CETRA^{2026} 27-30 May 2026, Bol, Island Brač, Croatia
9th International Conference on Road and Rail Infrastructure

REGRESSION ANALYSIS OF TRAIN OPERATING PARAMETERS INFLUENCING RAILWAY NOISE EMISSION BASED ON FIELD MEASUREMENTS

Marcin Wrótny
Lublin University of Technology, Faculty of Civil Engineering and Architecture, Poland

Abstract

The aim of this study is to assess the influence of train operating parameters on railway noise emission levels under real operating conditions. The analysis is based on field measurement data collected at five sites located in different regions of Poland, under uniform measurement conditions (microphone positioned 7.5 m from the track axis and 1.2 m above the railhead). In total, over 300 train pass-bys were included in the dataset, covering various train categories. Each pass-by was recorded with simultaneous measurements of train speed, the equivalent continuous A-weighted sound level (L_{Aeq}), and meteorological parameters. A set of regression models was developed to determine the effect of selected operating parameters on railway noise emission. The modelling approaches included classical linear regression (LM), weighted least squares regression (WLS), and regularized methods such as LASSO, Ridge, and Elastic Net. The models were evaluated in terms of their goodness of fit (R^2, RMSE), parameter stability, and the statistical significance of the effects (p-values). Additionally, an interaction term between train speed and train type was incorporated to better capture the real operational variability of railway noise emission processes.

Keywords: railway noise, train operating parameters, noise emission modelling, regularization methods

1 Introduction

Rail transport is an important part of the logistics system in Europe and worldwide. It enables efficient transport of passengers and goods over medium and long distances. In recent years, its development has accelerated due to the growing need for alternatives to road transport and increasing environmental awareness. Compared to road transport, railways produce lower carbon emissions per transport unit. However, despite these advantages, railway infrastructure also generates negative environmental impacts, with noise emission being one of the most significant. Long-term exposure to high noise levels can reduce the quality of life of people living near railway lines and may lead to health problems and increased stress [1, 2]. Railway noise is a complex phenomenon influenced by many factors. Train speed, rolling stock type, track condition, traffic intensity, and local environmental characteristics all affect the level of emitted sound. Because of this variability, accurate noise prediction requires analytical methods that can capture the main influencing factors while avoiding unnecessary model complexity. Railway noise has been widely studied in literature from different technical and environmental perspectives. Nevertheless, modelling this phenomenon remains challenging. Noise levels depend not only on operational parameters but also on infrastructure condition and terrain characteristics [3].

DOI: https://doi.org/10.5592/CO/CETRA.2026.1912

The large number of potential explanatory variables increases model complexity and highlights the need for effective variable selection and dimensionality reduction techniques. Traditionally, railway noise modelling has been mainly used for spatial planning, infrastructure design, and the preparation of strategic noise maps, especially in accordance with Directive 2002/49/EC (Environmental Noise Directive – END) [4]. This directive requires EU Member States to regularly assess environmental noise and to prepare action plans in areas where permissible limits are exceeded. Several standardized calculation methods have been developed for this purpose. Although methods such as RMR, Schall 03, and CNOSSOS-EU are widely used in regulatory and planning practice, they are less flexible when applied directly to empirical measurement data. Therefore, this study focuses on regression-based models, which allow the influence of specific variables on noise levels to be quantified and their predictive performance to be evaluated using indicators such as RMSE, MAE, and R^2.

2 Methods

2.1 Field measurement procedure

The purpose of the study was to quantify noise levels generated by passing trains and to identify the main factors influencing railway noise emission under real operating conditions. During each train pass-by, additional parameters were recorded, including train speed, passage time at the measurement cross-section, and rolling stock category. This approach enabled a direct link between the measured acoustic indicators and the technical and operational characteristics of the trains. A total of 323 individual train pass-byes were included in the dataset. Measurements were carried out at five locations distributed across five voivodeships in Poland. The selection of sites allowed for the inclusion of regional variability in infrastructure and operating conditions. All locations were situated along active railway lines and met the requirements for standardized acoustic measurements. Noise levels were recorded continuously at each measurement location throughout the entire observation period. Subsequently, during the data analysis stage, individual train pass-by events were identified and their corresponding time intervals were extracted. The equivalent sound levels were then determined separately for each identified pass-by, based on its specific duration. The measurement procedure followed the PN-EN ISO 3095:2013-12 standard [5], using the pass-by method. Noise measurements were carried out using a Class 1 sound level meter (SVAN 971, Svantek) equipped with a microphone and windscreen. The measurement system was calibrated before and after each measurement session using a Class 1 acoustic calibrator (SV 36, Svantek). Microphones were installed at a height of 1.2 m above the railhead and at a horizontal distance of 7.5 m from the track axis (figure 1). Meteorological parameters were recorded simultaneously to ensure consistent measurement conditions.

