

Modeling of acoustic pressure variability at thoroughfare

Leszek Radziszewski¹, Michał Kekez^{1,*}, Andrzej Bąkowski¹, and Alžbeta Sapietova²

¹Kielce University of Technology, Faculty of Mechatronics and Mechanical Engineering, Aleja Tysiąclecia Państwa Polskiego 7, 25314 Kielce, Poland

²University of Žilina, Faculty of Mechanical Engineering, Univerzitná 8215/1, 01026 Žilina Slovakia

Abstract. The work presents the analysis of variability of acoustic pressure, calculated from values of equivalent sound level, recorded in monitoring station. The models of acoustic pressure variability were built in the form of regression trees and random forest. The analysis of accuracy of obtained models was carried out. These models were subsequently used for reconstruction of equivalent sound level for periods of monitoring station inactivity.

1 Introduction

The noise of the vehicle is produced by the engine, exhaust system, contact of tires with road surface, and also by the contact of vehicle surface with air (aerodynamic noise). The level of road traffic noise depends on: vehicle speed, road geometry, type and condition of road surface, number of trucks and other “noisy” vehicles, including motorcycles with defective mufflers, and many other parameters [1]. The monitoring stations located near the streets register sound level which depends on vehicle noise and sounds coming from other sources located in the vicinity of the road (“background noise”). Monitoring stations sometimes stop registering data, and in order to impute the missing values, the models of sound level course can be applied. Our aim was to create method for construction of equivalent sound level model for any street. By using registered values we can calculate selected descriptors of registered signal, use them for building the training dataset, create the model (using regression trees or random forests), and finally apply the model for reconstruction of missing data for single days (preferably no more than one day per one calendar week). Moreover, some of the obtained models are very transparent and easy to interpret.

2 Measurement data

The equivalent sound level values were recorded in a noise monitoring terminal comprising a sound level meter (SVAN 958A, digital, four-channel, class-1, vibration and sound meter)

* Corresponding author: m.kekez@tu.kielce.pl

Reviewers: Juraj Gerlici, Michał Śledziński

and a weather station [2], located at Krakowska Street in Kielce, Poland. Measurements were run continuously (24 hours a day) and the RMS of the A-level sound was saved in the buffer in 1 second intervals with a resolution of 0.1 dB. The measurement results were recorded every minute. Based on these measurements, equivalent sound levels were calculated for the following 24h sub-intervals: day (6-18), evening (18-22) and night (22-6). The noise annoyance caused by long-term exposure to noise is often expressed by the A-weighted equivalent sound level ($L_{Aeq,T}$) in decibels (dB), defined by formula [3]:

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

where p_0 is the reference pressure equal to 20 μ Pa, and $p_A(t)$ is A-weighted acoustic pressure course. According to ISO standard, the value of $L_{Aeq,T}$ can also be calculated from [4]:

$$L_{Aeq,T} = 10 \cdot \log \left(\frac{1}{N} \sum_{i=1}^N 10^{0.1L_{A,i}} \right) \tag{2}$$

where $L_{A,i}$ is the A-weighted acoustic pressure level measured in the i -th time interval. The recorded values of $L_{Aeq,T}$ calculated for day sub-interval (6-18) in year 2013 are shown in Fig. 1. The figure shows only “almost complete” calendar weeks (with at most one day without sound level records) excluding “irregular” weeks (containing at least one of 13 national holidays). Data shown in this figure will be used for construction of the sound level models.

