68e-1
ARMA	1.86e-1	2.18e-1	2.46e-1	2.78e-1
Kalman	1.79e-1	1.93e-1	1.89e-1	1.90e-1
ANN	1.50e-1	1.54e-1	1.56e-1	1.63e-1
LR	1.68e-1	1.99e-1	2.09e-1	2.10e-1
LASSO	1.23e-1	1.42e-1	1.52e-1	1.56e-1
MRMTL	1.32e-1	1.48e-1	1.56e-1	1.58e-1
regMVMT	1.06e-1	1.15e-1	1.18e-1	1.19e-1
stMTMV-IJCAI	9.33e-2	9.66e-2	9.80e-2	9.90e-2
stMTMV	<b>8.99e-2</b>	<b>9.54e-2</b>	<b>9.71e-2</b>	<b>9.83e-2</b>

## 6.4 Evaluation on Model Components

To evaluate each component of the stMTMV model, we compared it with three different variants of stMTMV:

- **stMTMV-us**: In this variant, uniform spatial correlation is used to evaluate the importance of spatial correlation among tasks. We can derive it by setting  $S = I$ .
- **stMTMV-ws**: This is a derivation of stMTMV without group sparsity. We can derive it by setting  $\theta = 0$ .
- **stMTMV-sv**: This derivation is to evaluate the importance of spatio-temporal view alignment. We can derive it by setting  $\lambda = 0$ .

The experimental results are demonstrated in Figure 10. From this figure, it can be seen that stMTMV-us achieves the worst performance, which demonstrates the effectiveness of graph Laplacian component in the stMTMV model. This further verifies that the tasks are mutually correlated and the spatial autocorrelation plays an important role in the co-prediction tasks. Moreover, stMTMV-ws achieves the second worst performance, which justifies

the importance of group sparsity in the stMTMV model. This also provides evidence for the assumption that only a small set of features are predictive for the water quality prediction tasks. Compared to stMTMV-ws and stMTMV-us, the effect of spatio-temporal view alignment tend to be less strong, and this is observed by the superior performance of stMTMV-sv over other two variants. However, stMTMV outperforms stMTMV-sv since spatio-temporal view alignment can combine heterogeneous spatio-temporal information and further boost performance.

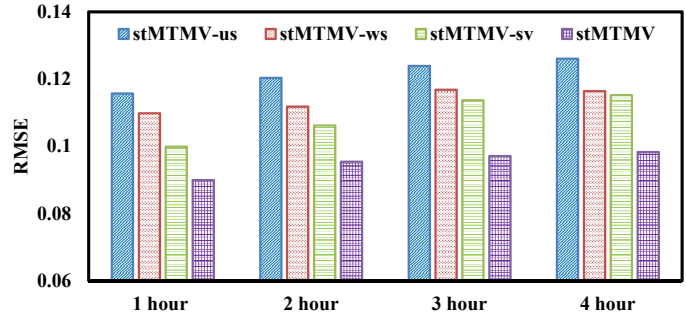


Fig. 10. Comparative predictive performance illustration of each component in stMTMV with respect to 1-4 hours.

## 6.5 Evaluation on Views

To demonstrate the descriptiveness of each view, we compared our stMTMV model over the following combinations.

- **t-view**: Only temporal view (*t-view*) is used.
- **s-view**: Only spatial-view (*s-view*) is used.
- **st-view-na**: Both spatio-temporal views are used, but there is no *s-t* view alignment within each station.
- **st-view**: Both spatio-temporal view are used and the *s-t* view alignment is employed for each station.

The results are presented in Figure 11. From this figure, we observe that: 1) the combinations of spatial and temporal views outperform each individual one. This observation reveals that the more views fed to our model, the better the performance that can be achieved. 2) the *st-view* outperforms *st-view-na*, which implies that aggregating information from spatial and temporal views can achieve better performance than concatenating them together. This also verifies that the heterogeneous information distributed across spatial and temporal views is usually complementary rather than conflicting, and appropriate aggregation of these can provide a better way to capture each station's characteristics comprehensively, and consequently boost the performance.

