International Journal of Engineering & Technology, 7 (2.28) (2018) 79-84

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Estimation of Daily Life Time Series Data Affected by Rainfall

Teruhisa Hochin[1]*, Hiroki Nomiya[2]

[1]Professor, Faculty of Information and Human Sciences, Kyoto Institute of Technology, JAPAN
[2]Associate Professor, Faculty of Information and Human Sciences, Kyoto Institute of Technology, JAPAN
**Corresponding author E-mail: hochin@kit.ac.jp*

Abstract

The amount of sewage flow, which is one of daily life data, was estimated for their efficient management. The amounts of flow of a typical day were tried to be adjusted to those of a day. The values for the adjustment were tried to be estimated by using the multiple regression analysis. This method is applied to the estimation of the ammonia nitrogen concentration, which is the major factor of the quality of sewage flow. The estimation results show that this method is applicable to the estimation of the ammonia nitrogen concentration, and that the amount of rainfall is dominant in estimating the ammonia nitrogen concentration.

Keywords: *Ammonia nitrogen concentration; Daily life data; Estimation; Multiple regression analysis; Rainfall*

1. Introduction

There are many time series data of daily life. Traffic volume, electricity usage, and air temperature change are such data. Sewage flow is one of those data. Sewage flow must be considered to in the sewage treatment.
Sewage treatment is the process of removing contaminants from wastewater [1]. Basic sewer systems were used even in ancient Mesopotamia, the Indus Valley civilization, Ancient Crete and Greece, and the Roman empire [1]. Sewage treatment is inevitable for modern society and is indispensable for the healthy life of people.
Energy requirement for sewage management is very high. A lot of energy is consumed. As efficient control of sewage management results in saving energy, it is considered to be very important. It is said that clarifiers and mechanized secondary treatment are more efficient under uniform flow conditions [1]. For making the flow uniform, the estimation of the load of sewage flow is required. If the load of sewage flow can be known in advance, the plan of driving the plant can be fit to the optimum one. The load of sewage flow is based on the amounts and the quality of sewage flow. They are strongly required to be estimated.
The amounts of sewage flow have been estimated for their efficient management [2], [3]. Here, non-rainy days were treated. Only the data good enough to process were treated. The day having good data is called a normal day. Normal days are divided into typical days and irregular ones. The latter have outliers, and the former do not. Typical days are classified into regular days and non-regular ones based on the existence of outliers. The amounts of flow on regular days are almost those of the typical day. The amounts of flow of a typical day were tried to be adjusted to those of a non-regular day. The values for the adjustment were tried to be estimated by using the multiple regression analysis. It was shown that these values could be estimated based on the temperature and the amount of rainfall [3]. As the load of sewage flow is based on the amounts and the quality of sewage flow as described before, the quality of sewage flow is required to be estimated.

This paper tries to estimate the ammonia nitrogen concentration of sewage, which is one of the major factors indicating the quality of sewage flow. The estimation method used is the one estimating the amounts of sewage flow. The estimation results show that the ammonia nitrogen concentration can be estimated. This shows the estimation method is effective to the estimation of the ammonia nitrogen concentration. The estimation results also show that the amount of rainfall is dominant in estimating the ammonia nitrogen concentration.
The remainder of the paper is as follows: Section 2 describes sewage management. Section 3 briefly reviews the method of estimating the amounts of sewage flow. Section 4 estimates the quality of sewage flow. Section 5 gives some considerations. Finally, Section 6 concludes the paper.

2. Sewage treatment

Sewage treatment is the process of removing contaminants from wastewater, primarily from household sewage [1]. Physical, chemical, and biological processes are included in order to remove contaminants and to produce treated wastewater which is environmentally safe.

2.1. Process steps

The end of the sewage treatment is to produce an effluent that will not harm the surrounding environment when it is discharged to the environment. It is strongly required to prevent pollution from being released into the surrounding environment, and to contaminate it. Sewage treatment generally has three stages: primary, secondary, and tertiary treatments.

