Predictive Analytics for Detecting Sensor Failure Using Autoregressive Integrated Moving Average Model

Karthick Thiyagarajan, Sarath Kodagoda, Linh Van Nguyen
Centre for Autonomous Systems
University of Technology Sydney
Sydney, Australia
Karthick.Thiyagarajan@uts.edu.au, Sarath.Kodagoda@uts.edu.au, VanLinh.Nguyen@uts.edu.au

***Abstract*— Sensors play a vital role in monitoring the important parameters of critical infrastructure. Failure of such sensors causes destabilization to the entire system. In this regard, this paper proposes a predictive analytics solution for detecting the failure of a sensor that measures surface temperature from an urban sewer. The proposed approach incorporates a forecasting technique based on the past time series of sparse data using an autoregressive integrated moving average (ARIMA) model. Based on the 95% forecast interval and continuity of faulty data, a criterion was set to detect anomalies and to issue a warning for sensor failure. The forecasted and faulty data were assumed Gaussian distributed. By using the probability density of the distribution, the mean and variance were computed for faulty data to examine the abnormality in the variance value of each day to detect the sensor failure. The experimental results on the sewer temperature data are appealing.**

Keywords—ARIMA model; anomalies detection; forecasting; sensor failure detection; surface temperature; time series; sewers.

I. Introduction

Predictive analytics algorithms integrated with cutting-edge sensor technology is an imperative component of smart monitoring systems mainly due to the reason that it enables the practitioners to foresee the future trends and more squarely it answers “What is likely to happen?” based on the historical or past sparse data. Recently, the relevance of predictive analytics to tackle real-world problems that are emerging from the sophisticated mainstream utilities is a paradigm of “How the advanced data analytics has taken the ascendency in delivering better solutions?”. In this context, this paper provides a predictive analytics solution for detecting the sensor failure based upon the real-time operational data sourced from an urban sewer system.

In recent years, urban sewerage infrastructures transporting wastewater suffers from Hydrogen sulphide (H_2S) induced concrete corrosion primarily owing to the physical and chemical activities of biogenic sulphuric acid producing bacteria that lives on the sewer walls [1, 2]. Due to the unavailability of technology to measure liable bacteria that causes corrosion, researchers have identified temperature as one of the three main proxy parameters to predict the rate of concrete corrosion [3-5]. For the aforementioned reason, a custom-made sensor system was installed in sewer pipes for monitoring the temperature variations on the concrete surface.

The environmental conditions of the sewer systems make hostile for engineers to enter the sewer pipes [6, 7]. For this reason, a monitoring station was constructed adjacent to the sensor deployment site, where the accessibility to electrical power is very limited. This circumstance leads to supplying power to the sensors by batteries and due to the power consumption of the monitoring system, the batteries were swapped once in a week with the recharged ones. The sewer environments are harsh and notoriously capable of spoiling sensors. Given the sensor units cannot be physically monitored all the time and they only can be remotely monitored through communication infrastructure a mechanism to monitor and estimate sensor failure is an important aspect. With the increase of the number of the sensor units in the infrastructure, the need of an automatic sensor failure detection system becomes paramount.

The sensors can behave differently over the time. There can be temporary spurious data due to change in sensor characteristics. Those are intermittent and possible to occur randomly. They should not be considered as sensor failures but need to be filtered out as spurious data. On the other hand, sensor failures will give rise to unexpected data over a time period. It is not always possible to wait for a long period before the sensor failure detection because fast failure detections will give rise to appropriate on-time intervene strategies.

There are large numbers of literature available for sensing and monitoring of various systems. The existing literature contains numerous forecast models such as Exponential Smoothing (ETS) [8], TBATS model [9] and Autoregressive and Moving Average (ARMA) model [10]. ARMA model is a well-known method to investigate the time series data [11]. Based on ARMA model, [12] formulated Autoregressive and Integrated Moving Average (ARIMA) model for predicting linear time series data by transforming the non-stationary data into stationary data before forecasting [13].

