Neural Comput & Applic (2018) 29:1045–1057
https://doi.org/10.1007/s00521-016-2625-9

ORIGINAL ARTICLE

Artificial neural network model for ground vibration amplitudes prediction due to light railway traffic in urban areas

G. Paneiro[1] · F. O. Durão[1] · M. Costa e Silva[1] · P. Falcão Neves[1]

Received: 4 March 2016 / Accepted: 12 October 2016 / Published online: 21 October 2016

Abstract The growth of density and circulation speed of railway transportation systems in urban areas increases the importance of the research issues of the produced environmental impacts. This study presents a field data analysis, obtained during monitoring campaigns of ground vibration, due to light railway traffic in urban areas, based on the artificial neural network (ANN) approach, using quantitative and qualitative predictors. Different ANN-based models, using those predictors, were evaluated/trained and validated. Using several criteria, including those that measures the possibility of ANN overfitting (RR^2) and complexity (AIC), the best ANN model was successfully obtained for Lisbon area. This model, with 16 input elements (quantitative and qualitative predictors), 2 neurons on the hidden layer with a hyperbolic tangent sigmoid transfer function, and 1 neuron on the output layer considering a linear transfer function, has 0.9720 for the coefficient of determination and 0.5293 for the sum squared error.

Keywords Ground vibrations · Artificial neural networks · Railway traffic · Qualitative predictors · Ratio of coefficients of determination

✉ G. Paneiro
gustavo.paneiro@tecnico.ulisboa.pt

F. O. Durão
fdurao@tecnico.ulisboa.pt

M. Costa e Silva
matilde.horta@tecnico.ulisboa.pt

P. Falcão Neves
pfalcaoneves@tecnico.ulisboa.pt

[1] DECivil/CERENA, Técnico Lisboa, Universidade de Lisboa, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal

Abbreviations

AIC	Akaike information criteria
ANOVA	Analysis of variance
ANN	Artificial neural network
FFT	Fast Fourier transform
FTA	Federal Transit Administration
FRA	Federal Railroad Administration
ISO	International Organization for Standardization
LFSS	Sum of the squared lack-of-fit error
MSE	Mean squared error
PESS	Sum of the squared pure error
SBIC	Schwarz–Bayesian information criteria
SSE	Sum of the squared error
SST	Total sum of squares

List of symbols

A	Ground vibration peak amplitude
B	Building type (qualitative predictor)
Cn	Category of a qualitative variable
$\boldsymbol{C}$	Coefficient of determination matrix between pairs of distinct columns of matrix $\boldsymbol{X}$
c	Vibration propagation velocity
E	Energy
f	Wave frequency
$f(s)$	Transfer function
G	Dominant geology (qualitative predictor)
hls	Hidden layer size
D	Distance
Kn	Indicator variables
PVS	Peak velocity sum
Q	Quality factor
$R2$	Coefficient of determination
R^2_{Max}	Maximum coefficient of determination
RR^2	Ratio of coefficients of determination

s	Input of the transfer function
T	Rail track type (qualitative predictor)
$\boldsymbol{t}(\boldsymbol{\gamma}^h\boldsymbol{x}+\boldsymbol{\theta}^h)$	Vector of transfer functions associated to two neurons of the hidden layer
V	Train circulation speed
$\boldsymbol{X}$	Input matrix
$\boldsymbol{x}$	Vector of the input variable
xi	Input variable
$\boldsymbol{Y}$	Output vector
Yi	Sum of a predicted mean response
α	Damping factor
ΔE	Dissipated energy
γ	Parameter (or weight)
$\boldsymbol{\gamma}^O$	Weights between the hidden layer and the output layer
γ_{ij}	Adjustable parameters
$\boldsymbol{\gamma}^h$	Weights between the inputs and the hidden layer
θ^O	Bias of the output vector
$\boldsymbol{\theta}^h$	Bias vector for the hidden layer
θj	Bias

1 Introduction

In the last 50 years [1], the use of railway transportation has grown, not only in terms of density but also in terms of circulation speed to answer the demands of modern societies. The consequent increase in ground vibration production and its related environmental impacts, produced by railway transport in urban areas, has been an important research issue, which has been increasingly attended at geotechnical projects in the past decade [2].

The ground-borne vibrations generated by railway trains circulation, which causes undesirable environmental concerns and can result in human distress and real estate losses, are the result of the vehicle forces that act into the railway track, and consequently, vehicles weight and irregularities/discontinuities at the wheel/rail interface play an important role on how these dynamic phenomena will propagate outward the railway track [3].

