Estimating Missing Data in Wireless Sensor Network Through Spatial-Temporal Correlation

Walid Atwa[1*], Abdulwahab Ali Almazroi[2], Eman A. Aldhahr[3], Nourah Fahad Janbi[4]
Department of Information Technology-College of Computing and Information Technology at Khulais,
University of Jeddah, Jeddah, Saudi Arabia[1, 2, 4]
Department of Computer Science and Artificial Intelligence-College of Computer Sciences and Engineering,
University of Jeddah, Jeddah, Saudi Arabia[3]

***Abstract*—Wireless sensor networks consist of a set of smart sensors with limited memory and wireless communication capabilities. These sensors get data from the environment and send them to an application center. However, data loss has happened due to the characteristics of sensors, which negatively affect the accuracy of applications. To solve this problem, we need to estimate the missing data for applications that depend on accurate data collecting. In this study, we present an algorithm that uses the most significant historical data to estimate the missing data based on spatial and temporal correlations. In the proposed algorithm, we combine the spatial correlation by using data from the closest sensor based on the missing pattern and the temporal correlation by referring to the closest data prior to the missing instance. The experimental results demonstrate that the proposed algorithm lowers estimation errors when compared to current algorithms for a variety of missing data patterns.**

Keywords—Wireless sensor networks; missing data estimation; spatial correlation; temporal correlation

I. Introduction

In recent years, WSNs have generated a lot of interest for a variety of purposes [1]. WSNs have become particularly important for physical exploration, as they are used in difficult-to-reach places such as aquatic settings, active calderas, and deep forests [2-5]. These networks are employed in real-world settings to gather environmental data, which is then used in computer systems. A WSN is composed of many smart, inexpensive sensors that have less processing and storage power than conventional sensors. These sensors can detect, measure, and gather information from their environment, process the information in an initial step, and then send it to users [2].

Most applications in WSN suffer from missing data because of sensor limitations and environmental factors. Thus, the analysis tools cannot be applied well without recovering these missing data accurately. If the missing data is ignored, the original information and data resources are lost. Thus, reduce the accuracy and reliability of analysis results. This problem can be effectively solved by data estimation algorithms that interpolate missing data accurately and efficiently. So, handling the missing data is particularly one of the most important challenges in data management for different applications of WSN.

Numerous estimation methods have been used to address the issue of missing data, which can be categorized into three main groups: spatial correlation, temporal correlation, and spatial-temporal correlation.

Spatial correlation methods rely on the assumption that data points closer in spatial distribution have a greater influence on interpolating missing data. These methods treat each sensor as independent and calculate weight based on the gap between the missing data and surrounding sensors. However, when data are unevenly distributed, the performance of these methods tends to be suboptimal.

Temporal correlation methods estimate the missing data from the same sensor using past data for that sensor. However, if a continuous sequence of data is lost, these methods fail to accomplish comprehensive reconstruction.

In recent studies, several techniques have been used to estimate missing data by utilizing both temporal and spatial correlations. These methods find the most significant spatial samples and temporal series for data and divide missing data into homogeneous spatial regions. However, they need the complete dataset to be computed, which leads to a large amount of redundant data and high computational complexity.

In this study, we introduce the Spatial and Temporal Correlation Estimation Algorithm (STCEA), for estimating the missing data in wireless sensor networks (WSNs). In this algorithm we address the missing data by analyzing the real-world datasets, identifying data loss, and identifying the patterns of data loss in WSNs. We demonstrate the improved efficacy and efficiency of STCEA by comparing its performance with other algorithms that address the problem of missing sensor data.

This study's remaining sections are arranged as follows: Section II reviews traditional strategies for estimating missing data. Section III introduce the proposed algorithm developed in this study. Section IV evaluates the algorithm through simulated experiments. Finally, Section V provides the conclusion of the study.