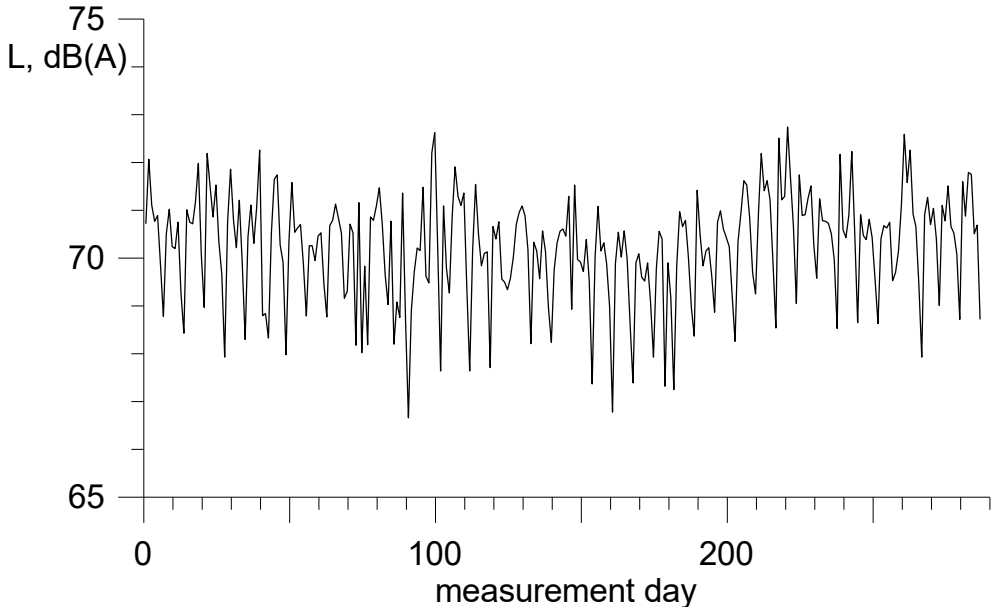


Fig. 1. $L_{Aeq,T}$ calculated from measurements for day sub-interval (6-18) for “regular” weeks in year 2013

3 Models of sound pressure and its level

Computational intelligence methods such as artificial neural networks, fuzzy systems, random forest regression, or regression trees can be applied to constructing the model. Two

approaches named as last above were employed in this study, i.e., the RandomForest algorithm in WEKA [5] and Cubist software [6].

Every training dataset used for construction of models consisted of 231 records, each of them describing one day d . Each record contained one output attribute y_1 (db_A – $L_{Aeq,T}$ value for day sub-interval) and at least one input attribute x_1 (day_of_week – number of day of the week for day d , $x_1 \in \{1,2, \dots,7\}$, where 1 denotes Monday, and 7 refers to Sunday). Other descriptors of A-weighted equivalent sound level were also calculated and subsequently used as input attributes. First of them is x_2 ($Leq_tbd - L_{Aeq,T}$ value calculated for all other 6 days in calendar week containing x_1 , except the given day x_1),

$$x_2 = 10 \log \left(\frac{1}{6} \sum_{i=1..7, i \neq x_1} 10^{0.1L_{A,i}} \right), \tag{3}$$

but when no $L_{A,i}$ value is available for any of these six days, the value from the same day of week from the nearest week is used instead. Three other descriptors, x_3 (dn2n1), x_4 (dn1p1), and x_5 (dp1p2) are calculated in the following way:

$$x_3 = L_{A,d-2} - L_{A,d-1}, \quad x_4 = L_{A,d-1} - L_{A,d+1}, \quad x_5 = L_{A,d+1} - L_{A,d+2} \tag{4}$$

In some training datasets, the equivalent sound level (in dB) was replaced by the A-weighted acoustic pressure expressed in pascals. In such datasets, x_6 was used instead of x_2 , and y_2 instead of y_1 :

$$x_6 = p_0 \cdot \sqrt{10^{0.1x_2}} \tag{5}$$

$$y_2 = p_0 \cdot \sqrt{10^{0.1y_1}} \tag{6}$$

3.1 Model no. 1

The model no. 1 uses dataset consisting of records containing only $\{y_1, x_1, x_2\}$ (or, in other words, db_A, day_of_week, and Leq_tbd) attribute values. At first, x_1 was defined as numeric attribute (model 1a), and Cubist regression tree software produced the following model (Fig. 2):

IF day_of_week > 5 THEN db_A = 33.846 – 1.295 day_of_week + 0.62 Leq_tbd
IF day_of_week ≤ 5 THEN db_A = 14.496 – 0.033 day_of_week + 0.8 Leq_tbd

Next, x_1 was defined as symbolic attribute with unordered nominal values (model 1b), and then Cubist produced another very simple and transparent model:

IF day_of_week = 7 THEN db_A = 33.936 + 0.49 Leq_tbd
IF day_of_week = 6 THEN db_A = 16.207 + 0.76 Leq_tbd
IF day_of_week ∈ {1, 2, 3, 4, 5} THEN db_A = 12.993 + 0.82 Leq_tbd

The accuracy of these models, as well as other models built by Cubist (including committee of 5 or 100 models) and RandomForest, is shown in Table 1. RandomForest models (always built using x_1 defined as numeric attribute, and with various number k of randomly selected input attributes) are rather “overfitted” to training data and have worse 10-fold cross validation accuracy than Cubist’s.