What concerns to vehicle type, Wilson et al. [4] demonstrated that proper design of bogie suspension can significantly reduce levels of ground vibration, and according to [5], vehicles with soft primary suspension produce lower levels of vibration than vehicles equipped with stiffer suspensions.

Kouroussis et al. [3] refer that railway train track can be divided into three main groups: high-speed, moderate-speed, and low-speed railway tracks. Vibrations generated by high-speed trains are mainly dependent on quasi-static track deflection, because these kinds of tracks are characterized by very high-quality rolling surfaces. On the other hand, light and low transit vehicles (which are vehicle types contemplated in the present paper) are characterized by a low speed and a relatively high density of singular rail surface defects, and thus, dynamic track deflection mainly contributes to ground wave generation. The moderate-speed trains present excitation mechanisms that are a combination of those experienced on both high-speed and urban railway lines.

An accurate estimate of ground vibrations generated by railway traffic requires the knowledge of input parameters, such as the dynamic track and ground characteristics, for which in situ testing may be required. For this, the Federal Railroad Administration (FRA) and the Federal Transit Administration (FTA) of the US Department of Transportation have developed a set of empirical procedures to predict vibration levels due to railway traffic [6, 7]. Also, the standard ISO 14837-1:2005 [8] offers guidance on prediction for a range of circumstances (e.g., to support the assessment of the effect on human occupants, considering the activities developed in the target structures, the effect on sensitive equipment, and the risk of damage to building structures) and covers all forms of wheel and rail systems, from light-rail to high-speed trains and freight, circulating on grade rail systems, on elevated structures, and in tunnels.

Several numerical models have been developed for the prediction of ground vibrations due to surface railway traffic [9–14] and in tunnels [15–17]. Kouroussis et al. [3] offer a good scoping of some vibration prediction models and point out particular aspects in each approach.

In the particular case of prediction of ground vibrations due to railway traffic in urban areas, it can be highlighted the work of [18] which established a link between numerical and empirical predictions as FRA procedures allows for the so-called hybrid predictions based on empirical data and numerical results.

Connolly et al. [19] developed a new scoping railway vibration prediction model (named ScopeRail), capable of predicting three international vibration metrics in the presence of layered soils, using a high-accuracy 3D model to populate a database of vibration records. From this database, the relationships between key railway variables were established by means of a machine learning approach, using an artificial multilayer perceptron neural network architecture. This model was later developed to instantly predict ground vibration levels in the presence of different train speeds and soil profiles [20].

Actually, artificial neural network models, based on nonlinear transfer functions and massive parallel networks composed of many computational elements connected by links of adjustable (variable) weights, are suitable when prediction models involve several predictors that influence the response variable. However, due to the nature of the

work developed by [19, 20], there is a need to gather a considerable amount of quantitative information for propagation media characterization. Because the access to the ground is, particularly in urban areas and for existing infrastructures, difficult or near impossible, quantitative data for propagation media characterization are equally difficult to obtain.

In this study, quantitative field data, acquired during monitoring works of ground vibrations due to railway traffic in urban areas, combined with qualitative data are used to build and validate mathematical models, based on the artificial neural network (ANN) approach, to predict ground vibration amplitude on buildings, with reduced prediction errors. In order to make use of the available qualitative data, a suitable codification methodology for the qualitative predictors is applied.

The best mathematical model is selected from several candidate models using criteria expressing a balance between adjustment quality (residual analysis) and model complexity (number of adjusted parameters/weights).

To illustrate this approach, this paper shows the application of ANN's to predict the vibrations amplitude due to railway traffic in Lisbon area, significantly reducing, by this way, the prediction errors, comparing to other alternative prediction methodologies [21].

All the computations were performed using MATLAB application, and all the computation issues related with ANN-based regression model used the specialized functions of the Neural Network Toolbox.

2 Characterization of sources, propagation media, and targets

From the generation point of view, it is widely known that, after the release of any kind of energy in the ground, there is a radial propagation of superficial and volumetric waves that reaches nearby structures and populations. Consequent vibration amplitudes will depend of several factors:

(a) Released energy amount on the source.
(b) Distance between the source and the point where the effects are measured.
(c) Transmission and dissipate properties of the ground.
(d) Building dynamic strength.

It becomes clear that mathematical models for vibrations prediction should include predictors related with the source, the propagation media and receivers. In this way, it is imperative to characterize them, but, particularly in urban areas, the application of techniques for the determination of quantitative data is not easy or almost impossible. For that, it is important to consider other forms to determine the explanatory variables for prediction models.