II. Related Work

Estimating missing data is a preprocessing step for cleaning and preparing incomplete datasets. Ignoring this problem can present significant challenges to WSN applications. If missing data is not addressed, a significant amount of sensor data may be lost. Thus, reducing the accuracy and reliability of the

*Corresponding Author.

application. Therefore, implementing algorithms to estimate missing data is important to address these challenges.

For handling missing data, some methods ignore the missing data that is not well-suited to the nature of WSNs [6]. While other methods re-query the data that consumes additional time and network bandwidth and does not ensure the original data is retrieved. As a result, estimating missing sensor data has become essential.

In this section, we summarize the different estimation algorithms used to deal with the missing data problem in the wireless sensors.

Different algorithms have been conducted on estimating the missing data in statistics, including Mean Substitution, Multiple Imputations, Expectation Maximization, Imputation by Regression, Bayesian Estimation, and Maximum Likelihood [7]. However, these algorithms are unsuitable in Wireless Sensor Networks (WSNs) because they assume that data is missing randomly. Thus, these algorithms are generally inefficient.

Data mining algorithms can be used to estimate the missing data in WSN, involves extracting knowledge from data and applying it to predict the missing values [8]. Algorithms based on association rules are employed to capture the relationships between sensor nodes and data. The goal is to identify all the association rules that meet the user-defined threshold. Examples of these algorithms: Window Association Rule Mining (WARM) [9], Closed Itemsets-based Association Rule Mining (CARM) [10], and Freshness Association Rule Mining (FARM) [11].

Estimating missing data in WSN based on association rules has several challenges. Sensor data in WSN is often changed over time. While association rules-based algorithms are typically static and can’t capture these dynamics effectively. Also, the relationships between sensors can be nonlinear. While association rules typically get linear relationships. These challenges highlight the limitations of using the association rules for estimating the missing data in WSN. To overcome these challenges different algorithms have been developed to address the problem of spatial-temporal missing data. These algorithms are classified into three categories: spatial correlation, temporal correlation, and spatial-temporal correlation.

The Grey System Estimation Algorithm (GSEA) is a spatial correlation algorithm that begins by evaluating the relationships between the target sensor with missing data and its neighboring sensors [12]. Then sort these sensors according to their correlation values, placing those with higher correlations closer to the target sensor. The algorithm calculates this correlation based on the distance between the missing sensor and its neighboring sensors. This algorithm works particularly well with environmental variables, such as voltage, humidity and temperature, which do not exhibit significant changes over small areas. However, GSEA encounters limitations when dealing with large areas of missing data.

Another estimating missing data based on the spatial correlation is Adaptive Multiple Regression (AMR) algorithm [13]. AMR utilizes the multiple regression model for estimating the missing values by considering data from multiple neighboring nodes simultaneously. AMR dynamically adjusts its estimation equation to account for changing correlations among sensor data, resulting in more accurate predictions. However, the algorithm employs a heuristic approach to select sample data and relevant sensors, which increases its computational complexity.

Time series prediction techniques usually build a structure for guessing missing data points at a given area using historical data from that area. The autoregressive integrated moving-average (ARIMA) model is one such technique [14]. However, this approach has two significant drawbacks. First, a lot of prediction models perform poorly because they don't make full use of the crucial features of spatio-temporal data. Second, these prediction approaches frequently fail to produce appropriate reconstruction when a complete consecutive set of data is missing [15].

In recent years, several studies consider both spatial and temporal correlations which are of particular interest to us e.g. Minimized Similarity Distortion (MSD) [16], Mining Autonomous spatial and Temporal (MASTER) [17], Data Reconstruction Algorithm (DRA) [18] and Temporal and Spatial Correlation Algorithm (TSCA) [19].

MSD [16] is an imputation method that reduces similarity distortion by considering various attributes of sensor datasets, not just spatial and temporal dimensions, to ensure comprehensive data segmentation. However, as the number of missing values increases, accurately identifying the correct neighboring unit becomes more challenging, leading to higher error rates. MASTER [17] is an online spatio-temporal mining algorithm that processes data incrementally in a single scan to estimate missing or corrupted sensor data streams.

