Research Article

Michał Kekez*

Model-based imputation of sound level data at thoroughfare using computational intelligence

<https://doi.org/10.1515/eng-2021-0051>

Received Sep 15, 2020; accepted Jan 08, 2021

Abstract: The aim of the paper was to present the methodology of imputation of the missing sound level data, for a period of several months, in many noise monitoring stations located at thoroughfares by applying one model which describes variability of sound level within the tested period. To build the model, at first the proper set of input attributes was elaborated, and training dataset was prepared using recorded equivalent sound levels at one of thoroughfares. Sound level values in the training data were calculated separately for the following 24-hour sub-intervals: day (6-18), evening (18-22) and night (22-6). Next, a computational intelligence approach, called Random Forest was applied to build the model with the aid of Weka software. Later, the scaling functions were elaborated, and the obtained Random Forest model was used to impute data at two other locations in the same city, using these scaling functions. The statistical analysis of the sound levels at the above-mentioned locations during the whole year, before and after imputation, was carried out.

Keywords: imputation, monitoring station, sound level, random forest, scaling functions

1 Introduction

Missing values in measurement data always hamper interpretation of results, regardless of the area of research [1]. The reasons for the lack of data can be analyzed using three models: MCAR (missing completely at random), MAR (missing at random), and MNAR (missing not at random). In the last two models, the missingness of data is related to the data observed, or caused by a malfunction of measurement path components, wrong decisions or ethical considerations. In consequence, missing data may lead to bias as

they do not appear in the sample completely randomly [1]. These factors have led to the development of various computational methods helping to overcome problems related to missingness of data [1]. In [2], the authors proposed classification of these methods into a weighting approach [3] and an imputation-based approach [1, 4]. Both approaches use additional information on the phenomena under study [1]. Weighting methods make adjustments due to missing data by modifying the base weights. Imputation methods use additional information to build the imputation model, on the basis of which the missing data is imputed [1, 5]. Imputation methods are divided into deductive and statistical [6]. Deductive methods use rules and relationships between variables for determining the missing data. Statistical methods use remaining part of dataset for reconstruction of the missing values. These methods can be divided into deterministic (imputation by the mean, and regression imputation) and stochastic (hot-deck, and stochastic regression imputation) [6]. However, they often do not give satisfactory results [7].

More sophisticated methods of imputation require building a model. When we consider time series data imputation only, autoregressive and computational intelligence (CI) methods [8] can be applied to building models. Such models can be often used also for time series forecasting [9]. Among autoregressive methods [10], autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) [10], autoregressive conditional heteroscedasticity (ARCH), and generalized autoregressive conditional heteroscedasticity (GARCH) [11] are used. Examples of machine learning and computational intelligence methods used for missing data imputation are [12]: K-nearest neighbor (KNN), fuzzy K-means (FKM), singular value decomposition (SVD), and Bayesian principal component analysis (BPCA) as well as regression trees [1] like classification and regression trees (CART) [13] or Cubist [14]. CI methods for modeling, e.g. neural networks, or fuzzy systems are often used together with optimization algorithms like [15]: genetic algorithms, particle swarm optimization (PSO), ant colony optimization, and memetic algorithms [16]. Hybrid connections of various methods are also used for imputation [8]: hybrid simulated annealing and genetic algorithms (HSAGA), hy-

*Corresponding Author: Michał Kekez: Faculty of Mechatronics and Mechanical Engineering, Kielce University of Technology, al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland; Email: mkekez@tu.kielce.pl

brid of autoassociative neural network, simulated annealing and genetic algorithm (AANN-HSAGA), hybrid of principal component analysis and artificial neural networks (PCA-ANN), hybrid of principal component analysis, neural network, simulated annealing and genetic algorithm (PCA-NN-HSAGA). Ensembles of models, like committees of neural networks [8] or committees of regression trees [1] are also used in imputation.

In various transportation-related problems, computational intelligence methods as well as autoregressive methods like ARIMA are used for data imputation [17, 18]. Neural networks were used for imputation in [19]. Machine learning was used for cleaning data collected in intelligent transportation systems [20]. Problems regarding road safety and modeling of transport processes were analyzed in [21] and the proposed tool for imputation was Random Forest. In [1], two kinds of CI methods, namely regression trees and Random Forest, were used for analysis of road traffic noise. However, imputation of road traffic noise at various locations using only one model required the development of a new methodology, as discussed in Section 2, 3 and 4. This allows building the initial model for the first location near thoroughfare and the immediate extension of this model for any new thoroughfare in the same city.