According to [22], many authors suggest a correlation between earthquake magnitudes and explosive blasts, particularly nuclear blasts, concluding that the amount of released energy is the most important factor in these dynamic events. Energy should be the most suitable parameter for source characterization.

In this context, an easily applicable parameter to characterize the dynamic event produced by train circulation is the kinetic energy at the considered monitoring point.

However, sometimes, train circulation speed is not possible to be directly measured and registered/recorded, especially in underground circulation. Figure 1 shows a schematic representation of a general track model presented by [23]. In this railway track model, the train consists of coaches, bogies, and axles, which are all modelled as a mass–spring–damper system. Springs and dampers connect them to each other. Vibration amplitude peaks due to axles are therefore expected to be identified from waveform analysis. Thus, train circulation speed can be simply calculated by knowing the distance between each axle.

One way to characterize the propagation media is the determination of attenuation. There are, mainly, three ways to quantify vibration attenuation with distance:

1. A dimensionless number n that is an exponent of the distance, D, between two points, 1 and 2, where peak

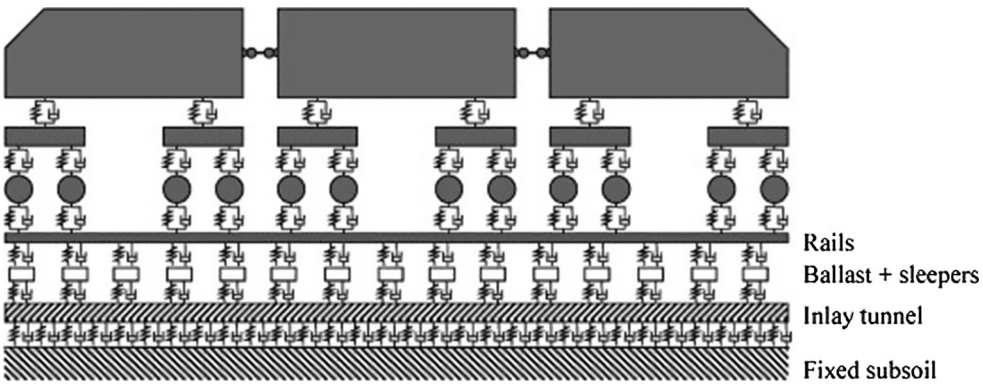

Fig. 1 Railway track model [23]

magnitude A is monitored. The relation between both magnitudes may be expressed by [21]:

$$A_2 = A_1 D^{-n}. \quad (1)$$

2. A factor α that involves more variables (propagation velocity, c, and wave frequency, f) in the relationship between two amplitudes in the form of [24]:

$$A_2 = \frac{A_1}{D} e^{-\alpha(D-1)}. \quad (2)$$

It can be demonstrated that [24]:

$$\alpha = \frac{f}{2c}\frac{\Delta E}{E}. \quad (3)$$

Here E is the energy supplied to the movement and ΔE is the dissipated part between the two points.

3. Through the form of a dimensionless quality factor Q, to measure the internal friction, which is the ratio between the energy of seismic wave, E, and the dissipated energy per wave cycle, ΔE, or [24]:

$$Q = 2\pi\frac{E}{\Delta E}. \quad (4)$$

As referred by [25], building type plays an important role in how vibrations are perceived. According to this author, when the frequency of the generated event approaches the building's natural frequency, vibration amplitude increases (resonant frequency).

However, as previously referred, the determination of these quantitative ground and target characteristics is often a very difficult or almost impossible task. For that and in contrast to most of the published works, it is suggested the application of qualitative predictors to characterize propagation media and target buildings.

For the cases of railway train circulation and for simplification, it is considered that propagation media consist on the railway track type and ground type (demanded by geology) and, for building characterization, a qualitative classification regarding the correspondent construction age. Each of these qualitative explanatory predictors may have many categories and to use them for mathematical models, codification is needed.

3 Applied methodology in data analysis

3.1 Qualitative predictors and multicollinearities

According to [26, 27], in order to represent qualitative predictor categories, binary indicator variables or dummy variables can be used. For the reference category, all indicator variables are 0, and for each one of the other exclusive categories, the indicator variable is 1. In a generalized form, Table 1 represents the codification of a qualitative explanatory variable with n categories.

Table 1 Binary codification of a qualitative predictor

Category	Indicator variables			
	$K1$	$K2$	…	Kn_{-1}
$C1$ (reference category)	0	0	…	0
$C2$	1	0	…	0
$C3$	0	1	…	0
…	…	…	…	…
Cn	0	0	…	1

Besides the application of other methodologies, these binary indicator variables are easy to use and are widely employed [26].