Accuracy of the models on the training and test sets was defined by the mean absolute error (MAE):

$$MAE = \frac{\sum_{i=1}^n |\bar{y}_1(i) - y_1(i)|}{n} \tag{7}$$

where $y_{1(i)}$ denotes real y_1 value taken from i -th record, $\overline{y_{1(i)}}$ denotes y_1 value calculated by the model, and n is the number of records.

Table 1. Accuracy in model no. 1, expressed by Mean Average Error (MAE) in dB

MAE [dB]	Test, 10-fold cross validation	Training data	Mondays	Tuesdays	Wednesdays	Thursdays	Fridays	Saturdays	Sundays
Cubist, model 1a	0.441	0.424	0.515	0.498	0.333	0.431	0.372	0.343	0.475
-,,- committee of 5	0.447	0.424	0.514	0.503	0.331	0.430	0.373	0.344	0.475
-,,- committee of 100	0.445	0.425	0.518	0.503	0.331	0.428	0.372	0.343	0.477
Cubist, model 1b	0.442	0.421	0.508	0.507	0.332	0.423	0.380	0.327	0.470
-,,- committee of 5	0.448	0.429	0.508	0.510	0.331	0.424	0.378	0.374	0.477
-,,- committee of 100	0.439	0.421	0.508	0.507	0.332	0.423	0.380	0.331	0.470
RandomForest $k:0,2$	0.548	0.176	0.224	0.197	0.136	0.148	0.187	0.131	0.211
RandomForest $k:1$	0.551	0.178	0.222	0.201	0.138	0.156	0.188	0.137	0.202

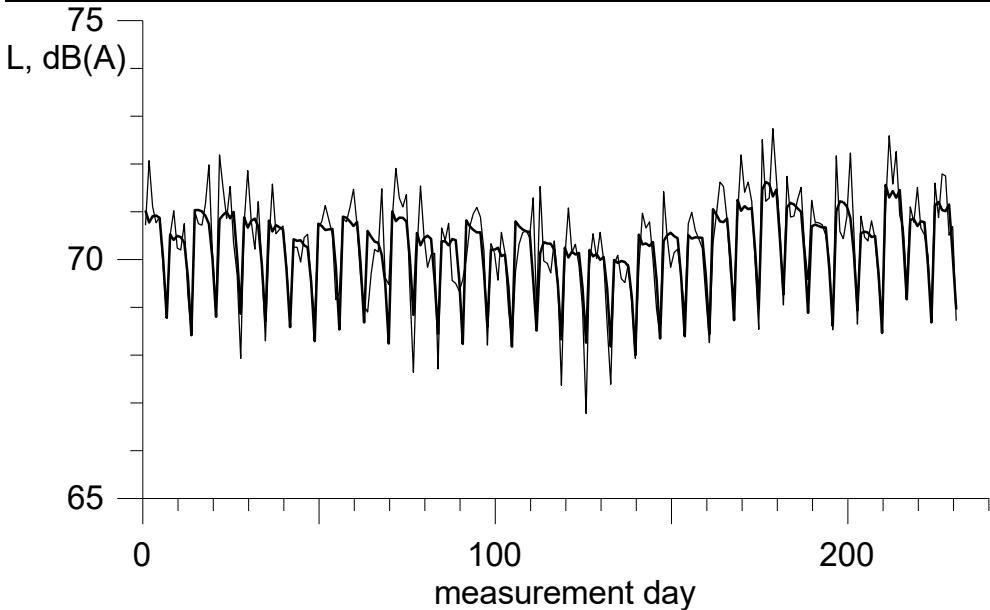


Fig. 2. $L_{Aeq,T}$ values: measured (thin line), and calculated by Cubist model 1a (thick line)

3.2 Model no. 2

The model no. 2 uses dataset consisting of records containing only $\{y_2, x_1, x_6\}$ (or, in other words, pA, day_of_week, and pA_tbd) attribute values – this means that sound pressure levels in decibels were replaced by acoustic pressure in pascals. At first, x_1 was defined as numeric attribute (model 2a), and Cubist regression tree software produced the following model (Fig. 3):

IF day_of_week > 4 THEN pA = 0.0915792 – 0.0065 day_of_week + 0.12 pA_tbd
IF day_of_week ≤ 4 THEN pA = 0.010382 – 0.00015 day_of_week + 0.89 pA_tbd