Primary treatment: The sewage is held in a quiescent basin for a while. Heavy solids sink to the bottom of the basin, while oil, grease, and lighter solids float to the surface. The sunk and floating materials are removed. The remaining liquid may be discharged.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Secondary treatment: Dissolved and suspended biological matter is removed. Indigenous, water-borne micro-organisms in a managed habitat are typically used. Removing the micro-organisms from the treated water is required before the discharge or the process of the tertiary treatment.

Tertiary treatment: The water processed through the primary and the secondary treatments is processed in order to be disposed into a highly sensitive or fragile ecosystem, which includes estuaries, low-flow rivers, coral reefs, and so on. The water treated is chemically or physically disinfected before discharged into a stream, river, bay, lagoon, or wetland.

2.2. Energy requirement

Around thirty percent of the annual operating costs is usually required for energy for the conventional sewage treatment plant [1]. The energy requirements depend on the type of treatment process as well as wastewater load. The wastewater load is based on the amounts and the quality of sewage flow.
In the conventional secondary treatment process, most of the electricity is used for aeration, pumping systems, and the equipment for the dewatering and drying of sewage sludge. Advanced wastewater treatment plants require more energy than the plants only achieving primary and secondary treatments.
It is said that clarifiers and mechanized secondary treatment are more efficient under uniform flow conditions [1]. In order to make the flow uniform, the estimation of the wastewater load is required. If the load of sewage flow can be known in advance, the plan of driving the plant can be fit to the optimum one.

2.3. Quality of sewage

Three major parameters for the quality of sewage are Ammonia Nitrogen, Biochemical Oxygen Demand, and Suspended Solids [4].

Ammonia nitrogen: Ammonia nitrogen is the nitrogen included in the ammonia (NH4) part of the sewage. It mainly comes from urine. Nitrogen is a major plant nutrient. Excessive amount of nitrogen in the effluent causes explosive plant growth and algae blooms in rivers, lakes, and ponds. This results in the death of fish and other aquatic life due to oxygen starvation because the plants remove the oxygen from the water during the night time. The removal of nitrogen from sewage effluent is required for the treatment of sewage and wastewaters.

Biochemical oxygen demand: The biochemical oxygen demand (BOD) is the measure of the amount of oxygen required by micro-organisms in the sewage to break down its organic material. The more BOD the sewage has, the more pollution it has. The sewage having high BOD causes severe depletion of the oxygen in the water if such sewage is discharged into a river.

Suspended solids: Suspended solids are the tiny particles floating around in the sewage liquor. These do not easily settle out over time. Suspended solids are reduced by settlement treatment stages and by biological treatment.

3. Estimation of amounts of sewage flow

The amounts of sewage flow have been estimated [2], [3]. Days are categorized based on the outliers of the amounts of sewage flow. The day type is also considered. The typical amounts of sewage flow of days are obtained by taking their mean values. The amounts of every day is estimated from the typical ones by using multiple regression analysis. These are described in the following.

3.1. Categorization of sewage flow

Normal days: Only non-rainy days are treated in order to make the analysis easy. Here, the day having the rainfall amount less than 5 mm per hour is considered as a non-rainy day.
The amount of flow may suddenly become very large. The day including such a sudden huge amount of flow is not treated in this paper. As the average amount of flow is around 2,000 as described later, the day including the flow amount of 3,000 or more is not used. That is, the days having the flow amount of less than 3,000 are used. These days are called normal days. We obtained 186 normal days. The amounts of flow are averaged. These averaged amounts of flow are called the tentative typical amounts of flow.

Typical and irregular days: Normal days are divided into two categories: typical days and irregular ones. Irregular days are the days including outliers from the normal days. The other days are typical days.
For selecting typical days, we adopted three criteria: (1) A typical day has less than 20,000 of the sum of the positive differences of flow amounts, (2) It has less than 2,000 of the sum of the absolute values of their negative differences, and (3) Correlation to the tentative typical amounts of flow is more than 70 percent. The amounts of flow of these days are averaged to obtain the typical amounts of flow.