2 Measurement data used for building the model

Road traffic and noise monitoring stations, located near thoroughfares in many cities, constantly record sound level, traffic volume, and vehicle speed and type. Recorded data can be used for various purposes, including calculation of long-term noise indicators L_{DEN} and L_N [22], environmental monitoring, and creation of acoustic maps. However, if monitoring stations cease to function partially or completely, missing values of sound level need to be imputed. When traffic data are present, imputation can be carried out by using deductive method, e.g. using CNOSSOS-EU [23] or Nordic prediction method [24]; otherwise the possible way of sound level imputation is to produce models for traffic volume [18, 25–31], and vehicle speed and type, impute missing traffic data using these models, and finally use previously mentioned deductive method. However, such multi-stage imputation decreases the quality of imputed data. The better solution is to create the model of sound level variability and use it for imputation. To present the proposed methodology, the data recorded in a noise monitoring station will be used.

Sound level values were recorded in a noise monitoring station, situated at the location number 1 (thoroughfare, namely Krakowska Street in Kielce, Poland), consisting of class-1 sound level meter, a road radar, and weather station [1, 12]. Measurements were made continuously and the RMS (root mean square) of the A-weighted sound level was saved in the buffer in 1 second intervals with a resolution of 0.1 dB. This allowed to calculate the most common indicator of noise annoyance [32], namely A-weighted equivalent sound level L_{Aeq} , expressed in dB(A), defined as [32, 33]:

$$L_{Aeq} = 10 \log \left[\frac{1}{T} \int_0^T \left(\frac{p_A(t)}{p_0} \right)^2 dt \right] \quad (1)$$

where T represents the total time of measurement (expressed in s), $p_A(t)$ – A-weighted sound pressure (in Pa), and p_0 – reference sound pressure of 20 μ Pa.

Based on the previously mentioned measurements, L_{Aeq} values were calculated for the three 24-hour sub-intervals: day (6-18), evening (18-22), and night (22-6), separately for each 24-hour period in the year, as shown in [1] and in Figure 1.

Figure 1: L_{Aeq} calculated from measurements made at the location number 1, in year 2013, for: day sub-interval (6-18), (solid line), and night sub-interval (22-6), (dash-dot line)

In Figures 1–6, numbers on the horizontal axis show consecutive 24-hour periods in the year, numbered from 1 to 365; night sub-interval (hours from 22 to 6) is counted as part of the 24-hour period ending at 6 a.m. The L_{Aeq} values for evening (hours from 18 to 22), presented in [1], were omitted in Figure 1 to improve the readability of the chart.

Data calculated from the measurements made in the year 2013 includes 905 records describing the equivalent sound level for a particular sub-interval: day (301 records), evening (302 records) or night (also 302 records). For each of the sub-intervals, the L_{Aeq} values are missing for almost all of the first 44 and last 26 days of the year (Figure 1). Median

values of non-missing L_{Aeq} values in 2013 are: 70.42 dB for day sub-interval, 68.79 dB for evenings, and 64.785 dB for nights [1].

3 Elaborated model

The training data for the model contains equivalent sound level (L_{Aeq}) values of six previous days (for the same sub-

interval of the 24-hour period) marked $l_1, l_2, \dots, l_6$, where l_i is the L_{Aeq} recorded i days earlier. The authors in [1] created 3 separate training datasets, one for each sub-interval of 24-hour period (or time of the day, in other words). Each training set consisted of records containing the values of one output attribute dB_A (equivalent sound A-level, expressed in dB) and 8 input attributes: *day_of_the_week* (taking values from 1 – Monday to 7 – Sunday), *day_of_the_year* (values in the range from 1 to 365), l_1, l_2, l_3, l_4, l_5 , and l_6 . There was no *time_of_day* attribute (taking values 0 for night, 1 for evening, and 2 for day) in the created sets, because it was held constant in the entire set. The training sets included all records (301 for days, 302 for evenings, and again 302 for nights). Testing was conducted by 10-fold cross validation [1]. Certain records in the training sets showed the missing values of some of $l_1, l_2, \dots, l_6$ of the input attributes, and contrary to model 1 in [1], these records were not removed from the training dataset.