When x_1 was defined as symbolic attribute (model 2b), Cubist produced another model:

IF day_of_week = 7 THEN pA = 0.0540169
IF day_of_week = 6 THEN pA = 0.0607477
IF day_of_week ∈ {1, 2, 3, 4, 5} THEN pA = 0.0111286 + 0.87 pA_tbd

The accuracy of models is shown in Table 2. RandomForest models (always built using x_1 defined as numeric attribute) are worse than Cubist’s when comparing 10-fold cross validation accuracy. After conversion from acoustic pressure in pascals to its level in decibels, it turned out that accuracies of all models no. 2 (Table 3) were slightly worse than accuracies of their counterparts in model 1.

Table 2. Accuracy in model no. 2, expressed by Mean Average Error (MAE) in Pa.

MAE [Pa]	Test, 10-fold cross validation	Training data	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Cubist, model 2a	0.00352	0.00343	0.00412	0.00414	0.00262	0.00338	0.00361	0.00305	0.00306
Cubist, model 2b	0.00378	0.00339	0.00407	0.00416	0.00262	0.00338	0.00304	0.00330	0.00315
RandomForest $k:0,2$	0.00418	0.00137	0.00180	0.00159	0.00114	0.00117	0.00155	0.00102	0.00128
RandomForest $k:1$	0.00420	0.00139	0.00183	0.00162	0.00111	0.00125	0.00157	0.00108	0.00129

Table 3. Accuracy of model no. 2, expressed by Mean Average Error (MAE) in dB.

MAE [dB]	Test, 10-fold cross validation	Training data	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Cubist, model 2a	0.464	0.451	0.515	0.506	0.331	0.420	0.453	0.429	0.501
Cubist, model 2b	0.505	0.447	0.508	0.509	0.331	0.421	0.380	0.463	0.516
RandomForest $k:0,2$	0.549	0.180	0.227	0.192	0.145	0.147	0.191	0.143	0.213
RandomForest $k:1$	0.552	0.183	0.229	0.195	0.141	0.156	0.194	0.152	0.215

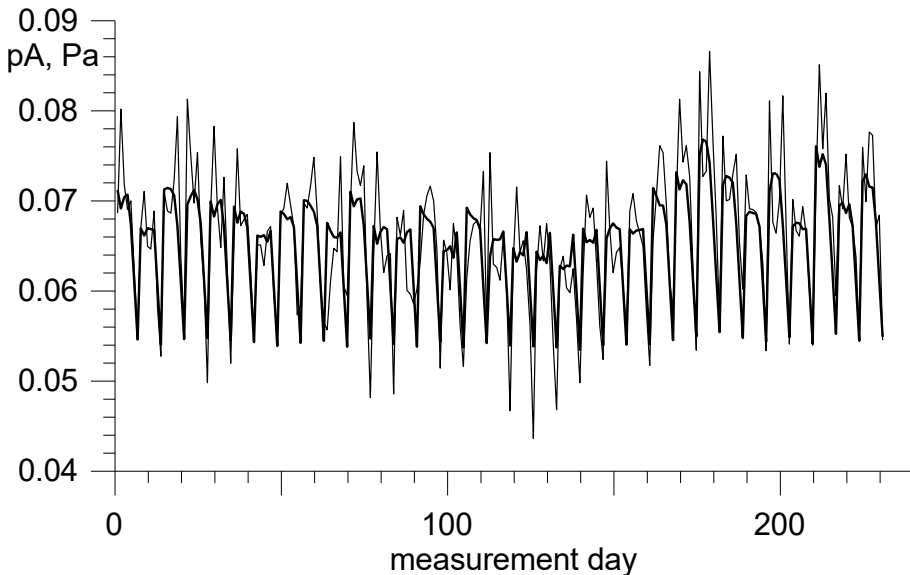


Fig. 3. p_A values: measured (thin line), and calculated by Cubist model 2a (thick line)

3.3 Model no. 3

The model no. 3 uses dataset consisting of records containing $\{y_1, x_1, x_2, x_3, x_4, x_5\}$ (or, in other words, pA, day_of_week, pA_tbd, dn2n, dn1p1, and dp1p2) attribute values. At first, x_1 was defined as numeric attribute (model 3a), and Cubist regression tree software produced the model consisting of two rules:

IF day_of_week > 5 THEN db_A = 33.853 – 1.296 day_of_week + 0.62 Leq_tbd
 IF day_of_week ≤ 5 THEN db_A = 16.114 – 0.109 day_of_week + 0.78 Leq_tbd + 0.085 dn1p1 + 0.075 dp1p2

Then x_1 was defined as symbolic attribute (model 3b), and Cubist produced another model:

IF day_of_week = 7 THEN db_A = 35.413 + 0.47 Leq_tbd – 0.071 dn2n1 + 0.073 dp1p2
 IF day_of_week = 6 THEN db_A = 9.177 + 0.86 Leq_tbd – 0.171 dn2n1 + 0.031 dp1p2
 IF day_of_week ∈ {1, 2, 3, 4, 5} THEN db_A = 17.9 + 0.75 Leq_tbd – 0.05 dn2n1 + 0.051 dp1p2

The accuracy of these models is shown in Table 4. Again, RandomForest models (e.g. Fig. 5) are better than Cubist’s (Fig. 4) on training set, but worse when comparing 10-fold cross validation accuracy. Model 3 contains more input attributes, and this fact allows RandomForest to build more accurate system in comparison to models 1 and 2 (both in terms of 10-fold cross validation and training data accuracy).

Table 4. Accuracy in model no. 3, expressed by Mean Average Error (MAE) in dB

MAE [dB]	Test, 10-fold cross validation	Training data	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Cubist, model 3a	0.441	0.419	0.500	0.510	0.310	0.426	0.370	0.344	0.475
-, - committee of 5	0.444	0.424	0.507	0.523	0.298	0.426	0.366	0.378	0.472
-, - committee of 100	0.453	0.427	0.537	0.515	0.292	0.422	0.368	0.379	0.477
Cubist, model 3b	0.446	0.415	0.526	0.492	0.309	0.417	0.386	0.305	0.470
-, - committee of 5	0.451	0.419	0.563	0.495	0.292	0.419	0.357	0.340	0.470
-, - committee of 100	0.450	0.417	0.563	0.485	0.285	0.418	0.377	0.324	0.468
RandomForest $k:0,3$	0.500	0.170	0.195	0.198	0.116	0.152	0.145	0.168	0.218
RandomForest $k:1$	0.501	0.180	0.217	0.185	0.117	0.168	0.152	0.165	0.257
RandomForest $k:2$	0.502	0.177	0.208	0.197	0.109	0.153	0.158	0.166	0.246
RandomForest $k:4$	0.512	0.172	0.202	0.193	0.125	0.161	0.158	0.146	0.222
RandomForest $k:5$	0.516	0.168	0.187	0.188	0.131	0.145	0.147	0.151	0.229

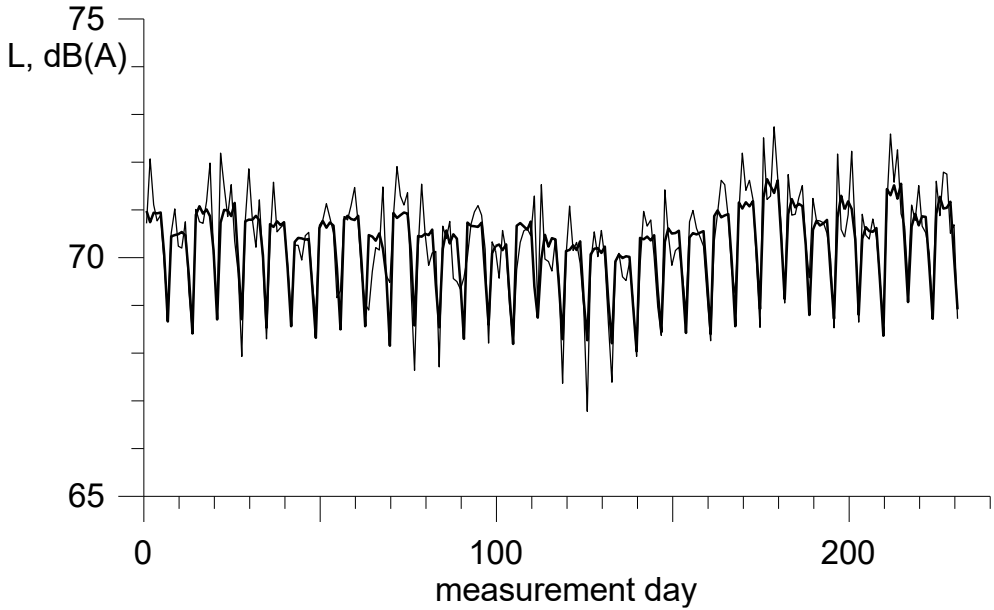


Fig. 4. $L_{Aeq,T}$ values: measured (thin line), and calculated by Cubist model 3b (thick line)