ARIMA based forecasting models are widely used to forecast weather patterns [14], stock market [11] and electricity loads [15, 16]. Also, ARIMA models were used in anomalies detection for specific applications in industrial sectors like

978-1-5090-6161-7/17/$31.00

refinery [17] and network security [18]. However, there is no literature reported the use of ARIMA model for forecasting and sensor failure detection in the wastewater industry. In this paper, we propose a predictive analytics solution using ARIMA model to forecast the time series of sparse data based on the past measurements and thereby detect the event of sensor failure occurrence. The spare data is the surface temperature measurements that were collected at uniform time intervals of one hour from a sewer monitoring station. The forecasting results of ARIMA model were compared with the results other two popular forecast models for examining the prediction performance. By using 95% confidence interval of forecasted data, anomalies were detected and a criterion was set to issue a warning for sensor failure. Gaussian distribution was implemented for the faulty dataset and forecasted dataset. Based on the mean and variance of the probability density function, the sensor failure was detected.

The remainder of this paper is structured as follows. The methodology for the formulation of predictive analytics using ARIMA model is described in Section II. The Section III presents and evaluates the results of ARIMA model with discussions. Finally, the conclusion is reported in Section IV with directions for future work.

II. Formulation of Autoregressive Integrated Moving Average Model

This section describes the methodology for the formulation of ARIMA model to forecast the time series sparse data obtained from the sewer systems. The time series of sparse data is given by X_t which has real numbers as in (1)

$$X_t = \{X_1, X_2, X_3, ...\} \text{ for all } t > 1 \tag{1}$$

The Autoregressive model of order p is denoted as *AR(p)*. In *AR(p)* model, the value of the future variable is assumed to be a linear combination of past value p of the variable with a constant and white noise [19, 20]. Mathematically, *AR(p)* model is defined as in (2)

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{2}$$

where c is a constant, φ_i is the parameter of the model, and i $(1 \le i \le p)$ is an integer and ε_t is the white noise error term. The constant term c in (2) can be omitted for simplicity. Given the time series X_t, the lag operator L can be defined as in (3)

$$L^i X_t = X_{t-i} \tag{3}$$

substituting (3) in (2) to write *AR(p)* in terms of a lag operator as in (4)

$$X_t = \sum_{i=1}^{p} \varphi_i L^i X_t + \varepsilon_t \tag{4}$$

re-arranging the expression (4) in terms of ε_t and given in (5)

$$\varepsilon_t = X_t - \sum_{i=1}^{p} \varphi_i L^i X_t \tag{5}$$

Therefore, *AR(p)* model in terms of lag operator L can be defined as in (6)

$$\varepsilon_t = \left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t \tag{6}$$

The Moving-average model of order q is denoted as *MA(q)*. In *MA(q)* model, the past errors are used as an explanatory variable [20, 21] . Mathematically, the *MA(q)* model is defined as in (7)

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{7}$$

where μ is the expectation of X_t often assumed to be zero, θ_i is the parameter of the model, and i $(1 \le i \le q)$ is an integer and ε_t , ε_{t-i} are the white noise error terms. For the time series of ε_t , the lag operator L can be defined as in (8)

$$L^i \varepsilon_t = \varepsilon_{t-i} \tag{8}$$

substituting (8) in (7) to write *MA(q)* in terms of a lag operator in (9)

$$X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i L^i \varepsilon_t \tag{9}$$

upon the simplification of expression in (9), X_t for *MA(q)* is given in (10)

$$X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{10}$$

The ARMA model combines the *AR(p)* autoregressive terms of (2) and *MA(q)* moving-average terms of (7) to form *ARMA(p,q)* [20, 21]. Mathematically, the *ARMA(p,q)* model is defined as in (11)

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{11}$$

where φ_i are the parameters of the autoregressive part of the model *AR(p)* and θ_i are the parameters of the moving-average part of the model *MA(q)*.

Since the *ARMA(p,q)* models are manipulated using the lag operator, *ARMA(p,q)* model combines the *AR(p)* autoregressive terms of (6) and *MA(q)* moving-average terms of (10) to form *ARMA(p,q)* expression as in (12)

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{12}$$

As mentioned in Section I, *ARMA(p,q)* model can only be applied for stationary time series data. However, for the application reported in this work, we need a model to describe the non-stationary time series. For this reason, ARIMA model was chosen for analysing the sparse data, which converts non-stationary time series to stationary by applying finite differencing of data points [21, 22].