## 6.6 Predictive Performance among Stations

To investigate the strength of stMTMV model among stations, we compared our stMTMV model against other baselines and its variants among different stations. Figure 12 illustrates the predictive performance comparison for stations. It can be seen that stMTMV performs no worse than other baselines as well as its variants for most of the stations. Figure 12(a) presents the comparison of stMTMV over other baselines, where stMTMV is significantly better than other models in at least 6 of the stations. In addition, the results in Figure 12(b) demonstrate that stMTMV stably outperforms stMTMV-us in all stations, which justifies the importance of spatial correlations among stations and further verifies that the tasks are mutually correlated and

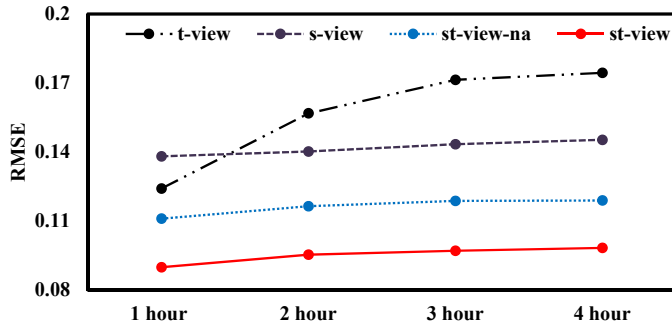


Fig. 11. Comparative predictive performance illustration of stMTMV over view combinations with respect to 1-4 hours.

the spatial autocorrelation plays an important role in the co-prediction tasks. Interestingly, stMTMV is slightly worse than stMTMV-*sv* in one or two stations in Figure 12(b), it might due to the inconsistency between spatial and temporal views in those stations. Moreover, as demonstrated in Figure 12(c), stMTMV consistently outperforms other view combinations, which, in turn, further provides reasonable justification for the heterogeneous information aggregation.

## 6.7 Predictive Accuracy

Furthermore, we also investigated the predictive accuracy of our stMTMV model with single-task learning and single-view learning methods. The results are demonstrated in Table 4. The predictive accuracy *Acc* is computed as follows:

$$Acc = 1 - \frac{1}{M} \sum_{l=1}^M \frac{\|y_l - \hat{y}_l\|_1}{\|y_l\|_1}, \quad (16)$$

where  $y_l$  is the ground truth and  $\hat{y}_l$  is the predicted value for the station  $l$ , and  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm. The predictive accuracy *Acc* is a similar metric as mean absolute percentage error (MAPE). From this table, we can have the observations as follows: 1) The predictive accuracy of all methods exhibits a decrease trend for the next 1-4 hours. However, as compared to other methods, the decrease in our method is quite small in magnitude, which demonstrates the robustness of our approach. This further verifies the assumption that the co-prediction among all the stations can boost the performance. 2) Our approach achieves higher accuracy as compared to either single-task or single-view methods, which justifies the effectiveness of combining the multi-task with multi-view algorithms together. 3) Since accuracy *Acc* is a similar metric as mean absolute percentage error (MAPE). Our model can achieve ~85% accuracy, which is acceptable for hourly-based water quality prediction.

TABLE 4  
Predictive accuracy comparison of our model with single-task and single-view methods over next 1-4 hours.

Predictive Accuracy	1 hour	2 hour	3 hour	4 hour
LR	80.1%	76.6%	71.4%	68.2%
<i>t-view</i>	83.2%	77.3%	70.6%	63.2%
<i>s-view</i>	76.0%	73.4%	71.6%	70.8%
stMTMV	<b>85.9%</b>	<b>85.0%</b>	<b>84.3%</b>	<b>84.1%</b>

## 6.8 Water Quality Predictions

Figure 13 depicts the predictive results of our method over the next 1 hour against the ground truth in Shenzhen from October 2012 to November 2012. In general, our model is very accurate in tracing the ground truth curves (including sudden changes) of the water quality in Shenzhen City, which demonstrates the effectiveness of our approach.