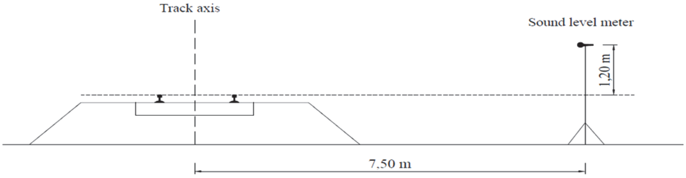

Figure 1 Schematic of the measurement point showing the location of the sound level meter

During the measurements, train speed was determined using a radar device installed in a consistent configuration across all measurement locations. The radar was positioned 7.5 m from the track axis at a height of 1.5 m above ground level.

To reduce the cosine error associated with radar measurements, speed readings were taken when the train was at a distance of at least 50 m from the measurement cross-section, ensuring a minimal angle between the direction of travel and the radar beam.

2.2 Measurement parameters and indicators

In railway noise analysis, individual train pass-bys are evaluated separately, since each train type and operating condition may result in different noise emission levels. For each recorded pass-by, the equivalent continuous A-weighted sound level was calculated for the duration of the acoustic event only. The same measurement procedure was applied at all research sites. This ensured consistency and allowed the creation of a uniform and reliable empirical dataset. The collected data were then used to extract the main explanatory variables for further statistical analysis. These variables formed the basis for the development of regression models and for assessing the influence of selected factors on railway noise levels. The variables included in the regression analysis are summarized in table 1.

Table 1 Variables used in the regression analysis

Symbol	Variable	Description
Y	L_{AeqT}	Equivalent continuous sound level of a single train pass-by [dB]
X1	Train speed	Train speed at the moment of pass-by [km/h]
X2	Passage time	Train passage time at the measurement cross-section [s]
X3	Train type	Rolling stock category (P_P, P_SN, P_SS, P_T, P_W)*
X4	Location	Measurement site (L_1 – L_5)
X5	Temperature	Average air temperature [°C]
X6	Humidity	Relative air humidity [%]
X7	Atmospheric pressure	Average atmospheric pressure [hPa]
X8	Wind speed	Average wind speed [m/s]

*Categories: high-speed train (P_P), self-propelled passenger train new type (P_SN), self-propelled passenger train old type (P_SS), freight train (P_T), passenger train with wagons (P_W)

2.3 Regression modelling approaches

Multiple linear regression (OLS) was applied to assess the combined influence of explanatory variables on railway noise levels. The method estimates parameters by minimizing the sum of squared residuals and offers straightforward interpretation, although it may be sensitive to multicollinearity and heteroscedasticity. To address these limitations, regularization techniques were also considered. LASSO performs both parameter estimation and automatic variable selection by shrinking some coefficients to zero, leading to more parsimonious models. Ridge regression reduces the effect of multicollinearity by shrinking coefficients without eliminating predictors. Elastic Net combines both approaches, enabling simultaneous variable selection and stabilization of correlated predictors. Additionally, Weighted Least Squares (WLS) was applied to account for heteroscedasticity. By assigning weights to observations, WLS improves the efficiency and reliability of parameter estimates when error variance is not constant.

3 Results

3.1 Descriptive analysis

Before developing the regression models, a descriptive analysis of the numerical variables was performed. The aim of this step was to examine the basic characteristics of the dataset, including value ranges, central tendencies, and variability. This preliminary assessment also allowed for the identification of potential deviations from normality, which could affect model selection and performance. Table 2 summarizes the main descriptive statistics for the dependent variable (Y) and selected numerical independent variables (X1, X2, X5, X6, X7, X8). The categorical variables X3 (Train type) and X4 (Location) were not included in this summary, as standard numerical measures are not applicable to non-numeric data.

Table 2 Descriptive statistics for numerical variables (n = 323)

Variable	Unit	Mean	Std. dev.	Median	Min–Max
Y	dB	85.4	4.7	84.9	74.9 – 99.7
X1	km/h	85.0	25.9	88.0	22.0 – 161.0
X2	s	14.0	12.3	8.0	4.0 – 60.0
X5	°C	18.9	4.9	19.0	4.0 – 27.0
X6	%	61.1	10.2	58.0	40.0 – 83.0
X7	hPa	1013.5	6.1	1014.3	1001.3 – 1023.0
X8	m/s	4.1	0.4	4.1	3.5 – 4.8

To examine the relationships between the noise level (Y) and the remaining numerical variables, a scatterplot matrix was prepared (figure 2). The matrix presents pairwise relationships between all quantitative variables (Y, X1, X2, X5, X6, X7, X8), enabling a visual assessment of potential linear or nonlinear trends, clusters of observations, and possible outliers. The clearest associations are observed between the dependent variable and train speed (X1) as well as passage time (X2). These relationships are consistent with the physical nature of railway noise generation, confirming that train movement characteristics play a key role in determining the emitted sound level.