After this codification, a rectangular matrix $\boldsymbol{X}$, of size $n \times k$, represents all observed values, and each matrix row represents an observation. The number, k, of matrix columns is equal to the sum of the number of quantitative predictors and the total number of dummy variables created for the categories codification.

It is well known [26, 27] that uniqueness and quality of the estimation exercise of the regression model parameters are critically dependent on the detection and elimination of multicollinearities or perfect linear relations between the columns of the $\boldsymbol{X}$ matrix (linearly dependent columns) or strongly correlated columns (correlation coefficients' absolute value greater than or equal to 0.99).

One simple technique, used in this study, for multicollinearity detection and elimination is based on the computation of a squared matrix $\boldsymbol{C}$, $k \times k$, consisting of the coefficients of determination between pairs of distinct columns of matrix $\boldsymbol{X}$. Thus, one of the columns from the columns pairs with absolute values higher than 0.99 is removed from matrix $\boldsymbol{X}$. Another alternative technique consists on the application of Gauss–Jordan elimination method [28], with the reduction of matrix $\boldsymbol{X}$ to a reduced row echelon form; it identifies the linearly independent columns with the automatic rank determination of the $\boldsymbol{X}$ matrix.

3.2 Artificial neural networks (ANN)

3.2.1 ANN architecture

Like other nonlinear regression models, ANNs can approximate, with the required precision, the following general unknown relationship:

$$Y_i = f(\boldsymbol{X}_i, \boldsymbol{\gamma}) + \varepsilon_i. \quad (5)$$

An observation (of the response variable) Yi is the sum of a predicted mean response $f(\boldsymbol{X}i, \boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a parameter

(or weight) vector, and the prediction error term εi. As usual, the error terms are assumed to be realizations of independent random variables of zero mean and constant variance.

A multilayer artificial neural network (ANN) consists, essentially, of several layers of nonlinear transformations of linear combinations of input variables. The output variables of a layer are the input variables of the next layer. The most common ANN architecture consists of two layers: the hidden layer and the output layer. The input variables of the hidden layer are the explanatory variables, and the output variables of the output layer are the dependent variables. The hidden layer consists of a configurable number of processing elements, also known as transfer functions or neurons. The number of processing elements of the output layer is equal to the number of dependent variables. The input of each processing element is a linear combination of the inputs of the layer including a constant term, known as bias. The coefficients of the linear combinations are the adjustable weights of the ANN.

The most used transfer functions in ANN's are the logarithmic sigmoid and the hyperbolic tangent sigmoid, defined as follows:

$$\text{Logarithmic sigmoid function: } f(s) = \frac{1}{1+e^{-s}} \tag{6}$$

$$\text{Hyperbolic tangent sigmoid function: } f(s) = \frac{e^{2s}-1}{e^{2s}+1} \tag{7}$$

where $s = \sum_{i=1}^{n} \gamma_{ij} x_i + \theta_j$ is the input of the transfer function of the j-th neuron, with $x_i, i = 1, 2, \ldots, n$, the input variables of the layer, $\gamma_{ij}, i = 1, 2, \ldots n$, the adjustable parameters, also known as weights, θj the bias and n the number of input variables.

Figure 2 is a representation of a neuron, input and output links, weights, and transfer function.

In order to increase neural network training efficiency and to avoid transfer function saturation, input and target datasets need to be normalized before their use [29]. In this work, because of the use of two different transfer functions and domains, two normalization methods are applied. For the ANNs with logarithmic sigmoid transfer functions, the normalized inputs/targets are zero mean and unity variance variables, which means that the domain falls in the interval between 0 and 1. For those ANNs that use hyperbolic tangent sigmoid functions, the normalized inputs/targets fall in the range $[-1, 1]$.

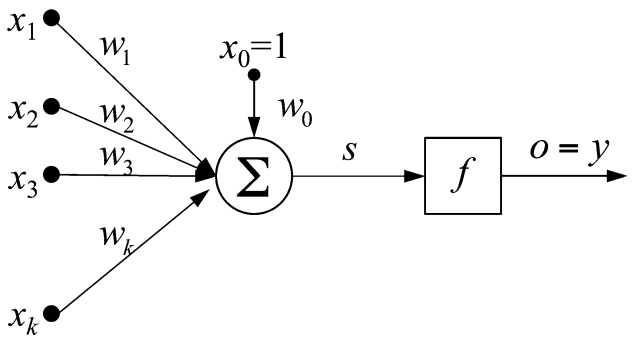

Fig. 2 General representation of a neuron of an artificial neural network

3.2.2 *Optimization algorithms*

The optimal values of the ANN weights and biases are the solution of a nonlinear parameter estimation problem consisting in the minimization of a metric of the prediction errors. In this study, sum of the squared errors (SSE) between the predicted outputs and the corresponding observed outputs for all pairs of input and output patterns will be used as the error metrics.