The model was constructed using Random Forest algorithm without random selection of attributes, implemented in Weka software [34]. The obtained model consists of 300 trees (100 for each sub-interval of the 24-hour period). In this method, the size of the trees may be large due to no pruning [35]. To assess the accuracy of prediction made by the model, mean absolute error (MAE), which is the arithmetic average of absolute values of differences between the predicted and real value, can be used [36]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y'_i - y_i| \quad (2)$$

where y_i denotes real dB_A value, y'_i denotes dB_A value calculated by the model, and n is the number of records. Accuracy of the model defined by MAE estimator (eq. 2) is very good because MAE on the training set does not exceed 0.27 dB, and MAE on the test set during 10-fold cross validation is not higher than 0.72 dB (Table 1) [1]:

Table 1: Accuracy of the model at location no. 1 [1]

Dataset or validation method	MAE of the model		
	Day	Evening	Night
Training dataset	0.21 dB	0.27 dB	0.23 dB
Ten-fold cross validation	0.56 dB	0.72 dB	0.62 dB

Values of L_{Aeq} calculated by the model for the location no. 1 are shown in Figure 2. Vertical dashed lines in Figure 2 separate days of the year (from 1 to 7 or 8, from 14 or 15 to 43 or 44, and from 340 or 341 to 365) for which the measurement data was missing.

Figure 2: L_{Aeq} calculated by the model for the location no. 1: (a) day sub-interval (6-18), (b) evening (18-22), and (c) night (22-6) in year 2013

The L_{Aeq} values calculated by the model, shown in Figure 2, have smaller variance than L_{Aeq} calculated from measurements (Figure 1). One can observe that minimum value of modeled L_{Aeq} at night, 60.76 dB, shown in Figure 2c is higher than corresponding measurement value of 59.12 dB in Figure 1. Similarly, maximum value of modeled L_{Aeq} at night, 66.73 dB is lower than corresponding value of 67.69 dB in Figure 1. However, the overall accuracy of the model, shown in Table 1 is quite good.

The model was used for imputation of data at location no. 1, for the whole year 2013. This means that missing L_{Aeq} values in measurement data were replaced by L_{Aeq} calculated by the model, while remaining part of data was not changed. Median values of L_{Aeq} sets after imputation are: 70.51 dB (day), 68.74 dB (evening), 64.72 dB (night). After the imputation of missing L_{Aeq} values by the model for the whole year 2013 at location no. 1 (Table 2), the median of L_{Aeq} did not change significantly (at most ± 0.09 dB). The quartiles Q_1 and Q_3 did not change more than ± 0.18 dB (Q_1) and ± 0.15 dB (Q_3).

Table 2: Selected parameters of the model at location no. 1 [1]

L_{Aeq} values	Q_1 quartile for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	69.58 dB	68.28 dB	64.293 dB
Calculated by the model	69.686 dB	68.394 dB	64.202 dB
After imputation	69.62 dB	68.27 dB	64.113 dB
L_{Aeq} values	Median for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	70.42 dB	68.79 dB	64.785 dB
Calculated by the model	70.517 dB	68.776 dB	64.78 dB
After imputation	70.51 dB	68.74 dB	64.72 dB
L_{Aeq} values	Q_3 quartile for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	70.9 dB	69.435 dB	65.31 dB
Calculated by the model	70.928 dB	69.179 dB	65.153 dB
After imputation	70.98 dB	69.29 dB	65.26 dB

4 Generalization of the model by using scaling functions

The idea of imputing traffic data at one location by the model built on data from nearby location was presented in [5]. In this section, the idea of imputing sound level data at given location by the model built for another location will be presented.

The elaborated model described in Section 3 can be adjusted to predict sound level values at another place in the same city, located close to any road of the same class as at location no. 1. For this purpose, the $l_1, l_2, \dots, l_6$ inputs of the model are modified by input scaling function (eq. 5), while y output of the model is modified by the output scaling function (eq. 6).

One can assume that at all thoroughfares in a given city, for a given 24-hour sub-interval and for a given day of week, traffic volume can be expressed as a product of a constant and a coefficient having the value specific to this road. When Nordic prediction model [24] is applied to calculate sound level at any of these thoroughfares (and when percentage of heavy vehicles is similar at all thoroughfares), we obtain the sound level expressed as a sum of a constant and a parameter having the value specific to this road. This led to the idea of output scaling function (eq. 6) in the form of the sum of a constant (obtained for location no. 1) and a parameter specific to given thoroughfare (calculated separately for each day of week and each of three 24-hour sub-intervals).