DRA [18] is a data reconstruction algorithm that accounts for both spatial and temporal correlations, iteratively minimizing the difference between the estimated and reconstructed data, treating the estimated data as the original. TSCA [19] selects sample data for each missing value estimation by leveraging spatial correlation through the calculation of distances between sensor nodes and the missing sensor and then uses past timestamp data for temporal estimation. The final estimated value is obtained by combining both spatial and temporal estimates.

As previously indicated, the most effective linear estimations in both dimensions are obtained by calculating the weights of the spatial and temporal contributions. However, this approach necessitates using the entire dataset for computation, leading to high computational complexity and excessive redundant data. Additionally, when there are consecutive missing data points, the accuracy of interpolation decreases, and it may become difficult to generate the final result.

III. Proposed Estimation Algorithm

We introduce the Spatial and Temporal Correlation Estimation Algorithm (STCEA) to solve the problem of missing data in wireless sensor networks (WSNs). This algorithm is designed for static networks with offline data, where the sensor locations are known. Algorithm 1 outlines the spatial and temporal correlations for STCEA algorithm.

In a WSN, sensor nodes are placed in a designated area and can be represented as ($SN1$, $SN2$, $SN3$, ..., SNm). These nodes periodically report data at times {$t1$, $t2$, ..., tn}. At time ti, the collected data form a time series $SN(ti)$ = ($SN1_{ti}$, $SN2_{ti}$, $SN3_{ti}$, ..., SNm_{ti}).

If the sensor node SNm loses data at tn, we can compute the missing data by using temporal and spatial correlations, which can be determined using the following formula:

$$Missing\ Data\ Estimation = \sum_{i=1}^{SN_m} w_i * S_{Value} + \left(1 - \sum_{i=1}^{SN_m} w_i\right) * T_{Value}$$

where, S_{Value} and T_{Value} denote the results of spatial correlation and temporal correlation, respectively. w_i is weight assigned to each relevant node SNi, determined by average correlation coefficient with the sensor node being estimated. The spatial and temporal results are then combined to derive the estimated value.

$SN1_{t1}$	$SN2_{t1}$	$SN3_{t1}$	$\cdots$	SNm_{t1}
$SN1_{t2}$	$SN2_{t2}$	$SN3_{t2}$	$\cdots$	SNm_{t2}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$SN1_{ti}$	***Missing***	$SN3_{ti}$	$\cdots$	SNm_{ti}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$SN1_{tn}$	$SN2_{tn}$	$SN3_{tn}$	$\cdots$	SNm_{tn}

Fig. 1. Sensor nodes in WSN in a periodical time.

For example, to estimate the missing data for sensor node $SN2$, we calculate the distance between $SN2$ and all other nodes ($SN1$, $SN3$, $SN4$,…, SNm), as illustrated in Fig. 1. We then choose the nodes whose distance is close to $SN2$. These chosen nodes, that exhibit robust spatial correlation with node $SN2$, form the collection $S_{Correlate}$. The spatial correlation estimation is then calculated as follows:

$$S_{Value} = \sum_{SNi} wi * SNj_{(t_{n-1})}$$

where, SNi represents a sensor node in $S_{Correlate}$. $SNj_{(tn-1)}$ denotes the value of SNj at the time moment immediately preceding tn. wi signifies the weight associated with Si, calculated using the nodes' average correlation coefficient.