## 7 RELATED WORK

### 7.1 Classical Model-based Approaches

It is worth mentioning that several research efforts have been dedicated to model-based approaches for urban water quality prediction [8] [3]. The main idea behind this kind of approaches is to utilize the first-order or higher-order kinetics to model the chlorine decay along the water distribution system. However, as the mechanisms of the chlorine decay is quite complicated and comprise of reactions with bulk fluid, pipe and natural evaporation, the accurate mathematical modelling of chlorine decay along the water supply system is a tough problem whose solution is not yet absolutely mastered [7]. Moreover, the developed decay model requires extensive human labors to perform model calibration with pipe networks, and it depends heavily on the pipe internal surface materials, temperatures, network structure, which makes it difficult to extend to other cities' water distribution systems. Compared to model-based approaches, data-driven based approaches demonstrate their advantages in both flexible and extendibility in many other urban applications [34] [5] [35], such as urban sensing [36], urban air forecast [34], destination prediction [4], and traffic prediction [37]. However, to the best of our knowledge, the literature on urban water quality prediction from the data-driven perspective is relatively sparse.

### 7.2 Classical Data-driven based Approaches

Several studies in the environmental science have been tried to analyze the water quality problems via data-driven based approaches, and those studies covers a range of topics, from the physical process analysis in the river basin, to the analysis of concurrent input and output time series [38]. Those data-driven approaches adopted in these studies can fall into the following three major categories: *k*-Nearest Neighbors (*k*-NN), Artificial Neural Network models (ANN) and Support Vector Machine models (SVM). For example, the work introduced by Karlsson *et al.* [39] addressed the classical rainfall-runoff forecasting problem by *k*-NN algorithm, and demonstrated promising results. Artificial Neural Network (ANN) have been widely employed to solve a wide variety of tasks, and can achieve good results [38]. For instance, the work introduced by Kalin [30] predicted the water quality indexes in watersheds through ANN and demonstrated its advantages over other numerical prediction methods. Support Vector Machines (SVMs) are typical supervised learning models that analyze data used for classification and regression [40]. For instance, Liong *et al.* [40] addressed the issue of flood forecasting using Support Vector Regression (SVR) which is an extension of SVM.

However, none of these approaches is applied into urban scenarios, which is quite different from our applications. Moreover, those existing approaches process the data from a single source, and can hardly integrate the data from different sources. Thus, their applications in the urban scenarios are restricted.

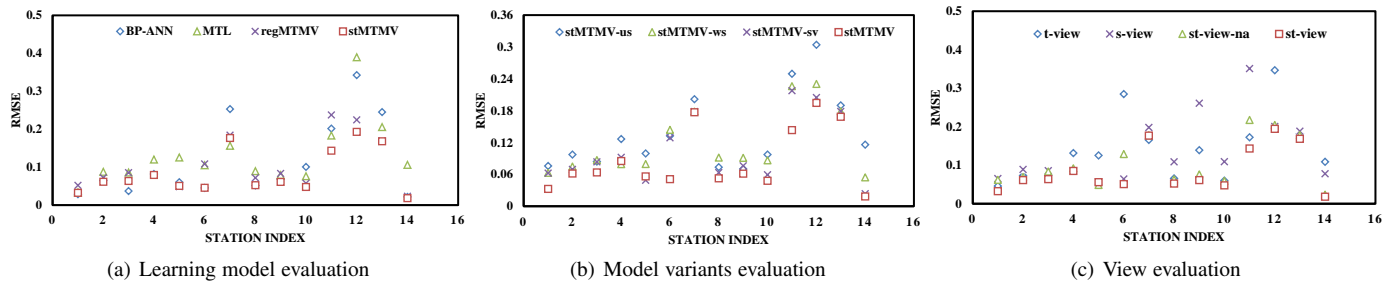


Fig. 12. Comparative predictive performance illustration of stMTMV over other predictive models, its model variants, and the view combinations among stations, respectively.