In order to obtain parameters of both scaling functions, at first $a(d, t)$ values, which are equivalent sound levels [33] for a given day of the week d , and for a given 24-hour sub-interval t , are calculated separately for each $d=1, 2, \dots, 7$, and for each $t=0, 1, 2$, with use of L_{Aeq} measurement data records from location no.1 (described as training data in Section 3):

$$a(d, t) = 10 \log \left(\frac{1}{n} \sum_{i=1}^n 10^{0.1 y_i} \right) \quad (3)$$

for all n records fulfilling the condition $day_of_the_week = d$ and $time_of_day = t$, and where $day_of_the_week$ is input attribute (1 – Monday, 2 – Tuesday, ..., 7 – Sunday), $time_of_day$ is also input attribute (0 – night, 1 – evening, 2 – day), and y_i denotes db_A value in measurement data for given location. Values of $a(d, t)$ for location no. 1 are shown in Table 3.

The $a(d, t)$ values for location no. 1 (calculated according to eq. 3) are denoted as $a_1(d, t)$. Then, the $a(d, t)$ values for a new location are calculated according to eq. 3 (using measurement data from this new location), and denoted as $a_2(d, t)$. Later, the parameters of scaling functions, namely

Table 3: Values of $a(d, t)$ in dB, for location no. 1

$a(d, t)$, in dB				d			
t	1	2	3	4	5	6	7
2	70.6	70.9	70.6	70.7	70.7	69.7	68.5
1	68.7	68.7	68.9	69.2	69.4	68.4	68.6
0	64.5	64.7	64.7	64.9	65.0	65.2	63.8

$p(d, t)$ values, are calculated separately for $d=1,2,\dots,7$, and $t=0,1,2$:

$$p(d, t) = a_2(d, t) - a_1(d, t) \quad (4)$$

The extended model, proposed in this section, for the given data record replaces the value of each input attribute l_i with the corresponding l'_i value, calculated using the so-called input scaling function:

$$l'_i = l_i - p(d', t) \quad \text{for } d' = (d + 6 - i) \bmod 7 + 1 \quad (5)$$

where d is *day_of_the_week* attribute value, and t is *time_of_day* attribute value in the given data record.

Then, the extended model produces its output y by using the so-called output scaling function:

$$y = y' + p(d, t) \quad (6)$$

where y' is the value of the output of elaborated model shown in Section 3, d is *day_of_the_week* attribute value, and t is *time_of_day* attribute value, in the given data record.

4.1 Application of the model for location no. 2

The values of L_{Aeq} at location no. 2 (thoroughfare, namely Jesionowa Street in Kielce, Poland) in year 2013 calculated from measurements are shown in Figure 3.

For over 130 days, the L_{Aeq} values are missing (Figures 3a, 3b, 3c). However, the L_{Aeq} data for the first and for the last 10 days of the year are present (contrary to data at location no. 1), with L_{Aeq} taking values often close to the year's minimum. The lowest value of L_{Aeq} for day sub-interval (Figure 3a) was 65.9 dB at 1st Jan, and highest was 76.6 dB at 20th Nov. The lowest values for evening and night sub-intervals were usually observed at national holidays.

In order to adjust the model (presented in Section 3) for location no. 2, the values of $a(d, t)$ (eq. 3) for this location were calculated, based on the measurement values. Next, the $p(d, t)$ values (eq. 4) for scaling functions were calculated. Then, the model with output scaling function

(a)

(b)

(c)

Figure 3: L_{Aeq} calculated from measurements made at the location no. 2, in year 2013, for: (a) day sub-interval (6-18), (b) evening (18-22), and (c) night (22-6) in year 2013

Figure 4: L_{Aeq} calculated by the model for the location no. 2, in year 2013, for: (a) day sub-interval (6-18), (b) evening (18-22), and (c) night (22-6) in year 2013

was used to predict the values of L_{Aeq} for location no. 2 (Figure 4).