To solve the nonlinear parameter estimation problem (the so-called ANN training), three nonlinear least squares optimization algorithms may be applied: scaled conjugate gradient method, Bayesian regulation algorithm [30], and, the most widely used, the Levenberg–Maquardt method. In this work, the Levenberg–Maquardt algorithm will be considered, taking into account the successful results obtained in several other knowledge areas.

As in any nonlinear optimization algorithm, the ANN training starts with the initialization of the weights and bias values. The most used initialization procedure assigns random values to the ANN parameters.

The objective function, sum of squared errors (SSE), of the optimization problem is usually characterized by multiple local minima, and the mentioned optimization algorithms only ensures local optimal solutions. To deal with this issue, multistart strategies or meta-heuristics methods [31] (simulated annealing, genetic algorithms, etc.) can be applied. A momentum parameter that may be used to prevent a training algorithm from converging prematurely to a local minimum or saddle point is provided in the Neural Network Toolbox for MATLAB. More information about multistart methods and strategies can be obtained in [32–35].

3.3 Model selection criteria

As seen before, the parameters of neural networks models are adjusted by a training method, using an optimization algorithm to minimize the metric of the prediction errors. However, one of the problems that may occur during neural network training is called overfitting.

To improve network generalization and avoid overfitting, many methods can be applied, but the *early stopping* method is one of the most used. A full description of this method can be found in Beale et al. [29].

In this way, the first criterion, for model selection, is the sum of squared errors (SSE) obtained for the test set. In this work, SSE was selected instead of mean square error (MSE) in order to be directly compared to the remaining alternative criteria.

The coefficient of determination ($R2$) is another popular criterion that can be used to evaluate the capability of the models to explain the non-random variability of the observed data. However, $R2$ cannot avoid overfitting issues because it does not take into account any information about the pure error, also called experimental error. According to several authors, e.g., [26], the analysis of variance (also called ANOVA table) summarizes the orthogonal decomposition of the total sum of squares, SST, (in terms of sum of squares and mean squares). An independent estimation of the variance of the pure (random) error or experimental error can be obtained when replications of the observations are available, supporting a further decomposition of the sum of the squared error (SSE) as the sum of squared lack-of-fit error (LFSS), an additional sum of squares, and sum of the squared pure error (PESS) [36]. One of the selection criteria (specially used in multiple linear regressions) that is commonly used to assess the adequacy of a model is the lack-of-fit testing based on F statistic. Large values of this statistic are evidence against the adequacy of the model.

Following this logic, it is possible to establish the maximum coefficient of determination (R^2_{Max}) for the adjustment and, therefore, avoid overfitting issues, using the relation given by Eq. (8).

$$R^2_{\text{Max}} = 1 - \frac{\text{PESS}}{\text{SST}} \tag{8}$$

Thus, a simple way to evaluate a fitted ANN, in terms of its coefficient of determination and avoiding overfitting, is to get the relation between the coefficient of determination and its correspondent maximum coefficient of determination from a given model. This ratio of coefficients of determination (RR^2), given by Eq. (9), gives us information about the capability of the fitted model to explain the non-random variability of the observed data, measured as a fraction, <1.0, of the maximum coefficient of determination.

$$\text{RR}^2 = \frac{\text{SST} - \text{SSE}}{\text{SST} - \text{PESS}} \tag{9}$$

A ratio RR^2 higher than 1 may mean an overfitting situation due to the considered ANN architecture.

Other useful model selection criteria for neural network modelling are Akaike information criteria *(AIC)* and Schwarz–Bayesian information criteria (SBIC). These are two popular criteria that not only measures of model adjustment quality but also provide penalties in terms of the model complexity. More detailed information about AIC and SBIC criteria can be obtained in [26]. For the present study, AIC criteria were used.

4 Case study

4.1 Background and predictors definition

The described methodology for modelling building and validation was applied in vibration amplitude monitoring of Lisbon's light-rail system. This rail system is mostly composed by underground rail tracks thought the city of Lisbon, Portugal. Field works consisted in 20 monitoring points, coincident to different buildings in the surroundings of the railway tracks (Fig. 3).