The notation *ARIMA(p,d,q)* refers to ARIMA model, which is relatively same as *ARMA(p,q)*. But, the time series of sparse data X_t given in *ARMA(p,q)* in (12) is replaced by (13) to form *ARIMA(p,d,q)* in (14)

$$X_t = (1-L)^d \; X_t \tag{13}$$

Finally, the *ARIMA(p,d,q)* which is a generalization of *ARMA(p,q)* is mathematically defined as in (14)

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d \; X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{14}$$

where φ_i is the i[th] parameter and p is the order of the autoregressive part of the model *AR(p)*, θ_i is the i[th] parameter and q is the order the moving-average part of the model *MA(q)*, d is the order of integrated part of the model.

The forecast interval, which is also known as confidence interval comprises of upper and lower bounds between which the predicted values of sparse data is expected to lay within the prescribed probability [23]. The forecast interval for *ARIMA(p,d,q)* model is given by (15).

$$\hat{y}_{T+h|T} \pm z_{\alpha/2}\sqrt{v_{T+h|T}} \tag{15}$$

where $v_{T+h|T}$ is the variance of $y_{T+h} \mid y_1, \ldots\ldots, y_T$, $z_{\alpha/2} = 1.96$ for 95% forecast interval (confidence interval). Irrespective of parameters and order of *ARIMA(p,d,q)*, for the integer h=1, the variance is expressed as in (16)

$$v_{T+h|T} = \hat{\sigma}^2 \tag{16}$$

In *ARIMA(p,d,q)*, the value d=1 for most cases as it controls the differencing. Based on the p, d and q integers values, the *ARIMA(p,d,q)* model can be reduced to different special cases as follows:

(1) When $d=0$, i.e. *ARIMA(p,0,q)*, the model reduces to *ARMA(p,q)*.

(2) When $p=0$ and $d=0$, i.e., *ARIMA(0,0,q)*, the model reduces to *MA(q)*.

(3) When $q=0$ and $d=0$, i.e., *ARIMA(p,0,0)*, the model reduces to *AR(p)*.

(4) When $p=0$ and $q=0$, i.e., *ARIMA(0,1,0)*, the model is known as Random Walk model [24].

(5) When $p=0$, $d=0$ and $q=0$, i.e., *ARIMA(0,0,0)*, the model is known as white noise model.

In the reported work, the *ARIMA(p,d,q)* model was implemented in R programming language using the package 'Forecast' [25].

III. Results and Discussions

In this section, the performance of sparse data prediction using ARIMA model will be examined and compared with other models such as ETS and TBATS. The sparse data from the sewer systems were recorded for 30 days from 04[th] November 2016 to 03[rd] December 2016. The total length of data points is DP_{total} = 720 and the data points of the first 24 days containing DP_{training} = 576 will be used as the training data. Remaining data points $DP_{testing}$ = 144 will be used for evaluating the performance of spare data prediction. Fig. 1 shows the plot of input training data.

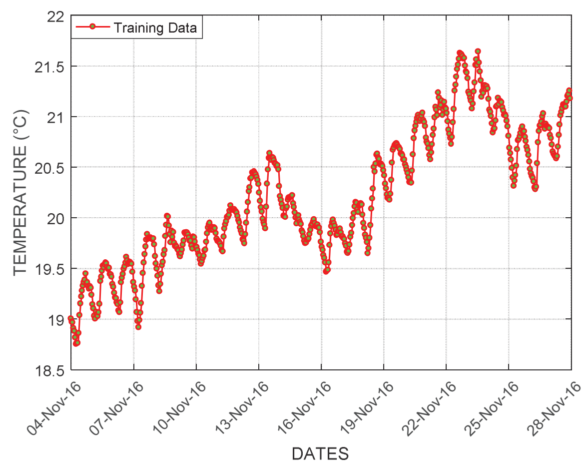

Fig. 1. Plot of sparse data for training the forecast model

Fig. 2 illustrates the time series forecasting of sparse data from 00:00 hours of 28[th] November 2016 to 23:00 hours of 03[rd] December 2016 using ARIMA, TBATS and ETS models.