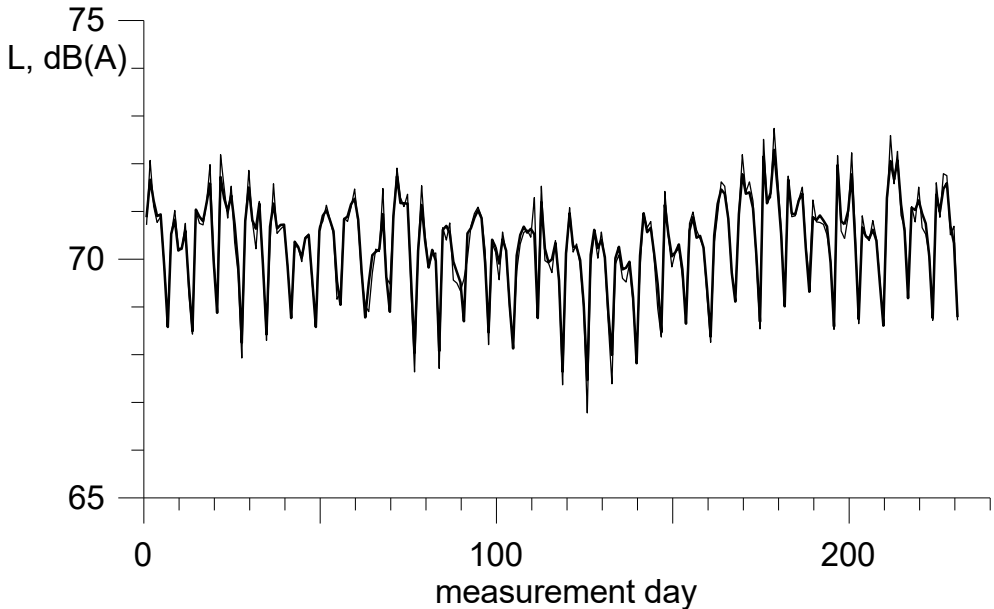


Fig. 5. $L_{Aeq,T}$ values: measured (thin line), and calculated by RandomForest with random selection of $k=5$ input attributes, model 3a (thick line)

4 Conclusions

All presented models (1, 2, and 3) have good accuracies (mean average error of prediction of $L_{Aeq,T}$ is lower than 0.6 dB for all datasets and all days of week). These models can be

successfully applied to imputation of one missing value of $L_{Aeq,T}$ of day sub-interval during every week. To calculate all descriptors used in models, it is recommended to wait until the end of the week containing missing value. When this condition is not met, the x_2 value will be calculated using data partially from other week, and the x_4 and x_5 attributes will have unknown values, but the models presented in this paper will still allow imputation of missing $L_{Aeq,T}$ values, presumably with lower accuracy.

Cubist achieves better accuracy than RandomForest when evaluated by 10-fold cross validation, but worse when evaluated on training dataset. Best accuracy on training dataset (Mean Average Error 0.168 dB) is achieved by model 3a produced by RandomForest with random selection of $k = 5$ input attributes. However, better estimate of accuracy is 10-fold cross validation, where the best result (MAE 0.439 dB) is achieved by model 1b produced by Cubist committee of 100 regression trees. The best transparency is obtained by Cubist model 1b, which contains only 2 simple rules.

References

1. S. Abo-Qudais, A. Alhiary. Effect of Traffic Characteristics and Road Geometric Parameters on Developed Traffic Noise Levels. *Canadian Acoustics*, **32**(4): 43-50, (2005)
2. A. Bąkowski, L. Radziszewski, Z. Skrobaccki. Assessment of uncertainty in urban traffic noise measurement. *Procedia Engineering*, **177**: 281-288 (2017)
3. W. Batko, B. Stępień. Type A Standard Uncertainty of Long-Term Noise Indicators. *Arch. Acoust.*, **39**(1): 25-36 (2014)
4. ISO 9612:2009, *Acoustics – Determination of occupational noise exposure – Engineering method*.
5. <http://www.cs.waikato.ac.nz/ml/weka/>
6. <https://www.rulequest.com/cubist-info.html>