Absence of L_{Aeq} data for the first and for the last week in the learning set of the model resulted in lower accuracy of the model for these two weeks at location no. 2 (Figure 4). As a result, minimum value of modeled L_{Aeq} for evening, 68.92 dB, shown in Figure 4b is higher than corresponding measurement value of 65.97 dB in Figure 3b. Similarly, maximum value of modeled L_{Aeq} for evening, 71.98 dB, is lower than corresponding value of 74.48 dB in Figure 3b. However, the overall accuracy of the model, shown in Table 5, is fairly good.

The model was used for imputation of data at location no. 2, for the whole year 2013. Median values of imputed L_{Aeq} sets are: 72.333 dB (day), 71.03 dB (evening), 67.46 dB (night). After the imputation of missing L_{Aeq} values by the model for the whole year 2013 at location no. 2 (Table 4), the median of L_{Aeq} did not change significantly (less than ± 0.06 dB). The quartiles Q_1 and Q_3 did not change more than +0.18 dB (Q_1) and -0.39 dB (Q_3).

To assess the accuracy of the model with scaling functions, parameters of that function were computed again, based on training data containing only about 2/3 of the

Table 4: Selected parameters of the model at location no. 2

L_{Aeq} values	Q_1 quartile for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	71.57 dB	70.35 dB	66.15 dB
Calculated by the model	72.089 dB	70.621 dB	66.639 dB
After imputation	71.7 dB	70.39 dB	66.323 dB
L_{Aeq} values	Median for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	72.28 dB	71.05 dB	67.43 dB
Calculated by the model	72.524 dB	71.014 dB	67.617 dB
After imputation	72.333 dB	71.03 dB	67.46 dB
L_{Aeq} values	Q_3 quartile for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	72.94 dB	71.885 dB	68.308 dB
Calculated by the model	72.851 dB	71.280 dB	68.130 dB
After imputation	72.755 dB	71.495 dB	68.024 dB

Table 5: Accuracy of the model with scaling functions at location no. 2

Dataset	Mean absolute error (MAE)		
	Day	Evening	Night
Training data	0.805 dB	0.923 dB	0.978 dB
Test data	0.775 dB	0.793 dB	1.105 dB

whole dataset. Using that model, mean absolute error (eq. 2) was calculated on training data and on the remaining test data. The MAE was in the range from 0.8 to 1.1 dB (Table 5).

4.2 Application of the model for location no. 3

The values of L_{Aeq} at location no. 3 (thoroughfare, namely Lodzka Street in Kielce, Poland) in year 2013 calculated from measurements are shown in Figure 5.

For about 120 days, mainly from May to August and in November and December, the L_{Aeq} values are missing (Figure 5). The lowest values of L_{Aeq} for every 24-hour sub-

Figure 5: L_{Aeq} calculated from measurements made at the location no. 3, in year 2013, for: (a) day sub-interval (6-18), (b) evening (18-22), and (c) night (22-6) in year 2013

interval were recorded at national holidays and in January and December.

In order to adjust the model (presented in Section 3) for location no. 3, the values of $a(d, t)$ (eq. 3) and $p(d, t)$ (eq. 4) for this location were calculated. The values of L_{Aeq} at location no. 3 in year 2013 calculated by the model with output scaling functions are shown in Figure 6.

The minimum values of modeled L_{Aeq} for day, evening, and night (70.56, 66.38, and 61.92 dB, respectively), shown in Figure 6, are higher than the corresponding measure-

Figure 6: L_{Aeq} calculated by the model for the location no. 3, in year 2013, for: (a) day sub-interval (6-18), (b) evening (18-22), and (c) night (22-6) in year 2013

ment values of 65.7, 61.4, and 59.3 dB, respectively (Figure 5). Similarly, the maximum values of modeled L_{Aeq} for day, evening, and night (73.52, 71.56, and 68.27 dB, respectively) are lower than the corresponding measurement values of 74.8, 73.2, and 69.9 dB, respectively (Figure 5). However, the overall accuracy of the model, shown in Table 7, is fairly good.

The model was used for imputation of data at location no. 3, for the whole year. Median values of L_{Aeq} after imputation are: 72.6 dB (day), 70.6 dB (evening), 67.1 dB (night). After the imputation of missing L_{Aeq} values by the model for the whole year 2013 at location no. 3 (Table 6), the median of L_{Aeq} did not change significantly (at most -0.1 dB). The quartiles Q_1 and Q_3 did not change more than $+0.4$ dB (Q_1) and -0.3 dB (Q_3).