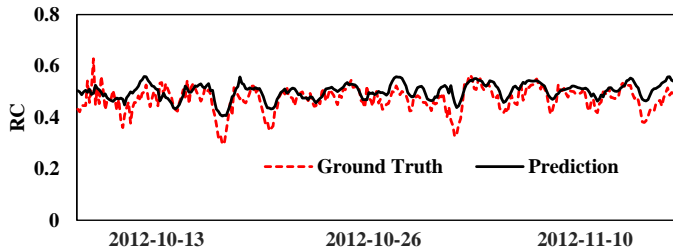


Fig. 13. Predictions of stMTMV against the ground truths.

### 7.3 Graph-based Approaches

Graphs are employed extensively as a ubiquitous data structure within various research areas, e.g., recommender systems, social networks. Many problems in those areas can be modeled as graphs, where interactions between different nodes can be captured by edges [41] [42]. With the great success of deep learning approaches in many fields, e.g., computer vision, natural language processing, there is increasing interest in extending deep learning approaches to graph-related applications [43] [44] [45] [46] [47]. Graph Convolutional Network (GCN) approaches [46] [47] [44] are an emerging set of approaches that can handle spatial relationships by generalizing the operation of 'convolution' from traditional data (grids or images) to graph structured data. For example, Geng *et al.* [44] proposed a novel spatiotemporal multi-graph convolution network (ST-MGCN) for ride-hailing demand forecasting, and the evaluations on two real-world datasets shows consistent improvement of 10% over other baselines. Another interesting work introduced by Li *et al.* [45] proposed a Diffusion Convolutional Recurrent Neural Network (DCRNN) for traffic forecasting by modeling the spatial and temporal dependency of the traffic flow as a diffusion process on a directed graph, and observe improvement of 12% – 15% over other baselines on two real-world large scale road network traffic datasets. There are many other interesting work about this topic, e.g., Gated Attention Networks (GaAN) [43], Spatio-Temporal Graph Convolutional Networks (STGCN) [47], data-efficient Graph Convolutional Network (GCN) [46], and those methods achieve promising results over other graph-based approaches.

However, Graph Convolutional Network (GCN) approaches may not be appropriate for our applications. First, the number of pipe intersections (nodes) in the urban water system is huge and much more than the number of nodes in normal GCN settings, which makes it difficult to adopt the existing GCN approaches. Moreover, the water pipe network is much more complicated than

the normal graph, for example, there are many different paths connect with two different nodes, which is not similar as only one edge between two nodes in GCN. Therefore, in this work, we consider a different learning paradigm for modeling the spatio-temporal dependency of the urban water quality.

### 7.4 Multi-task Multi-view Learning Approaches

Multi-task learning is a learning paradigm that jointly learns multiple related tasks and has demonstrated its advantages in many urban applications, such as transportation and event forecasting [17] [5]. In particular, it is more effective in handling those with insufficient training samples [18] [19] [48]. However, most of the existing approaches only explore the task relatedness, but ignore the consistency information among different views within a task.

Multi-view learning has been proposed to leverage the information from diverse domains or from various feature extractors, and combining the heterogeneous properties from different views can better characterize objects and achieve promising performance [4]. Nevertheless, existing multi-view learning approaches discard the label information from other related tasks, which usually leads to suboptimal performance.

Thus, multi-view multi-task learning is proposed to explore both task relatedness and view relatedness simultaneously within a learning framework [49] [16] [15] [1]. For example, He *et al.* [49] proposed a graph-based iterative framework (*GraM*<sup>2</sup>) for multi-view multi-task learning and obtained impressive results in text categorization applications. However, as far as we know, the literature on spatio-temporal based multi-task multi-view learning is relatively sparse. To the best of our knowledge, our approach is the first work on spatio-temporal based multi-task multi-view learning, which can incorporate spatio-temporal heterogeneities via a multi-task multi-view learning framework and is able to applied to other spatio-temporal based applications.