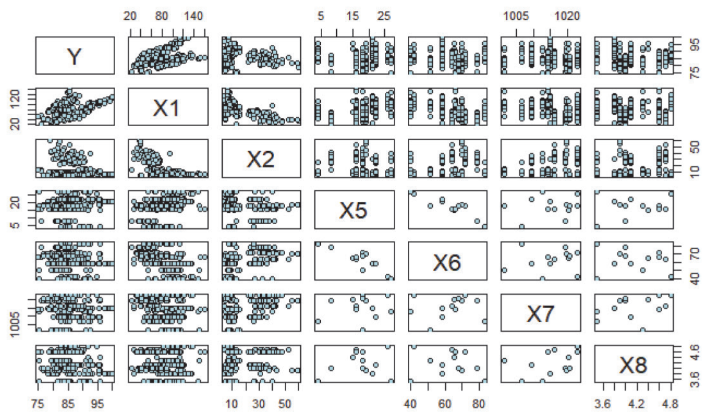

Figure 2 Scatterplot matrix of the dependent variable Y (noise level) and independent variables (X1, X2, X5–X8)

3.2 Categorical variable analysis

After examining the distributions of the numerical variables and their mutual relationships, the analysis was extended to categorical factors, namely train type (X3) and measurement location (X4). The objective was to evaluate their influence on the noise level (Y). As these variables are categorical in nature, nonparametric analysis of variance methods were applied to compare noise level distributions across different groups. This approach allowed for the identification of statistically significant differences without assuming normality of the data. Figure 3 presents boxplots illustrating: a) the distribution of the equivalent sound level according to rolling stock category and b) the distribution of noise levels across measurement locations.

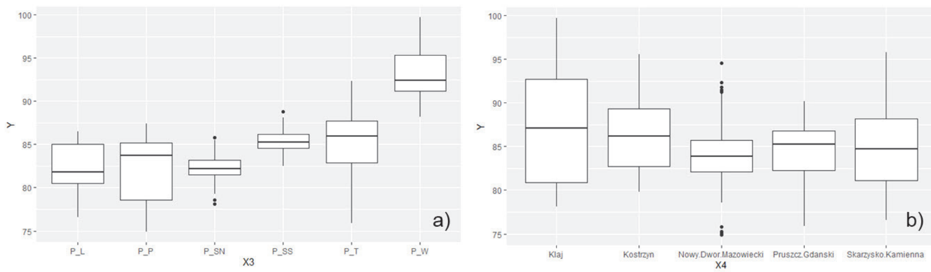

Figure 3 Distribution of equivalent sound levels (L_{AeqT}) across rolling stock categories (a) and measurement locations (b)

Clear differences in noise levels were observed between rolling stock categories. The highest levels were recorded for P_W (passenger trains with wagons) and P_T (freight trains), while the lowest were associated with P_SN (new-type self-propelled passenger trains), confirming the strong influence of train type. The second part of figure 3 shows substantial variability across measurement locations. At some sites (L_1), sound levels exceeded 90 dB, whereas at others (L_3) noticeably lower values were observed, likely reflecting local environmental and operational conditions. To verify these differences, the Kruskal–Wallis test was applied. In both cases, very low p-values led to rejection of the null hypothesis of identical distributions:

- train type (X3): $x^2 = 192.88$, df = 5, p-value $< 2.2^{-16}$
- location (X4): $x^2 = 12.65$, df = 4, p-value = 0.01.

These results confirm that both train type and measurement location significantly affect noise levels and should be included as predictors in further modelling.

3.3 Regression modelling results

The modelling process began with classical multiple linear regression (OLS), including all numerical and properly coded categorical predictors. The model showed a good fit and statistical significance. Meteorological variables (humidity, temperature, atmospheric pressure, wind speed) were not significant when controlling for operational factors and were therefore excluded. Residual diagnostics indicated no deviations from normality or autocorrelation, but heteroscedasticity was detected using the Breusch–Pagan and Goldfeld–Quandt tests. As variance-stabilizing transformations were ineffective, alternative regression methods were applied. LASSO and Elastic Net eliminated meteorological variables, confirming their limited relevance, while Ridge regression reduced multicollinearity but retained all predictors.