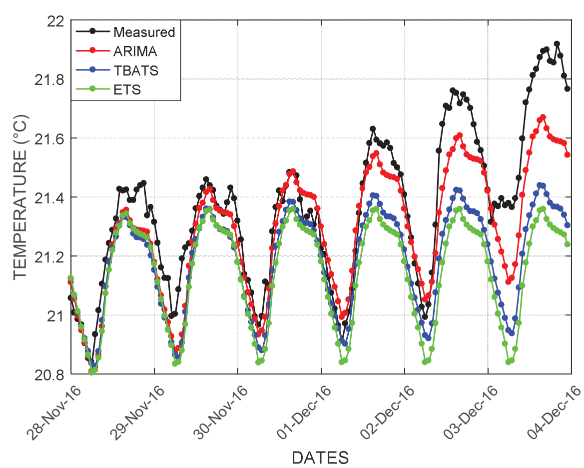

Fig. 2. Forecasted sparse data from the ARIMA, TBATS and ETS model

To evaluate the performance of the forecast models and to choose a suitable model for the sparse data from sewers, statistical performance metrics such as Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Performance Deviation (MAPD) and Root Mean Square Error (RMSE) were used to compute the forecast errors.

Table I lists the statistical performance metrics for the different models.

TABLE I. STATISTICAL PERFORMANCE METRICS

Statistical Metrics	Time Series Model		
	ARIMA	***ETS***	***TBATS***
MAE (ºC)	0.0962	0.2194	0.1801
MPE (%)	0.3491	1.0087	0.8247
MAPD (%)	0.4481	1.0203	0.8374
RMSE (ºC)	0.1228	0.2722	0.2263

It can be observed from the Table I that the MAE and RMSE of TBATS model were smaller than those of ETS model. Thence, the prediction performance of TBATS was better than ETS. But, the MAE and RMSE of TBATS model were higher than those of ARIMA model. So, based on MAE and RMSE the prediction performance of ARIMA model was better than the other two models. Like aforementioned performance metrics, the MPE and MPAD of TBATS model were smaller than those of ETS model and higher than those of ARIMA model. In summary, based on the four different performance metrics for forecasting the sparse data sourced from sewers, ARIMA model had better performance and it is used as a forecasting model to detect the sensor failure. Fig. 3 displays the 95% confidence interval for the forecasted data of ARIMA model. It can be observed from Fig. 3 that the forecasted data and the measured data lay within the upper and lower bounds of the confidence interval.

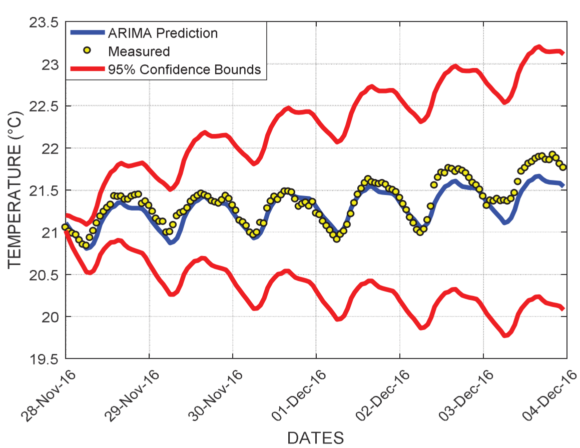

Fig. 3. Forecasted data using ARIMA model with 95% confidence interval

During the sensor testing phase at UTS-CAS Robotics Laboratory, the sensor was malfunctioned and a set of faulty data was logged. Those faulty data was manually added to the time series of sparse data from sewers as a testing dataset for the ARIMA model to detect anomalies and sensor failure. Fig. 4 illustrates the testing dataset to detect anomalies.

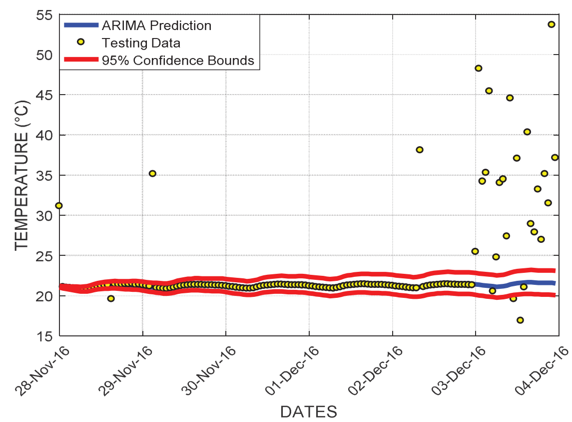

Fig. 4. Testing dataset to detect anomalies using ARIMA model.