The evaluation result is derived from a comprehensive assessment of the variability in sample data. Temporal correlations are derived from the stability of environmental variables, which change over short time for the same node. Therefore, we select the closest two time points t_{i-1}, t_{i-2} to the missing time ti. Thus, we evaluate the temporal correlations as follows:

$$T_{Value} = SN_{(t_{i-1})} + cr_{ti}$$

where, cr_{ti} is the change rate of data at time ti

$$cr_{ti} = \frac{SN_{t_{i-1}} - SN_{t_{i-2}}}{t_{i-1} - t_{i-2}}$$

Algorithm 1. Spatial-Temporal Correlation Algorithm

Input:
Matrix of the sensor node data $SN_{m\times n}$
SN_{miss}: estimated sensor node
Output:
Estimation of the missing sensor node data
Begin
1. $S_{Value} = 0$;
2. Calculate the distance between the missing sensor node and all other nodes
3. Add the closest sensor nodes in $S_{Correlate}$
4. **For** each $SN_i \in S_{Correlate}$
5. $wi = \frac{\psi(SN_{miss}, SN_i)}{|S_{Correlate}|}$
6. $S_{Value} = S_{Value} + wi$
7. **End For**
8. Select the closest two time points t_{i-1}, t_{i-2} for SN_{miss}
9. Compute the change rate of data at time ti
10. $cr_{ti} = \frac{SN_{t_{i-1}} - SN_{t_{i-2}}}{t_{i-1} - t_{i-2}}$
11. $T_{Value} = SN_{(t_{i-1})} + cr_{ti}$
12. $Missing\ Data\ Estimation = \sum_{i=1}^{SN_m} w_i * S_{Value} + \left(1 - \sum_{i=1}^{SN_m} w_i\right) * T_{Value}$

End

Here is an example from a real dataset collected by the Intel Berkeley Research Lab [20], which records time, node ID, humidity, temperature, light, and voltage values in every thirty seconds, as shown in Table I. We select a portion of the dataset, remove some readings, and then compare the estimated values with the actual recorded values, replacing the missing data with "NaN" to indicate the absence of values.

In this example we select two records with the bold values to be the records with missing values, replacing them with *NaN* values. This case represents element sequence loss pattern since sensor 31 has missing data at two continuous times 180, 150. STCEA starts from the last missing record to avoid the incremental error, estimating each attribute separately.

TABLE I. THE REAL DATASET FROM INTEL BERKELEY RESEARCH LAB

Time (seconds)	Node id (integer)	Temperature (celsius)	Humidity (0-100%)	Light (lux)	Voltage (volts)
150	34	18.8712	40.2976	60.72	2.67532
150	27	19.7434	38.1897	79.12	2.69964
150	31	19.028	40.4328	150.88	2.69964
150	6	19.9002	37.5737	121.44	2.63964
150	29	19.3612	39.2123	180.32	2.68742
150	10	19.3612	39.8235	75.44	2.67532
180	25	18.93	38.8039	97.52	2.68742
180	30	18.4596	41.5789	114.08	2.68742
180	10	19.273	39.9252	75.44	2.67532
180	26	18.8614	40.9055	121.44	2.68742
180	31	18.9496	40.3652	150.88	2.69964
180	32	18.734	39.9929	121.44	2.69964
180	21	19.5964	37.162	114.08	2.69964

After applying the proposed algorithm, we obtain the following values: 19.1162 and 40.1424 for temperature and humidity respectively, for the missing data record at time 180. Additionally, we obtain 19.1542 and 40.2124 for temperature and humidity respectively, for the missing data record at time 150. Based on the earlier findings, it is clear that there is only a slight difference between the obtained values and the actual values.

IV. Experimental Results

A. Datasets

We evaluate the proposed algorithm using two real-world datasets: the Intel Lab dataset [20] and the air quality dataset [21]. The Intel Lab dataset includes data from 54 sensor nodes at the Intel Research Berkeley Lab. These sensors recorded temperature, humidity, light, and voltage every 30 seconds. The sensor layout in the lab is shown in Fig. 2. The air quality dataset consists of records from 34 monitoring stations in Beijing, collected on an hourly basis, as shown in Fig. 3. This dataset includes a total of 2,891,393 air quality records gathered from May 2014 to April 2015. Each entry represents an air quality record, with columns for Station ID, Time, PM2.5, PM10, NO2, CO, O3, and SO2. Tables II and III display the missing data ratios for the Intel Lab and air quality datasets, with missing values indicated as NULL in the data files.