To assess the accuracy of the model with scaling functions at location no. 3, the same procedure as for location no. 2 was applied. The MAE was in the range from 0.7 to 1.1 dB (Table 7).

Table 6: Selected parameters of the model at location no. 3

L_{Aeq} values	Q_1 quartile for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	71.6 dB	69.7 dB	65.725 dB
Calculated by the model	71.859 dB	69.809 dB	65.978 dB
After imputation	71.7 dB	69.733 dB	66.1 dB
L_{Aeq} values	Median for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	72.7 dB	70.7 dB	67.1 dB
Calculated by the model	72.699 dB	70.318 dB	66.828 dB
After imputation	72.6 dB	70.6 dB	67.1 dB
L_{Aeq} values	Q_3 quartile for 24-hour sub-interval		
	Day	Evening	Night
Before imputation	73.3 dB	71.3 dB	67.7 dB
Calculated by the model	72.977 dB	70.745 dB	67.357 dB
After imputation	73.1 dB	71.0 dB	67.593 dB

Table 7: Accuracy of the model with scaling functions at location no. 3

Dataset	Mean absolute error (MAE)		
	Day	Evening	Night
Training data	0.706 dB	0.773 dB	0.827 dB
Test data	0.935 dB	1.094 dB	1.001 dB

5 Conclusions

The presented model with scaling functions was successfully applied to imputation of missing L_{Aeq} values at three various locations at thoroughfares in the same city. After imputation by the model with scaling functions, the Q_1 and Q_3 quartiles slightly changed (no more than ± 0.4 dB), and the median values did not change more than ± 0.1 dB. To evaluate the quality of the model, 10-fold cross validation (for location no. 1) and train and test sets (for locations no. 2 and 3) were applied. The accuracy of the model at location no. 1 was not worse than 0.72 dB (MAE value), while at locations no. 2 and 3 the MAE did not exceed 1.11 dB.

The accuracy of the model presented in Section 3 and 4 is sufficient for many practical purposes.

References

- [1] Kekez M, Radziszewski L, Bąkowski A. Application of selected AI models to reconstruction of noise levels at thoroughfare during selected year. 2019 20th International Carpathian Control Conference (ICCC), Krakow-Wieliczka, Poland. 2019. doi: 10.1109/CarpathianCC.2019.8765915.
- [2] Frątczak E, Kamiński A, Kordos J, editors. Statystyka – zastosowania biznesowe i społeczne. Warszawa: Wyższa Szkoła Menedżerska w Warszawie; 2014. Polish.
- [3] Särndal CE, Lundström S. Estimation in surveys with nonresponse. Chichester: John Wiley & Sons; 2005.
- [4] Little JA, Rubin D. Statistical analysis with missing data. Hoboken: John Wiley & Sons; 2002.
- [5] Henrickson K, Zou Y, Wang Y. Flexible and robust method for missing loop detector data imputation. Transportation Research Record: Journal of the Transportation Research Board. 2015;2527:29–36.
- [6] Piasecki T. Metody imputacji w badaniach gospodarstw domowych. Wiadomości Statystyczne. 2014;9:1–20. Polish.
- [7] van Buuren S. Flexible imputation of missing data. 2nd ed. Boca Raton: Chapman & Hall/CRC; 2018.
- [8] Marwala T. Computational intelligence for missing data imputation, estimation, and management. Hershey, New York: Information Science Reference/IGI Global; 2009.
- [9] Kalimoldayev M, Drozdenko A, Koplyk I, Marinich T, Abdildayeva A, Zhukabayeva T. Analysis of modern approaches for the prediction of electric energy consumption. Open Engineering. 2020;10:350–361.
- [10] Box G, Jenkins G, Reinsel G, Ljung G. Time series analysis: forecasting and control. 5th ed. New Jersey: John Wiley & Sons; 2016.
- [11] Engle RF. Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. Econometrica. 1982;50(4):987–1008.
- [12] Schmitt P, Mandel J, Guedj M. A comparison of six methods for missing data imputation. Journal of Biometrics & Biostatistics. 2015;6:224. doi:10.4172/2155-6180.1000224
- [13] Painsky A, Rosset S. Cross-validated variable selection in tree-based methods improves predictive performance. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(11):2142–2153.
- [14] McCandless T, Dettling S, Haupt SE. Comparison of implicit vs. explicit regime identification in machine learning methods for solar irradiance prediction. Energies. 2020;13(3):689:1–14.
- [15] Engelbrecht AP. Fundamentals of computational swarm intelligence. Chichester: John Wiley & Sons; 2005.
- [16] Moscato P, Mathieson L. Memetic algorithms for business analytics and data science: a brief survey. In: Moscato P, de Vries N, editors. Business and consumer analytics: new ideas. Cham: Springer; 2019. p. 545–608.
- [17] Chen W, Guo F, Wang FY. A survey of traffic data visualization. Intelligent Transportation Systems, IEEE Transactions on. 2015;16(6):2970–2984.