Regular and non-regular days: Moreover, the typical days are divided into regular days and non-regular days. For selecting regular days, we also adopted three criteria. The first two criteria are the same as those of (1) and (2) adopted in selecting typical days. The last one is the criterion that correlation to the typical amounts of flow is more than 70 percent. It is slightly different from that of (3) in selecting typical days.

3.2. Day type

As our daily life differs depending on the week day or the weekend, days are categorized into the week day, Saturday, Sunday, and the national holiday. These are called types of days. Typical, regular, and non-regular days are decided in considering the type of day.

3.3. Estimating amounts of sewage flow

3.3.1. Adjusting data of non-regular days

As the amounts of flow of the regular days are almost the same as those of the typical day, the amounts of flow of the typical day can be used as those of the regular days.
As it is not true for those of the non-regular days, those are tried to be obtained by adjusting the amounts of flow of the typical day to those of a non-regular one. Here, we consider the scaling and the offset. The transformation of the amount of flow of the typical day d(t) to that of a non-regular day d'(t) at the time t is represented by (1).

$$d'(t) = m * d(t) + ofst \quad (1)$$

Here, m is for scaling, and *ofst* is an offset value.
By using the least-squares method, we could decide the values of m and *ofst* for each day. For the days not considering the type of day, the mean correlation of the amounts of the original non-regular days and those of the adjusted ones is 0.95. For the days considering the type of day, the mean correlation of the amounts of the original non-regular days and those of the adjusted ones is 0.96.

Table 1: Explanatory Variables

Type	*m*	*ofst*
Considering no day type	Square root of the amount of rainfall of the previous day	Square root of the amount of rainfall of the previous day
	Square root of the mean temperature	Square root of the mean temperature
	Day type code	Day type code
Considering day type	Amount of rainfall of the previous day	Amount of rainfall of the previous day
	Second power of the highest temperature	Highest temperature

3.3.2. Estimating data of non-regular days

Although the values of *m* and *ofst* could precisely be obtained, these values can be calculated only after the amounts of flow of a day are known. As the aim is to estimate the amounts of flow on a future day, it is not enough to only obtain the values of *m* and *ofst*. The values of *m* and *ofst* are estimated by using the regression analysis.

Estimation without considering day types: The value *m* can be estimated better by using the square root of the amount of the rainfall of the previous day, the square root of the mean temperature, and the day type code as explanatory variables. The explanatory variables adopted here are listed in Table 1. The multiple correlation coefficient of the estimation was improved to 0.62.
The value of *ofst* can be estimated better by using the same values as those used in estimating the value of m as explanatory variables. The multiple correlation coefficient of the estimation was improved to 0.64.

Estimation considering day types: The value *m* can be estimated better by using the amount of the rainfall of the previous day and the second power of the highest temperature as explanatory variables. The multiple correlation coefficient of the estimation was improved to 0.62.
The value of *ofst* can be estimated by using the amount of the rainfall of the previous day and the highest temperature as explanatory variables. The multiple correlation coefficient of the estimation was 0.60.

4. Estimation of quality of sewage flow

4.1. Data used

We use sewage flow data collected in Oct. 1st, 2015 to Oct. 31st, 2016. The average ammonia nitrogen concentration of sewage flow is hourly measured. There are twenty-four data points a day. The day having a missing value is excluded from the day treated here.
The weather data including weather, temperatures, humidity, and rainfall amount are additionally used. These are used as explanatory variables of the regression analysis.

4.2. Approach

This paper follows the procedure in handling the amounts of sewage flow [2], [3] as described in Section 3. First, the days are divided into several groups according to the ammonia nitrogen concentration. These are normal, typical, irregular, regular, and non-regular days. Next, the ammonia nitrogen concentrations of non-regular days are adjusted from the mean ammonia nitrogen concentration of the regular days. Finally, the parameters used in adjusting the ammonia nitrogen concentrations are estimated from various information including weather one.