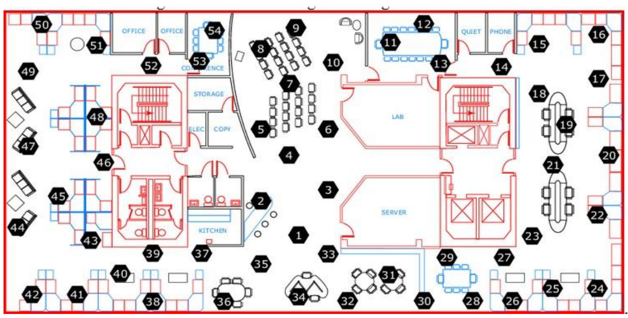

Fig. 2. Sensor arrangement diagram.

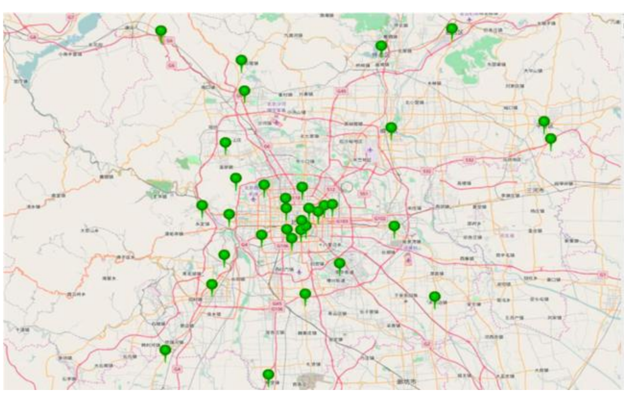

Fig. 3. Air quality stations.

TABLE II. RATIO OF MISSING DATA OF SIX POLLUTANTS IN AIR QUALITY DATASET

Missing Data	PM25	PM10	NO2	CO	O3	SO2
Ratio of Missing	13.3%	14.6%	16%	15.1%	15.4%	15.2%

TABLE III. RATIO OF MISSING DATA OF FOUR WEATHER CONDITIONS IN INTEL-LAB DATASET

Missing Data	Temperature	Humidity	Light	Voltage
Ratio of Missing	15.2%	16.6%	15.3%	15.7%

B. Evaluation Metrics

For evaluating the performance of proposed algorithm, we apply Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as evaluation metrics. MAE is calculated as the average of the absolute differences between the actual values and the predicted values.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

where, y_i and x_i are the prediction and true value respectively, and n is the total number of points.

RMSE measures the differences between true missing value and estimated values, RMSE is defined as follow:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}}$$

where, x_i and $\hat{x}_i$ is the real missing data and the computed value of the missing data respectively, and N is the number of non-missing points.

C. Result Analysis

To assess the effectiveness of the STCEA algorithm, we compare it with four existing algorithms: Adaptive Multiple Regression (AMR) [13], Minimized Similarity Distortion (MSD) [16], TSCA [19], and DRA [18].

Table IV and Table V show the results of the algorithms on datasets of the Air quality and Intel-lab dataset respectively. The results are based on the data missing ratios detailed in Table II and Table III for each dataset. The MAE and RMSE values of our method are significantly lower than those of the AMR, MSD, TSCA, and DRA methods, demonstrating a substantial improvement in interpolation accuracy. The STCEA algorithm achieves minimal errors and maintains consistent stability across different datasets.

As shown in Table IV and Table V, the estimation errors of the AMR algorithm are higher than other algorithms. This is because AMR algorithm computes the missing based on the spatial correlation. While MSD, TSCA and DRA computes the missing based on the spatial correlation and temporal correlation, so their estimation errors lower than AMR. Also, the evaluated results show that TSCA performed better than MSD and DRA but does not fully address the impact of missing patterns before interpolation, which worsens the results for some datasets. This variability in performance highlights the different missing data patterns present in each dataset.