To address heteroscedasticity, Weighted Least Squares (WLS) regression was implemented, resulting in improved variance stability and homoscedastic residuals. A comparison of all models (OLS, WLS, LASSO, Ridge, Elastic Net) based on R^2, RMSE, and MAE (figure 4) showed largely consistent results, with WLS providing the best balance between predictive performance and compliance with model assumptions

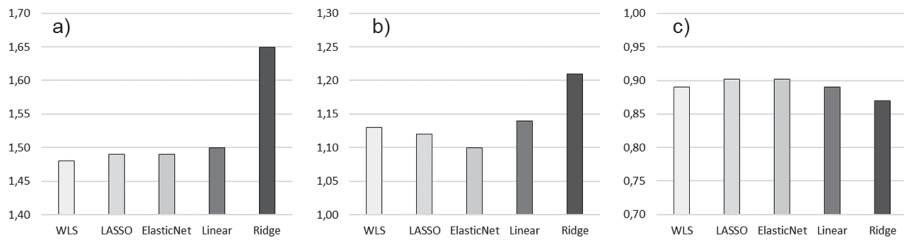

Figure 4 Comparison of regression models based on: a) RMSE, b) MAE, c) R^2

4 Discussion and conclusion

In the conducted study, five regression techniques were evaluated: classical linear regression (LM), weighted least squares (WLS), LASSO, Ridge, and Elastic Net. All models were estimated using a reduced dataset from which statistically insignificant environmental variables had been removed in order to enhance parameter stability and overall model consistency. The selection of the final model was not based exclusively on goodness-of-fit indicators. In addition to predictive accuracy, attention was given to interpretability, compliance with regression assumptions, robustness, and practical applicability in environmental assessments. The regularization methods (LASSO, Ridge, and Elastic Net) did not lead to a meaningful improvement in predictive performance compared with classical regression. Ridge regression showed slightly lower performance (R^2 = 0.87, RMSE = 1.65), while LASSO and Elastic Net, although capable of variable selection, retained most predictors already present in the linear model. As a result, these approaches did not provide substantial additional value and introduced greater interpretational complexity. In contrast, the weighted least squares (WLS) model proved to be the most appropriate alternative. This method was implemented in response to the heteroscedasticity detected in the OLS residual analysis. The application of weights significantly improved the variance structure of residuals. The Breusch–Pagan test no longer indicated heteroscedasticity, confirming better compliance with regression assumptions. For these reasons, WLS regression was selected as the final model. It combines high explanatory power with improved statistical validity while preserving clear interpretation of regression coefficients. The model can therefore be effectively applied to predict railway noise levels based on key operational and locational variables. The final form of the WLS regression model is presented below:

$$\begin{aligned} Y=73,\,14+0,\,13\cdot X1-0,\,04\cdot X2-5.06\cdot X3P_P-3,\,94{*}X3P_{SN}- \\ -\,1,\,33\cdot X3P_{SS}+6,\,63\cdot X3P_T+7,\,68\cdot X3P_W-0,\,31\cdot X4L_2+ \\ +\,0,\,22\cdot X4L_3-0,\,81\cdot X4PL_4+0,\,40\cdot X4L_5 \end{aligned} \tag{1}$$

where X1 denotes train speed (km/h), X2 represents passage time (s), and X3P_* are binary (dummy) variables corresponding to train categories (P – high-speed Pendolino, SN – new-type self-propelled passenger train, SS – old-type self-propelled passenger train, T – freight train, W – passenger train with wagons). Variables X4L_* are dummy variables representing measurement locations, with L_1 treated as the reference category and the remaining locations expressing effects relative to this baseline.

References

[1] Wrótny, M., Bohatkiewicz, J.: Impact of Railway Noise on People Based on Strategic Acoustic Maps, Sustainability, 12 (2020) 14, 5637, DOI: 10.3390/su12145637

[2] Wrótny, M., Bohatkiewicz, J.: Traffic Noise and Inhabitant Health – A Comparison of Road and Rail Noise, Sustainability, 13 (2021) 13, 7340, DOI: 10.3390/su13137340

[3] Thompson, D.: Railway Noise and Vibration: Mechanisms, Modelling and Means of Control, Elsevier, 2008.

[4] Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 Relating to the Assessment and Management of Environmental Noise, Official Journal of the European Communities, L 189/12, 2002.

[5] Polish Committee for Standardization: PN-EN ISO 3095:2013-12 – Railway Applications – Acoustics – Measurement of Noise Emitted by Railbound Vehicles, 2013.

[6] Nur, A.R., Jaya, A.K., Siswanto, S.: Comparative Analysis of Ridge, LASSO, and Elastic Net Regularization Approaches in Handling Multicollinearity for Infant Mortality Data in South Sulawesi, Jurnal Matematika Statistika dan Komputasi, 20 (2023) 2, pp. 311–319, DOI: 10.20956/j.v20i2.31632