4.3. Categorization of Days

4.3.1. Normal days

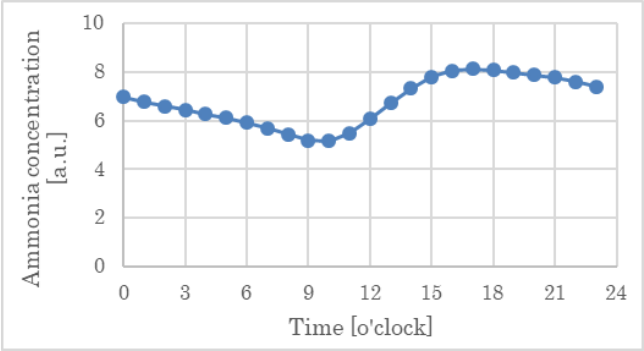
As the average amount of ammonia nitrogen concentration is between 5 and 10 [a.u.] as described later, the day including the ammonia nitrogen concentration between 3 and 12 [a.u.] is treated as a normal day. We obtained 234 normal days. 36 days are not normal ones, which are called abnormal days. The amounts of ammonia nitrogen concentration of the normal days are averaged. These mean amounts of flow are called the tentative typical amounts of ammonia nitrogen concentration.

4.3.2. Typical and irregular days

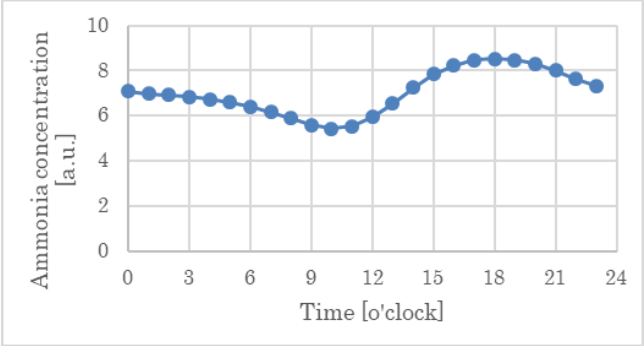
Normal days are divided into two categories: typical days and irregular ones. Irregular days are the days including outliers from the normal days. The other days are typical days.
For selecting typical days, we adopted three criteria: (1) A typical day has less than 2σ of the positive differences of ammonia nitrogen concentration, (2) It has less than 2σ of the sum of the absolute values of their negative differences, and (3) Correlation to the tentative typical amounts of flow is more than 70 percent, where σ is the standard deviation of ammonia nitrogen concentrations. We obtained 197 typical days. The remaining 37 days are irregular days. The ammonia nitrogen concentrations of these days are averaged to obtain the typical ammonia nitrogen concentration.

4.3.3. Regular and non-regular days

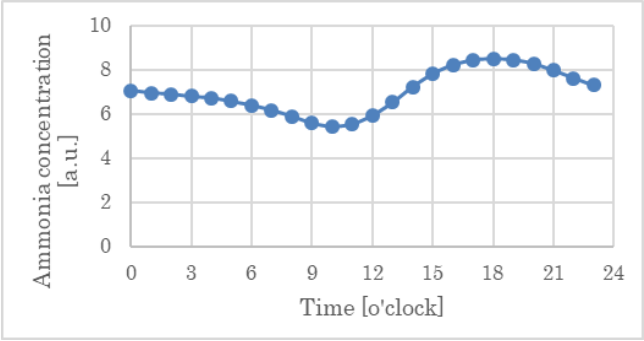
The typical days are divided into regular days and non-regular days. For selecting regular days, we also adopted three criteria. These are (1) A regular day has less than σ of the positive differences of ammonia nitrogen concentration, (2) It has less than σ of the sum of the absolute values of their negative differences, and (3) Correlation to the typical ammonia nitrogen concentration is more than 70 percent, where σ is the standard deviation of ammonia nitrogen concentrations. 155 days are selected as regular days. The remaining 42 days are selected as non-regular days.
Typical ammonia nitrogen concentration transitions considering day types are shown in Fig. 1. As the ammonia nitrogen concentration of the regular days are almost the same as those of the typical day, the ammonia nitrogen concentration of the typical day can be used as those of the regular days.
The ammonia nitrogen concentration transitions of the weekday, Saturday, Sunday, and the holiday are almost the same as shown in Fig. 1. The lowest (highest, respectively) concentration is around ten (eighteen) o'clock. The peak concentration of the weekday is earlier than those of the other types of days. The degree of ammonia nitrogen concentration of the weekend and the