In the monitoring campaigns, Instantel's standard tri-orthogonal geophones, with frequency range from 2 to 250 Hz, vibration velocity ranges up to 254 mm/s, and resolution of 0.00788 mm/s, were installed near the vibration sources (tracks) and in the buildings, in order to obtain vibration amplitudes and correspondent seismograms for each event (Fig. 4). The signals were stored, first, in Instantel Minimate Plus devices, which can store a maximum of approximately 52 events in the complete data recording, and in the post-processing stage, all these events were downloaded to a computer, using its dedicated Instatel's Blastware software.

Besides the 172 samples collected from the monitoring system, where vibration amplitude was recorded for each event, it was possible to collect additional data, in each monitoring point, about the train track type, building type, and dominant geology from the geological map from the Lisbon region, which corresponds to the categories of each qualitative predictor.

The vibration amplitude is here characterized by the peak velocity sum (PVS in mm/s) which corresponds to the maximum vector sum of the three measured vibration amplitude components (longitudinal, transversal, and vertical) in a time instant of the recorded event.

About the quantitative variables, these are the distance (D, in m) from the source point and the monitoring point and the train circulation speed, (V, in km/h), obtained indirectly by seismograph analysis for each monitored event.

The use of train circulation speed instead of its mass (or kinetic energy as stated before, which is an important parameter that influences the vibration level) is justified by the fact that monitored trains are from the same type, and therefore, their mass is approximately equal, which delivers a constant.

Referring to the qualitative predictors, the categories for the rail track type qualitative predictor are (5 categories and 4 dummy variables. $T1$–$T4$): ballasted (reference

Fig. 3 Location of the monitoring points in Lisbon area (image obtained in DigitalGlobe 2015, http://www.earth.google.com)

category), concrete, concrete with damping system, floating trays track, and concrete viaduct.

The considered categories for dominant geology are (9 categories and 8 dummy variables, $G1$–$G8$): volcanic complex (reference category), loam and limestone, sand, clay and sandstone, clay and loam, clay, limestones, sandstones, Benfica's formation, and Forno do Tijolo's clays.

Regarding the track unevenness and defects, the correspondent predictor could not be considered. However, a suitable inspection of the track unevenness and the presence of track defects can be done in order in the future to reflect these parameters in the prediction model.

For the building type, the considered categories are (3 categories and 2 dummy variables, $B1$ and $B2$): centenary building (reference category), old building, and current building.

After qualitative predictor codifications and multicollinearity elimination, as referred in Sect. 3.1, the input $\boldsymbol{X}$ matrix is defined as well as the output vector $\boldsymbol{Y}$.

4.2 Artificial neural network candidate models

According to the previous description, ANN-based nonlinear regression was performed considering the general architecture represented in Fig. 5. The number of input and output units in a neural network is generally determined by the dimensionality of the dataset. For the present study the input consists of the quantitative predictors (distance and train circulation speed) and the coded categories of the qualitative predictors, resulting in input vectors of 16 components. The output corresponds to the values of PVS.

The number of hidden units is a free design parameter that can be adjusted to give the best predictive performance. Note that the dimension of the hidden layer controls the total number of parameters (weights and biases) in the network, and it might expect that in a maximum likelihood setting there will be an optimum dimension that gives the best generalization performance, corresponding to the trade-off (optimum balance) between under-fitting and over-fitting. However, it can be said that, in order to preserve the degree of freedom for the regression, the maximum number of parameters for the fitted function should be lower than the number of observations, in order to preserve the degrees of freedom of the adjusted function. For this, the optimum dimension of the hidden layer, in order to guarantee a fitting correspondent to the global minimum, can lead to a very complex model.