From Tables IV and V, it is evident that STCEA consistently outperforms other algorithms in estimation accuracy. This is due to STCEA's ability to estimate missing data by leveraging spatial and temporal correlations, as well as the functional relationships among sensor data. Therefore, STCEA demonstrates the most stable estimation performance.

TABLE IV. PERFORMANCE COMPARISON ON AIR QUALITY DATASET

Dataset	Methods	MAE	RMSD
PM25	STCEA	**6.3244**	**0.2314**
	AMR	16.3245	0.8424
	MSD	14.4313	0.7420
	TSCA	11.2324	0.5343
	DRA	12.2424	0.7424
PM10	STCEA	**7.3245**	**0.2144**
	AMR	17.3435	0.8243
	MSD	13.5456	0.7464
	TSCA	12.2456	0.6724
	DRA	12.2425	0.8245
NO2	STCEA	**5.2426**	**0.1935**
	AMR	13.3567	0.7462
	MSD	11.3344	0.6355
	TSCA	9.4342	0.6724
	DRA	11.2457	0.7891
CO	STCEA	**6.2425**	**0.1358**
	AMR	15.2426	0.8388
	MSD	12.2863	0.8198
	TSCA	10.3536	0.6012
	DRA	12.0012	0.7534
O3	STCEA	**9.3535**	**0.0724**
	AMR	17.3531	0.8274
	MSD	15.4789	0.7835
	TSCA	11.2453	0.6357
	DRA	14.3536	0.9383
SO2	STCEA	**11.2426**	**0.0513**
	AMR	19.3682	0.9246
	MSD	16.4647	0.8375
	TSCA	12.5670	0.7234
	DRA	15.2450	0.8833

TABLE V. PERFORMANCE COMPARISON ON INTEL-LAB DATASET

Dataset	Methods	MAE	RMSD
Temperature	STCEA	**9.1345**	**0.1425**
	AMR	17.2474	0.9435
	MSD	15.3531	0.8353
	TSCA	12.5368	0.6463
	DRA	15.8643	0.8124
Humidity	STCEA	**10.3452**	**0.1391**
	AMR	17.3537	0.9352
	MSD	15.6278	0.9242
	TSCA	12.2781	0.6240
	DRA	14.6468	0.8724
Light	STCEA	**10.2424**	**0.1313**
	AMR	19.3632	0.7257
	MSD	16.3536	0.6955
	TSCA	13.3536	0.6435
	DRA	16.3536	0.9101
Voltage	STCEA	**10.2525**	**0.0925**
	AMR	18.3536	0.8242
	MSD	16.3699	0.8082
	TSCA	11.9735	0.6136
	DRA	15.4641	0.8133

V. CONCLUSION AND FUTURE WORK

STCEA algorithm is an effective solution to the challenges of missing data in wireless sensor networks (WSNs). By integrating both spatial and temporal correlations, the algorithm is capable of identifying substantial data loss and detecting underlying patterns, facilitating accurate partial reconstruction of missing data. We thoroughly evaluated the algorithm's performance using real world data and compared its accuracy with several existing missing data methods. The experimental results demonstrate that STCEA consistently outperforms other approaches in terms of estimation accuracy. Moving forward, future research will focus on further enhancing the STCEA algorithm to handle more complex scenarios, such as when all sensors in a given time series experience data loss, ensuring broader applicability and robustness in real-world WSNs.

ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Saudi Arabia, under grant No. (UJ-23-DR-273). The authors, therefore, acknowledge with thanks the university's technical and financial support.

REFERENCES

[1] G. B. Tayeh, A. Makhoul, C. Perera, and J. Demerjian, "A spatial-temporal correlation approach for data reduction in cluster-based sensor networks", IEEE Access, 7, pp 50669-50680, 2019.

[2] Y. Gong, Z. Li, J. Zhang, W. Liu, Y. Yin, Y. and Y. Zheng, "Missing value imputation for multi-view urban statistical data via spatial correlation learning", IEEE Transactions on Knowledge and Data Engineering, 35(1), pp.686-698, 2021.