- [18] Sławińska M. The problem of imputation of the missing data from the continuous counts of road traffic. *Archives of civil engineering*. 2015;61(1):131-145.
- [19] Alippi C, Boracchi G, Roveri M. On-line reconstruction of missing data in sensor/actuator networks by exploiting temporal and spatial redundancy. *Proceedings of the International Joint Conference on Neural Networks*; 2012 Jun 10-15; Brisbane, Australia. IEEE; 2012. p. 1-8. doi: 10.1109/IJCNN.2012.6252689.
- [20] Megler VM, Tufte K, Maier D. Improving data quality in intelligent transportation systems. *arXiv: 1602.03100*. 2016 [10 p.]. Available from: <https://arxiv.org/abs/1602.03100>
- [21] Wang Y, Zhang W, Henrickson K, Ke R, Cui Z. Digital roadway interactive visualization and evaluation network applications to WSDOT operational data usage. *Research Report WA-RD 854.1*, University of Washington, 2016.
- [22] Batko WM, Stępień B. Type A standard uncertainty of long-term noise indicators. *Archives of Acoustics*. 2014;39(1):25-36.
- [23] Kephelopoulou S, Paviotti M, Ledee FA. Common noise assessment methods in Europe (CNOSSOS-EU). *Publications Office of the European Union*. 2012.
- [24] Nielsen HL. Road traffic noise: Nordic prediction method. *Nordic Council of Ministers*. 1997.
- [25] Sharma S, Lingras P, Zhong M. Effect of missing values estimations on traffic parameters. *Transportation Planning and Technology*. 2004;27(2):119-144.
- [26] Zhong M, Lingras P, Sharma S. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. *Transport Research Part C: Emerging Technologies*. 2004;12(2):139-166.
- [27] Zhong M, Sharma S, Lingras P. Genetically designed models for accurate imputations of missing traffic counts. *Transport Research Record*. 2004;1879(1):71-79.
- [28] Zhong M, Sharma S, Lingras P. Genetically-designed time delay neural networks for multiple-interval urban freeway traffic flow forecasting. *Neural Information Processing – Letters and Reviews*. 2006;10(8-9):201-209.
- [29] Lingras P, Sharma S, Zhong M. Prediction of recreational travel using genetically designed regression and time delay neural network models. *Transport Research Record*. 2002;1805:16-24.
- [30] Zhong M, Sharma S, Lingras P. Short-term traffic prediction on different types of roads with genetically designed regression and time delay neural network models. *Journal of Computing in Civil Engineering*. 2005;19(1):94-103.
- [31] Zhong M, Sharma S, Lingras P. Refining genetically designed models for improved traffic prediction on rural roads. *Transportation Planning and Technology*. 2005;28(3):213-236.
- [32] Bąkowski A, Radziszewski L, Dekys V. Modelling of road traffic noise. *Matec Web of Conferences*. 2018;157:02001. doi: 10.1051/mateconf/201815702001.
- [33] ISO 9612:2009. *Acoustics – Determination of occupational noise exposure – Engineering method*.
- [34] Frank E, Hall MA, Witten IH. *The WEKA workbench*. Online appendix for “Data mining: practical machine learning tools and techniques”. 4th ed. Morgan Kaufmann; 2016.
- [35] Breiman L, Cutler A. Random forests. [Internet] c2004-2016 [cited 2020 Sep 15] Available from: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
- [36] Hyndman R, Athanasopoulos G. *Forecasting: principles and practice*. 2nd ed. Melbourne: OTexts; 2018. [cited 2020 Nov 10] Available from: <https://otexts.com/fpp2/>