## 8 CONCLUSION AND FUTURE WORK

This paper presents a novel data-driven approach to forecast the water quality of a station by fusing multiple sources of urban data. We evaluate our approach based on Shenzhen's water quality and various urban data. The experimental results demonstrate the effectiveness and efficiency of our approach. Specifically, our approach outperforms the traditional RC decay model [3] and other classical time series predictive models (ARMA, Kalman) in terms of RMSE metric. Meanwhile, as our approach consists of two components, each of the components demonstrates its effectiveness through extensive



experiments and analysis. In particular, the first component is the influential factors identification, which explores the factors that affect the urban water quality via extensive experiments and analysis in Section 3 and 4. The second one is a spatio-temporal multi-view multi-task learning (stMTMV) framework that consists of multi-view learning and multi-task learning. The experiments have shown that stMTMV has a predictive accuracy of around 85% for forecasting next 1-4 hours, which outperforms the single-task methods (LR) by approximately 11% and the single-view methods (*t-view* and *s-view*) by approximately 11% and 12%, respectively. The code has been released at: <https://www.microsoft.com/en-us/research/publication/urban-water-quality-prediction-based-multi-task-multi-view-learning-2/>

In future, we plan to deal with the water quality inference problems in the urban water distribution systems through a limited number of water quality monitor stations.

## ACKNOWLEDGMENTS

This work was supported by the China National Basic Research Program (973 Program, No. 2015CB352400), NSFC under grant U1401258, NSCF under grant No. 61572488. This research was also supported in part by A\*STAR SERC PSF under grant 152120008 from the National University of Singapore. We also thank Yipeng Wu for sourcing the data in this study.