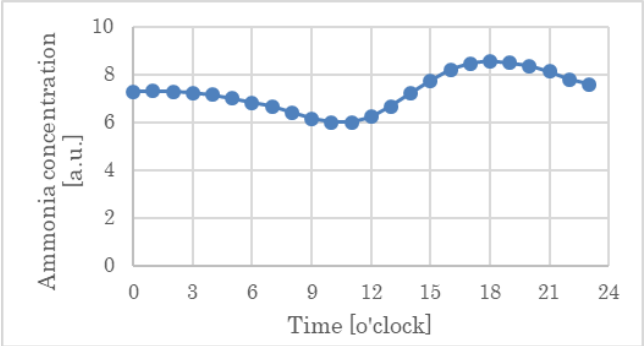
(a) Week day



(b) Saturday



(c) Sunday



(d) National Holiday

Fig. 1: Typical ammonia nitrogen concentration transition of typical days

holiday is higher than that of the weekday. Especially, the ammonia nitrogen concentration of the holiday is higher than the weekend as well as the weekday.

4.3. Adjusting non-regular data

As the ammonia nitrogen concentrations of the non-regular days are not the same as those of the typical day, those are tried to be obtained by adjusting the ammonia nitrogen concentration of the typical day to them. Here, we consider the scaling and the offset as in the adjustment of the amounts of sewage flow. Transformation is represented with (1).

By using the least-squares method, we could decide the values of *m* and *ofst* for each day.

4.4. Estimating non-regular data

The values of *m* and *ofst* are estimated through the regression analysis as in estimating the amounts of sewage flow [2], [3].

The value *m* can be estimated by using the square root of the rainfall of the previous day as an explanatory variable. The multiple correlation coefficient is 0.56.

The value of *ofst* can be estimated by using the square root of the amount of rainfall of the previous day, and the square root of the sum of the rainfalls of the previous day, and the day before the previous day, and the two days before it. The multiple correlation coefficient is 0.69.

The explanatory variables adopted for the estimations are shown in Table 2. These are of the amounts of rainfalls.

Estimation result having the minimum error as well as the original, the typical, and the adjusted ammonia nitrogen concentration transitions are shown in Fig. 2. Although the original and the typical transitions are quite different, the adjusted transition shows that the typical one was successfully adjusted to the original one. The estimated transition shows that the original transition is also successfully estimated from the typical one.

Estimation result having average error as well as the original, the typical, and the adjusted ammonia nitrogen concentration transitions are shown in Fig. 3. The tendency of the original, the typical, and the adjusted transitions is almost the same as the previous one: The typical transition is successfully adjusted to the original one. The estimated transition is a little bit different from the original one.

Table 2: Explanatory Variables

Parameter	*Explanatory variables*
m	Square root of the amount of rainfall of the previous day
ofst	Square root of the amount of rainfall of the previous day Square root of the sum of the rainfalls of the previous day, and the day before the previous day, and the two days before it

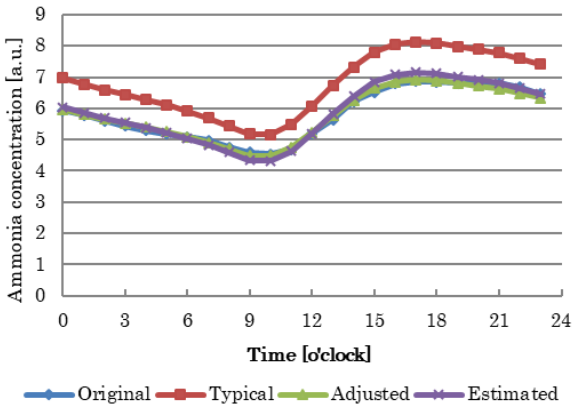


Fig. 2: Estimation of ammonia nitrogen concentration transition having the minimum error.