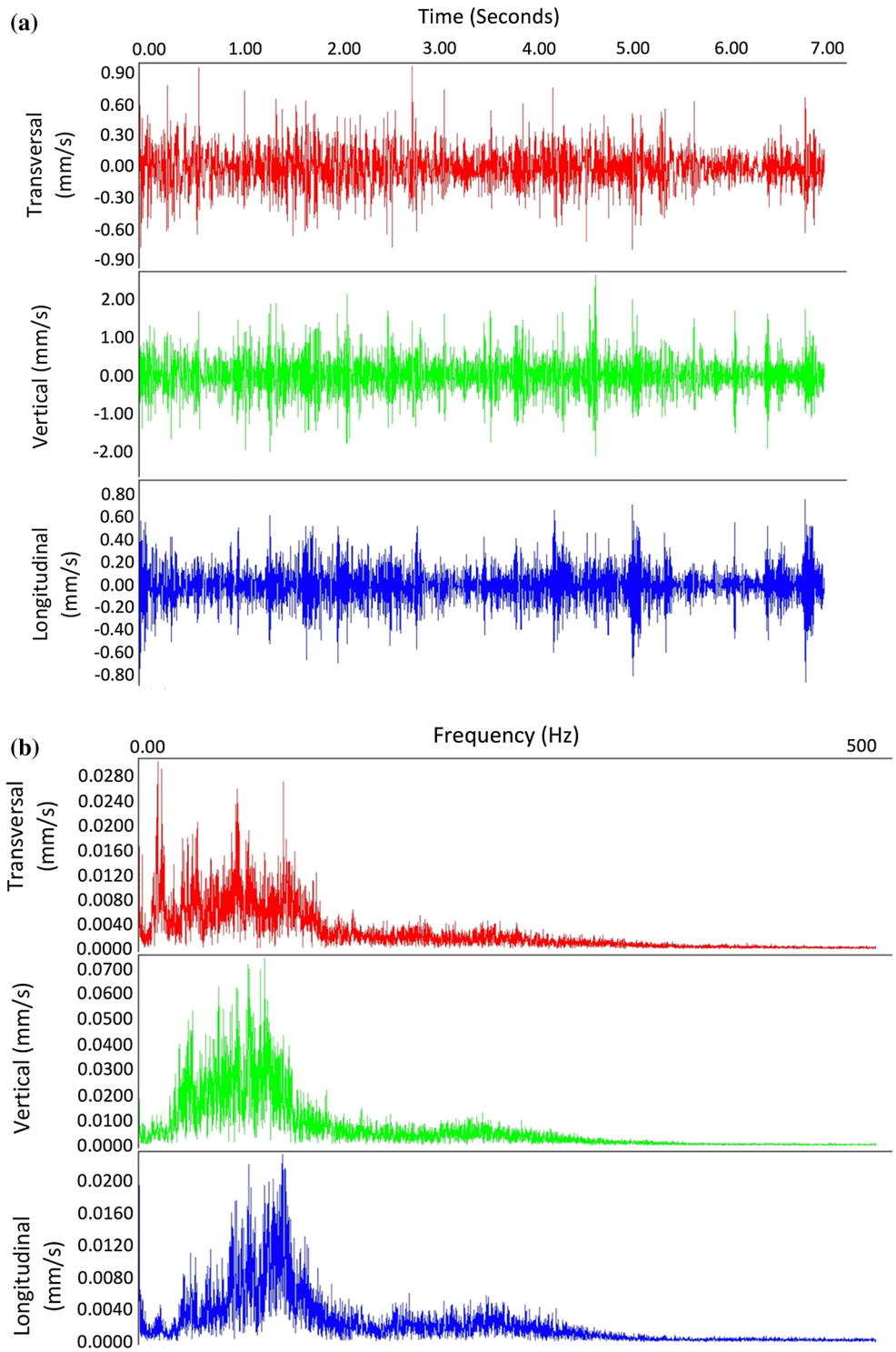

Fig. 4 Time history **a** of the three registered components and the correspondent frequency FFT **b** of a recorded event

From the architectural point of view, a combination of transfer function types and the number of neurons in one hidden layer determined the candidate ANN models.

Concerning the transfer function type associated with the neurons in the hidden layer, logarithmic sigmoid and hyperbolic tangent sigmoid functions were considered. For the output layer, the transfer function associated with the single neuron was always the linear identity function (output = input, $f(s) = s$). Finally, the specifications of the candidate ANN models were completed with the definition of the number of neurons in the hidden layer (from one to four neurons).

In order to avoid large complexity, the considered ANN architectures have a maximum hidden layer size of four

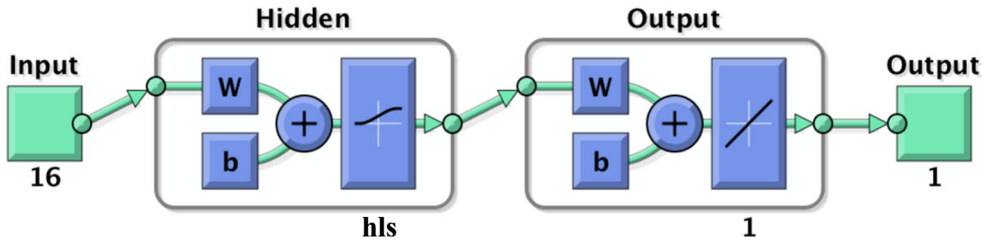

Fig. 5 MATLAB representation of the general architecture of the ANN-based model, using logarithmic sigmoid transfer function in the hidden layer. Here, hls represents the number of neurons on the hidden layer. The transfer function of the output layer is the linear identity function (adapted from [29])

neurons because the model complexity has a high tendency to increase with higher sizes of the hidden layer. As it can be showed in the next paragraphs, for hls > 2, there is an increase in the AIC criterion, which means that the increase in complexity does not overcome adjustment improvements.