## REFERENCES

- [1] Y. Liu, Y. Zheng, Y. Liang, S. Liu, and D. S. Rosenblum, "Urban water quality prediction based on multi-task multi-view learning," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2016.
- [2] W. H. Organization, *Guidelines for drinking-water quality*, 2004, vol. 3.
- [3] L. A. Rossman, R. M. Clark, and W. M. Grayman, "Modeling chlorine residuals in drinking-water distribution systems," *Journal of environmental engineering*, vol. 120, no. 4, pp. 803–820, 1994.
- [4] Y. Zheng, "Methodologies for cross-domain data fusion: An overview," *IEEE Transactions on Big Data*, vol. 1, no. 1, pp. 16–34, 2015.
- [5] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 3, pp. 38:1–38:55, 2014.
- [6] H. Cohen, "Free chlorine testing," <http://www.cdc.gov/safewater/chlorine-residual-testing.html>, 2014, accessed on 5 August 2016.
- [7] P. Castro and M. Neves, "Chlorine decay in water distribution systems case study—lousada network," *Electronic Journal of Environmental, Agricultural and Food Chemistry*, vol. 2, no. 2, pp. 261–266, 2003.
- [8] L. A. Rossman and P. F. Boulous, "Numerical methods for modeling water quality in distribution systems: A comparison," *Journal of Water Resources planning and management*, vol. 122, no. 2, pp. 137–146, 1996.
- [9] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 2003, pp. 2–11.
- [10] G. Luo, K. Yi, S.-W. Cheng, Z. Li, W. Fan, C. He, and Y. Mu, "Piecewise linear approximation of streaming time series data with max-error guarantees," in *Proceedings of the IEEE International Conference on Data Engineering*, 2015, pp. 173–184.
- [11] E. O. Brigham and E. O. Brigham, *The fast Fourier transform*. Prentice-Hall Englewood Cliffs, NJ, 1974, vol. 7.
- [12] C. S. Burrus, R. A. Gopinath, and H. Guo, "Introduction to wavelets and wavelet transforms: a primer," 1997.
- [13] M. W. LeChevallier, T. Evans, and R. J. Seidler, "Effect of turbidity on chlorination efficiency and bacterial persistence in drinking water," *Applied and environmental microbiology*, vol. 42, no. 1, pp. 159–167, 1981.
- [14] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [15] Y. Liu, L. Zhang, L. Nie, Y. Yan, and D. S. Rosenblum, "Fortune teller: Predicting your career path," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 201–207.
- [16] J. Zhang and J. Huan, "Inductive multi-task learning with multiple view data," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining*, 2012, pp. 543–551.
- [17] L. Zhao, Q. Sun, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Multi-task learning for spatio-temporal event forecasting," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 1503–1512.
- [18] Y. Liu, L. Nie, L. Han, L. Zhang, and D. S. Rosenblum, "Action2activity: Recognizing complex activities from sensor data," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2015, pp. 1617–1623.
- [19] Y. Zhang and D.-Y. Yeung, "A convex formulation for learning task relationships in multi-task learning," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2010, pp. 733–742.
- [20] A. M. Madlala, S. Bissoo, S. Singh, and L. Christov, "Xylanase-induced reduction of chlorine dioxide consumption during elemental chlorine-free bleaching of different pulp types," *Biotechnology Letters*, vol. 23, no. 5, pp. 345–351, 2001.
- [21] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of machine learning research*, vol. 7, no. Nov, pp. 2399–2434, 2006.
- [22] R. K. Ando and T. Zhang, "Learning on graph with laplacian regularization," *Advances in neural information processing systems*, vol. 19, p. 25, 2007.
- [23] P.-N. T. Jianpeng Xu, Lifeng Luo and J. Zhou, "Gspartan: a geospatio-temporal multi-task learning framework for multi-location," in *SIAM International Conference on Data Mining*, 2016.
- [24] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic geography*, pp. 234–240, 1970.
- [25] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [26] J. Zhou, J. Chen, and J. Ye, "Malsar: Multi-task learning via structural regularization," *Arizona State University*, 2011.
- [27] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.
- [28] F. J. Fabozzi, S. M. Focardi, S. T. Rachev, and B. G. Arshanapalli, "Autoregressive moving average models," *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*, pp. 171–190, 2014.
- [29] A. C. Harvey, *Forecasting, structural time series models and the Kalman filter*, 1990.
- [30] L. Kalin, S. Isik *et al.*, "Prediction of water quality parameters using an artificial neural networks model," in *Proceedings of the World Environmental and Water Resources Congress*, 2010, pp. 3145–3153.
- [31] G. A. Seber and A. J. Lee, *Linear regression analysis*, 2012, vol. 936.
- [32] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [33] T. Evgeniou and M. Pontil, "Regularized multi-task learning," in *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 109–117.
- [34] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine-grained air quality based on big data," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 2267–2276.
- [35] Y. Liang, K. Ouyang, L. Jing, S. Ruan, Y. Liu, J. Zhang, D. S. Rosenblum, and Y. Zheng, "Urbanfm: Inferring fine-grained urban flows," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.
- [36] F. Zhang, N. J. Yuan, D. Wilkie, Y. Zheng, and X. Xie, "Sensing the pulse of urban refueling behavior: A perspective from taxi mobility," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 3, p. 37, 2015.
- [37] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 25–34.
- [38] S. Haykin and N. Network, "A comprehensive foundation," *Neural Networks*, vol. 2, no. 2004, 2004.
- [39] M. Karlsson and S. Yakowitz, "Nearest-neighbor methods for nonparametric rainfall-runoff forecasting," *Water Resources Research*, vol. 23, no. 7, pp. 1300–1308, 1987.
- [40] P.-S. Yu, S.-T. Chen, and I.-F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3, pp. 704–716, 2006.