[3] A. A. Almazroi, N. Ayub, "Nature-inspired approaches for clean energy integration in smart grids", Alexandria Engineering Journal, 105, pp 640-654, 2024.

[4] A. Mustafi, A. I. Middya, and S. Roy, "Fuzzy-based missing value imputation technique for air pollution data", Artificial Intelligence Review, 56(2), pp.1-38, 2023.

[5] A. A. Almazroi, F. S Alsubaei, N. Ayub, N. Z. Jhanjhi, Inclusive Smart Cities: IoT-Cloud Solutions for Enhanced Energy Analytics and Safety, International Journal of Advanced Computer Science & Applications, 15(5), 2024.

[6] W. N. El-Sayed, H. M. El-Bakry, and S.M El-Sayed, "Integrated data reduction model in wireless sensor networks", Applied Computing and Informatics, 19(1/2), pp.41-63, 2023.

[7] A. Mirzaei, S.R. Carter, A.E Patanwala, and C.R Schneider, "Missing data in surveys: Key concepts, approaches, and applications", Research in Social and Administrative Pharmacy, 18(2), pp.2308-2316, 2022.

[8] Y. Gong, T. He, M. Chen, B. Wang, L, Nie, and Y. Yin, "Spatio-Temporal Enhanced Contrastive and Contextual Learning for Weather Forecasting", IEEE Transactions on Knowledge and Data Engineering, 2024.

[9] M. Halatchev, and L. Gruenwald, "Estimating missing values in related sensor data stream", In: COMAD, pp. 83–94, 2005.

[10] N. Jiang and L. Gruenwald, "Estimating missing data in data streams", In: DASFAA, pp. 981–987, 2007.

[11] L. Gruenwald, H. Chok, and M. Aboukhamis, "Using Data Mining to Estimate Missing Sensor Data", 7th IEEE ICDM Workshop on Data Mining, pp 207-212, 2007.

[12] F. Liu, Z. You, W. Shan, and J. Liu, "A grey system based missing sensor data estimation algorithm", in Proceedings of the 2nd International Conference on Computer Science and Network Technology, pp. 482–486, IEEE, 2012.

[13] L. Pan, H. Gao, H. Gao, and Y. Liu, "A spatial correlation based adaptive missing data estimation algorithm in wireless sensor networks", International Journal of Wireless Information Networks, 21(4), pp. 280–289, 2014.

[14] C. Yozgatligil, S. Aslan, C. Iyigun, "Comparison of missing value imputation methods in time series: The case of Turkish meteorological data", Theor. Appl. Climatol, 112, pp 143–167, 2013.

[15] Y. Li, Z. Li, "Efficient missing data imputing for traffic flow by considering temporal and spatial dependence", Transp. Res. Part C Emerg. Technol, 34, pp 108–120, 2013.

[16] K. Niu, F. Zhao, and X. Qiao, “A missing data imputation algorithm in wireless sensor network based on minimized similarity distortion”, in Proceedings of the 6th International Symposium on Computational Intelligence and Design, pp. 235–238, 2013.

[17] H. Chok, and L. Gruenwald, “Spatio-Temporal Association Rule Mining Framework for Real-time Sensor Network Applications”, In: ACM, 2009.

[18] L. Kong, M. Xia, X. Liu, M. Wu, and X. Liu, “Data Loss and Reconstruction in Sensor Networks”, In Proc. IEEE INFOCOM, pp. 1654–1662, 2013.

[19] Z. Gao, W. Cheng, X. Qiu, and L. Meng, “A Missing Sensor Data Estimation Algorithm Based on Temporal and Spatial Correlation”, In: International Journal of Distributed Sensor Networks, 2015.

[20] S. Madden, Intel Berkeley research lab data, http://www.select.cs.cmu.edu/data/labapp3/index.html.

[21] Y. Zheng, X. Yi, M. Li, “Forecasting fine-grained air quality based on big data”, In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015.