A total of DP_{faulty} = 28 were used as a faulty data out of testing data $DP_{testing}$ = 144. The data points that are lying outside of the 95% confidence interval of the forecasted data using ARIMA model were regarded as anomalies. Based on the upper and lower bounds of the model, 92.8% of the anomalies were detected and 7.2 % of faulty data were lying within the 95% confidence interval. A criterion was set to issue the sensor failure warning. For each faulty data lying outside the 95% confidence interval, weight $w = 0.2$ value was given. For five continuous faulty data, the sensor failure value $SF = w*n$, where n is the number of occurrence of faulty data continuously. If the value of $SF >= 1$, then the sensor failure warning is reported. For the testing dataset shown in Fig. 4, the sensor failure warning was reported at the following times in the Date/Month/Year Hours-Minutes-Seconds format: (i) 03/12/2016 04:00:13 (ii) 03/12/2016 11:00:14 and (iii) 03/12/2016 20:00:12. It can be observed from the Fig. 3 and Fig. 4 that the 95% confidence interval tends to widen as the time progress. This phenomenon could be reduced when the real-time data is supplied to the predictive analytics model after each measurement.

Gaussian distribution, which is also known as normal distribution is used for representing the random variables, where the distribution of the variables are not known in advance [26]. The Gaussian distribute was implemented on the faulty data and forecasted data of ARIMA model by using the probability density function of the normal distribution in [17].

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2 \pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{17}$$

where μ is the mean, σ is the standard deviation and σ^2 is the variance of the Gaussian distribution. The μ is a real number and the variance is always $\sigma^2 > 0$. The Gaussian distribution is noted in terms of $N(\mu, \sigma^2)$. Fig. 5 displays the Gaussian distribution of the faulty data and the forecasted data from the ARIMA model from 28-Nov-16 to 03-Dec-16.

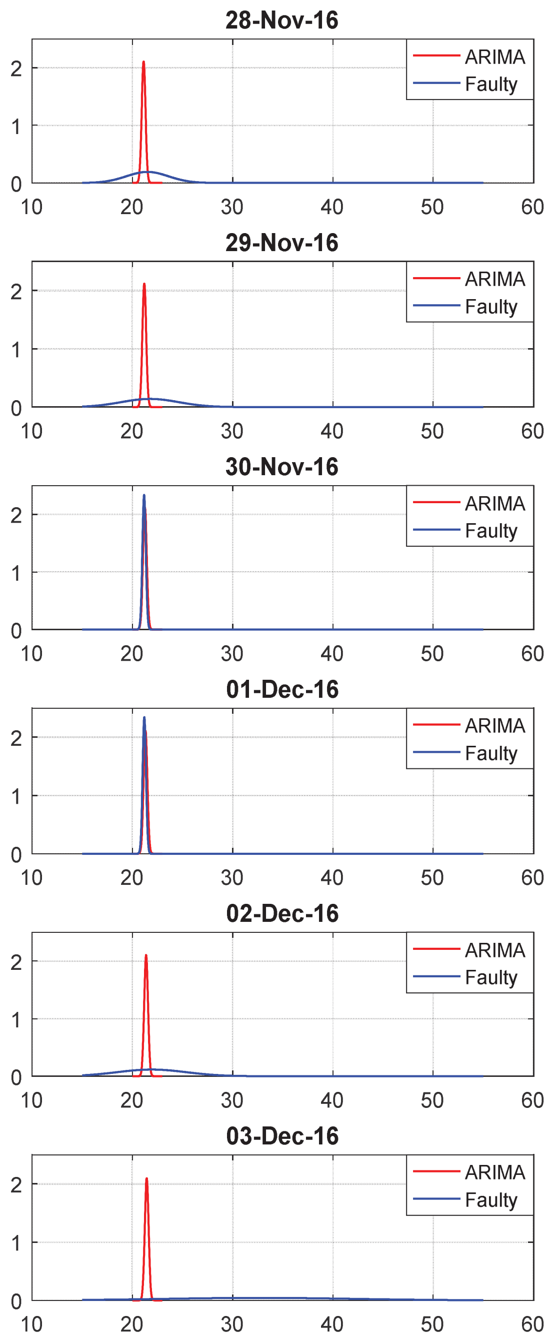

Fig. 5. Gaussian distribution of faulty and forecasted data.