- [41] Y. Liu, H. Li, A. Garcia-Duran, M. Niepert, D. Onoro-Rubio, and D. S. Rosenblum, "Mmkg: Multi-modal knowledge graphs," in *European Semantic Web Conference*, 2019, pp. 459–474.
- [42] H. Li, Y. Liu, N. Mamoulis, and D. S. Rosenblum, "Translation-based sequential recommendation for complex users on sparse data," *IEEE Transactions on Knowledge and Data Engineering*, 2019.
- [43] J. Zhang, X. Shi, J. Xie, H. Ma, I. King, and D.-Y. Yeung, "Gaan: Gated attention networks for learning on large and spatiotemporal graphs," *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pp. 339–349, 2018.
- [44] X. Geng, Y. Li, L. Wang, L. Zhang, Q. Yang, J. Ye, and Y. Liu, "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.
- [45] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.
- [46] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph convolutional neural networks for web-scale recommender systems," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 974–983.
- [47] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.
- [48] D. Preoŕiuc-Pietro, Y. Liu, D. Hopkins, and L. Ungar, "Beyond binary labels: political ideology prediction of twitter users," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2017, pp. 729–740.
- [49] J. He and R. Lawrence, "A graph-based framework for multi-task multi-view learning," in *Proceedings of the International Conference on Machine Learning*, 2011, pp. 25–32.



**Ye Liu** Dr. Ye Liu received his Ph.D degree from National University of Singapore. Before that he received the M.Sc. degree from Peking University. He is currently a Research Fellow in the National University of Singapore. His research interests lie mainly in the areas of urban computing, ubiquitous computing, knowledge graph, data mining, machine learning, artificial intelligence and their applications in human activity analysis, social network analysis, knowledge graph, urban computing.



**Yuxuan Liang** Yuxuan Liang received the B.Sc. degree in computer science from Xidian University. He is currently a research assistant at National University of Singapore now. His research interests include deep learning, urban computing and spatio-temporal data mining.



**Kun Ouyang** Kun Ouyang obtained his B.E. degree in the Internet of Things, in Wuhan University in 2015. He is currently a PhD candidate in Computer Science at National University of Singapore. His research interests include human mobility analytics, spatial-temporal data mining and deep learning.



**Shuming Liu** Dr. Shuming Liu obtained his PhD in Environmental Engineering in 2004 and now is an associate professor at Tsinghua University. He is now the vice dean of School of Environment (SOE) and Director of Smart Water Research Centre, SOE, Tsinghua University. His research interests mainly include leakage control, early warning system, water quality management, data driven modeling and optimal management of water distribution system etc. He has published more than 120 peer-reviewed papers.



**David S. Rosenblum** David S. Rosenblum is Professor of Computer Science and Dean of the School of Computing at the National University of Singapore (NUS). He received his PhD from Stanford University and joined NUS in April 2011, after holding positions as Research Scientist at AT&T Bell Laboratories (Murray Hill); Associate Professor at the University of California, Irvine; Principal Architect and Chief Technology Officer of PreCache (a technology startup funded by Sony Music); and Professor of Software Systems at University College London. His research interests span many problems in software engineering, distributed systems and ubiquitous computing. He is a Fellow of the ACM and IEEE and a Senior Member of the Singapore Computer Society.



**Yu Zheng** Dr. Yu Zheng a Vice President of JD Group and Chief Data Scientist at JD Digits. He also leads the Urban Computing Business Unit as the president and serves as the director of JD Intelligent City Research. Before joining JD Digits, he was a senior research manager at Microsoft Research. His research interests include big data analytics, spatio-temporal data mining, and ubiquitous computing. He currently serves as the Editor-in-Chief of ACM Transactions on Intelligent Systems and Technology and an Editorial Board

Member of IEEE Transactions on Big Data. Zheng has served as chair on over 10 prestigious international conferences, e.g. ICDE 2014 and ICDM 2015.