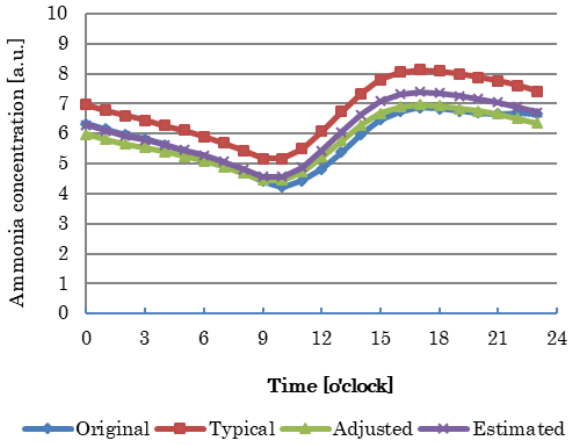


Fig. 3: Estimation of ammonia nitrogen concentration transition having the average error.

5. Considerations

The estimation results show that the ammonia nitrogen concentration could successfully be estimated. The estimation method is the one used in estimating the amounts of sewage flow. This shows that this method can be used in estimating the ammonia nitrogen concentration.

This paper showed the estimation considering the day type. When the day type is not considered, few non-regular days could be obtained even though there are the days whose transitions of ammonia nitrogen concentration are quite different from those of the typical one. This kind of irregularity may disappear because of the irregularity due to the differences of day types.

The explanatory variables adopted are of the amounts of rainfalls as shown in Table 2. We used other kinds of values of weather information including the mean, the highest, and the lowest temperatures, and the degree of humidity as the explanatory variables in the estimation. The values adopted were, however, those of the rainfall. In the estimation of the amounts of sewage flow, the amount of rainfall is also used as shown in Table 1. It may be said that the rainfall strongly affects the sewage flow. The temperature as well as the amounts of rainfall is used in estimating the amounts of sewage flow. This may be the difference between the estimations of the amounts of sewage flow and its ammonia nitrogen concentration.

Representing time series data in the frequency domain may be effective [5], [6]. Major information of the transition of the ammonia nitrogen concentration may be obtained by focusing the low frequency components of the ammonia nitrogen concentration transition. This kind of representation may derive some important information from the ammonia nitrogen concentration transition.

Fundamental movements are derived from the movements of a body [7], [8], [9], which is a kind of time series data. Signal processing methods are used in capturing them. By using these methods in representing the ammonia nitrogen concentration transition, some pattern commonly appearing in the concentration transition may be clarified. This paper did not treat the days having outliers. The signal processing methods enable us to treat these irregular days.

Li *et al.* analysed road traffic data to find out variables closely related to fatal accidents [10]. They investigated the relationship between fatal rate and attributes including weather, surface condition, and light condition. Weather affects a lot of things including fatal accidents and sewage flow. Chatterji and Sridhar analysed airspace system delay [11]. They showed that weather is the most important causal factor for delays followed by equipment and runway delays. Dyer and Baier reported that rain affected the estimation model of the field work conditions based on soil tractability [12]. The day rain and the night rain affect the estimation error of the model. Coops *et al.* estimated the potential forest productivity based on the monthly weather records [13]. Weather is inseparable from our daily life.

Roush *et al.* tried to predict gross domestic product by using autoregressive models [14]. The model predicts a current state from previous values in time-series data. As the sewage flow is a kind of time-series, predicting the amounts and the ammonia nitrogen concentration of the sewage flow through time-series predicting models may be effective. Gunu *et al.* examined a lot of predictive models including various regressions to predict collage graduation rate [15]. We should try to examine a lot of models to improve the accuracy of the estimation.

6. Conclusion

This paper estimated the ammonia nitrogen concentration of the sewage as the quality estimation of the sewage. The estimation method estimating the amounts of the sewage flow [2], [3] is used. The estimation result showed that this method could be applied to the estimation of the ammonia nitrogen concentration of sewage. It also showed that the amount of the rainfall is dominant in the sewage flow. As the amounts of the sewage flow could be already estimated [2], [3], the wastewater workload can be estimated because the wastewater workload is based on the amounts and the quality of sewage flow. By estimating the wastewater workload, the operation plan of the sewage plant could be formed a priori. When the amount and the ammonia nitrogen concentration of the sewage flow is as expected, the extra electric power for operating the sewage plant is not needed. This results in the reduction of the electric power.