Table II presents the mean and variance of faulty data and forecasted data from ARIMA model using probability density function of the Gaussian distribution. The mean of forecasted data and faulty data were denoted as μ_m and μ_f, and the variance of forecasted and faulty data was denoted as $\sigma^2{}_m$ and $\sigma^2{}_f$.

TABLE II. STATISTICAL ANALYSIS FOR SENSOR FAILURE

Dates	Mean and Variance from Gaussian Distribution			
	μ_m	μ_f	$\sigma^2{}_m$	$\sigma^2{}_f$
28-Nov-16	21.1301	21.5057	0.0358	4.3514
29-Nov-16	21.1969	21.7596	0.0354	8.1305
30-Nov-16	21.2569	21.1771	0.0356	0.0292
01-Dec-16	21.3170	21.1984	0.3858	0.0290
02-Dec-16	21.3773	21.9302	0.0360	11.6694
03-Dec-16	21.4377	32.6495	0.0362	87.9919

The $N(\mu, \sigma^2)$ of training sparse data from 00:00 hours of 04-Nov-16 to 23:00 hours of 27-Nov-16 is N (20.1984, 0.4322).

For the dates 28-Nov-16, 29-Nov-16 and 02-Dec-16, the Gaussian distributions of faulty data are N (21.5057, 4.3514), N (21.7596, 8.1305), N (21.9302, 11.6694) and the distributions of forecasted data are N (21.1301, 0.0358), N (21.1969, 0.0354), N (21.3773, 0.0360). Both the Gaussian distributions did not possess similar trend as the previously mentioned dates. This effect is due to the presence of faulty data in the testing dataset. However, the $N(\mu, \sigma^2)$ of faulty data is not significantly higher than the $N(\mu, \sigma^2)$ of forecasted data and $N(\mu, \sigma^2)$ of training data.

For the dates 30-Nov-16 and 01-Dec-16, the Gaussian distribution of faulty data are N (21.1771, 0.0292), N (21.1984, 0.0290) and the distribution of forecasted data are N (21.2569, 0.0356), N (21.3170, 0.3858). In the case of those dates, both the distributions tend to be similar, which can be observed from Fig 5. The reason is due to the fact that there is an absence of faulty data in the testing dataset for those dates.

For the date 03-Dec-2016, the Gaussian distribution of faulty data is N (32.6495, 87.9919) and the distribution of forecasted data is N (21.4377, 0.0362). In contrast to the Gaussian distribution of training data and forecasted data, this date has significantly higher mean and variance than any other days. According to the Fig. 5, the distribution of the faulty data of the 03-Dec-2016 is almost flat compared to the distribution of the forecasted data for that particular date. The abnormal mean and variance on 03-Dec-16 implies the possible sensor failure on that day. Henceforth, a predictive analytics solution to detect sensor failure implemented based on the forecasting ARIMA model and the Gaussian distribution for the sparse data from sewers shows convincing results.

IV. CONCLUSIONS

In this paper, we have proposed a predictive analytics solution for the time series of sparse data obtained from the urban sewer system. The forecasting performance of ARIMA model that uses sparse data was examined and compared with ETS and TBATS model. MAE, MPE, MAPD and RMSE were used as a performance criterion to evaluate the models discussed in this paper. Based on the statistical metrics, the prediction performance of ARIMA model was better than the ETS and TBATS model. Also, the prediction performance of TBATS model was better than the ETS model. Overall, this paper suggests ARIMA model for forecasting the surface temperature measurements obtained from the sewer systems. Using ARIMA model, this paper proposes an approach to detect the anomalies and issuing sensor failure warning based on a criterion set to the 95% confidence interval. Gaussian distribution was implemented for the faulty data and the forecasted data from ARIMA model. Based on the probability density of the distribution, the abnormality in the mean and variance was examined for each day of the faulty dataset to detect the possible sensor failure.

As future work, the authors are intended to develop a comprehensive framework for integrating the sparse data and predictive analytics platform, where the system gets updated with each measurement to form a real-time predictive analytics platform for the application motivated in this paper and the results will be published in due course.

ACKNOWLEDGMENT

This publication is an outcome from the Data Analytics on Sewers Project funded by Sydney Water Corporation, Melbourne Water Corporation, Water Corporation (WA) and South Australian Water Corporation. The research participants are Data61-CSIRO, University of Technology Sydney (UTS) and University of Newcastle (UoN). Karthick Thiyagarajan receives UTS President Scholarship and International Research Scholarship from The University of Technology Sydney, Australia.

REFERENCES

[1] G. Jiang, J. Keller, P. L. Bond, and Z. Yuan, "Predicting concrete corrosion of sewers using artificial neural network," *Water research,* vol. 92, pp. 52-60, 2016.

[2] G. Jiang, J. Keller, and P. L. Bond, "Determining the long-term effects of H 2 S concentration, relative humidity and air temperature on concrete sewer corrosion," *Water research,* vol. 65, pp. 157-169, 2014.

[3] K. Thiyagarajan, S. Kodagoda, and J. K. Alvarez, "An instrumentation system for smart monitoring of surface temperature," in *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2016, pp. 1-6.

[4] T. Wells and R. Melchers, "Modelling concrete deterioration in sewers using theory and field observations," *Cement and Concrete Research,* vol. 77, pp. 82-96, 2015.

[5] K. Thiyagarajan, S. Kodagoda, and N. Ulapane, "Data-driven machine learning approach for predicting volumetric moisture content of concrete using resistance sensor measurements," in *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, 2016, pp. 1288-1293.

[6] K. Thiyagarajan and S. Kodagoda, "SMART monitoring of surface temperature and moisture content using multisensory data fusion," in *Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), 2015 IEEE 7th International Conference on*, 2015, pp. 222-227.

[7] K. Thiyagarajan and S. Kodagoda, "Analytical Model and Data-driven Approach for Concrete Moisture Prediction," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, 2016, p. 1.

[8] R. J. Hyndman, M. Akram, and B. C. Archibald, "The admissible parameter space for exponential smoothing models," *Annals of the Institute of Statistical Mathematics,* vol. 60, pp. 407-426, 2008.

[9] A. M. De Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American Statistical Association,* vol. 106, pp. 1513-1527, 2011.

[10] P. Whitle, *Hypothesis testing in time series analysis* vol. 4: Almqvist & Wiksells, 1951.

[11] P. Mondal, L. Shit, and S. Goswami, "Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices," *International Journal of Computer Science, Engineering and Applications,* vol. 4, p. 13, 2014.

[12] G. E. Box and G. C. Tiao, "Intervention analysis with applications to economic and environmental problems," *Journal of the American Statistical association,* vol. 70, pp. 70-79, 1975.

[13] T. Kuzin and T. Borovicka, "Early Failure Detection for Predictive Maintenance of Sensor Parts."

[14] A. Geetha and G. M. Nasira, "Time series modeling and forecasting: Tropical cyclone prediction using ARIMA model," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 3080-3086.

[15] P. Saatwong and S. Suwankawin, "Short-term electricity load forecasting for Building Energy Management System," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2016 13th International Conference on*, 2016, pp. 1-6.

[16] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE transactions on power systems,* vol. 18, pp. 1014-1020, 2003.

[17] M. R. Saybani, T. Y. Wah, A. Amini, and S. R. A. S. Yazdi, "Anomaly detection and prediction of sensors faults in a refinery using data mining techniques and fuzzy logic," *Scientific Research and Essays,* vol. 6, pp. 5685-5695, 2011.

[18] H. Z. Moayedi and M. A. Masnadi-Shirazi, "Arima model for network traffic prediction and anomaly detection," in *2008 International Symposium on Information Technology*, 2008, pp. 1-6.

[19] J. Kihoro, R. Otieno, and C. Wafula, "Seasonal time series forecasting: A comparative study of ARIMA and ANN models," *AJST,* vol. 5, 2004.

[20] K. W. Hipel and A. I. McLeod, *Time series modelling of water resources and environmental systems* vol. 45: Elsevier, 1994.

[21] R. Adhikari and R. Agrawal, "An introductory study on time series modeling and forecasting," *arXiv preprint arXiv:1302.6613,* 2013.

[22] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*: John Wiley & Sons, 2015.

[23] J. S. Armstrong, *Principles of forecasting: a handbook for researchers and practitioners* vol. 30: Springer Science & Business Media, 2001.

[24] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing,* vol. 50, pp. 159-175, 2003.

[25] R. Hyndman, "Forecasting functions for time series and linear models, R package version 6.1," ed, 2015.

[26] G. Casella and R. L. Berger, *Statistical inference* vol. 2: Duxbury Pacific Grove, CA, 2002.