The estimation accuracy of the parameter m is not high. Improving the estimation accuracy is included in future work. Only the data of non-rainy typical days were treated. Treating irregular days and rainy ones is in future work. The adjustment adopted in the estimation of the amounts of sewage flow does not consider the shift of the peak time of the amounts of flow. Considering the shift of the peak is also included in future work.

References

[1] Wikipedia, Sewage Treatment, https://en.wikipedia.org/wiki/Sewage_treatment (accessed 2017-03-19).

[2] T. Hochin and H. Nomiya, "On the Estimation of Time Series Data of Daily Life," *Proceedings of 15th IEEE/ACIS International Conference on Software Engineering Research, Management and Applications (SERA 2017)*, (2017), pp. 351-356

[3] T. Hochin, H. Nomiya, "Improvement and Evaluation of Estimation of Time Series Data of Daily Life," *International Journal of Networked and Distributed Computing*, Vol. 5, No. 4, (2017), pp. 233-242

[4] Water Technology Engineering, "Wastewater Treatment Explained," https://www.wte-ltd.co.uk/wastewater_treatment_explained html (accessed 2018-01-19)

[5] T. Hochin, Y. Yamauchi, H. Nakanishi, M. Kojima, H. Nomiya, "Indexing of plasma waveforms for accelerating search and retrieval of their subsequences," *Fusion Engineering and Design*, Vol. 85, No. 5, (2010), pp. 649-654

[6] R. Agrawal, C. Faloutsos, and A. N. Swami, "Efficient similarity search in sequence databases," *Proceedings of the 4th Int'l Conf. on Foundations of Data Organization and Algorithms (FODO'93)*, (1993), pp. 69-84

[7] T. Hochin and H. Nomiya, "Analysis of Puases Toward Transmitting Traditional Skills," *Proceedings of 14th ACIS Int'l Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2013)*, (2013), pp. 414-419

[8] T. Hochin and H. Nomiya, "Deriving Fundamental Movements Based on Pauses for Transmitting Traditional Skills," *Proceedings of 1st ACIS International Symposium on Applied Computers & Information Technology (ACIT 2013)*, (2013), pp. 343-348

[9] T. Hochin and H. Nomiya, "Considerations on Pauses and Fundamental Movements for Transmitting Traditional Skills," *Proceedings of 15th ACIS Int'l Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2014)*, (2014), pp.287-292

[10] L. Li, S. Shrestha, G. Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques," *Proceedings of 15th IEEE/ACIS International Conference on Software Engineering Research, Management and Applications (SERA 2017)*, (2017), pp. 363-370

[11] G. B. Chatterji, B. Sridhar, "National Airspace System Delay Estimation Using Weather Weighted Traffic Counts," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, (2015), pp. 1-17

[12] J. A. Dyer, W. Baier, "Weather-based estimation of field workdays in fall," *Canadian Agricultural Engineering*, Vol. 21, No. 2, (1979), pp. 119-122

[13] N. C. Coops, R. H. Warning, J. J. Landsberg, "Estimation of potential forest productivity across the Oregon transect using satellite data and monthly weather records," *International Journal of Remote Sensing*, Vol. 22, No. 18, (2011), pp. 3797-3812

[14] J. Roush, K. Siopes, G. Hu, "Predicting Gross Domestic Product Using Autoregressive Models," *Proceedings of 15th IEEE/ACIS International Conference on Software Engineering Research, Management and Applications (SERA 2017)*, (2017), pp. 317-322

[15] E. A. Gunu, C. Lee, W. K. Gyasi, R. M. Roe, "Modern Predictive Models for Modeling the College Graduation Rates," *Proceedings of 15th IEEE/ACIS International Conference on Software Engineering Research, Management and Applications (SERA 2017)*, (2017), pp. 39-45