As referred before, the Levenberg–Maquardt algorithm was chosen to solve the nonlinear least squares optimization problems.

4.3 Determination of the quasi-global minimum model

One of the main problems in ANN fitting procedure, besides the overfitting issue, is the determination of the model parameters that corresponds to the global minimum of the sum of squared errors. Most of the times, the training procedure by Levenberg–Maquardt or other training algorithms, with early stopping method, guarantees only the convergence to the nearest local minimum, which may be not the global minimum. Due to the random nature of the weights and biases initialization procedure, the nearest local minimum may differ significantly, in terms of the adjusted values of weights and biases, from one initialization to another initialization. Nevertheless, the literature about these issues does not offer too much alternatives besides the retraining of the considered ANN architecture and, as previously referred, the number of neurons in the hidden layer should be optimized in order to ensure the best generalization performance.

To overcome these issues, not only the overfitting but, especially, the complexity and local minimum problems, the eight defined ANN candidate architectures were retrained several times applying a multistart strategy.

Figure 6 shows the result for the minimum SSE obtained for different number of starts. It can be seen that for one neuron in the hidden layer, the minimum SSE is generally larger then the other architectures (with a higher number of neurons in the hidden layer). However, it can also be seen that, for more than 100 starts, the minimum SSE tends to be constant which means that

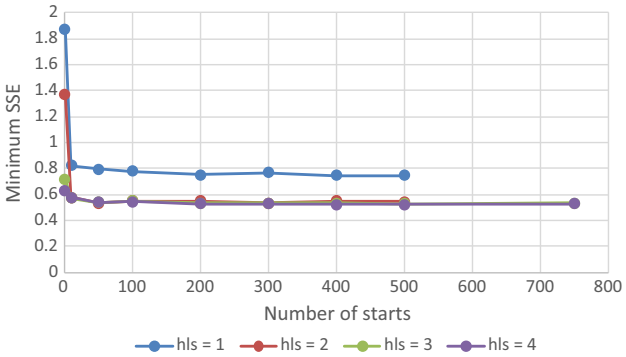

Fig. 6 Evolution of the minimum SSE obtained for different ANN architectures and for different numbers of starts (different starting points)

for more than 100 restarts, the quasi-global minimum can be obtained.

For this reason and in order to select the best ANN architecture that considers a quasi-global minimum, for each architecture, a multistart strategy with 1000 starts and retrains were considered for each ANN model.

4.4 Selection of the best fitted ANN model

Table 2 presents the results in terms of the values of $R2$, SSE minimum AIC, and RR^2 for each trained architecture.

For RR^2 determination, the sum of the squared pure error was estimated, knowing that the degrees of freedom of the pure error is 96, providing an independent estimation of the variance of the pure experimental/observational error. In this way, the obtained sum of the squared pure error is 0.477.

According to Table 2, considering the RR^2 criterion, it can be seen that, in all cases, the correspondent values are lower than 1, which means that there is no overfitting. The best fitted models are models 4, 6, and 7 which have a RR^2 value of 0.9972, meaning that only 0.28 % of the non-random data variability is not explained by the selected model.

However, considering the AIC criteria, the minimum value for this information criterion is achieved with model 6 (AIC value of -920.00) which means that, when

Table 2 Obtained results from the trained ANN architectures

Model	Tranfer function	hls	*df*	*R*2	SSE minimum	AIC	RR^2
1	Logsig	1	35	0.9606	0.7465	−895.2258	0.9854
2		2	51	0.9711	0.5469	−914.4284	0.9962
3		3	67	0.9712	0.5458	−880.7794	0.9963
4		4	83	0.9721	0.5277	−852.5205	0.9972
5	Tansig	1	35	0.9600	0.7580	−892.6054	0.9848
6		2	51	0.9720	0.5293	−920.0044	0.9972
7		3	67	0.9721	0.5285	−886.2764	0.9972
8		4	83	0.9719	0.5322	−851.0690	0.9970

df degrees of freedom of the regression

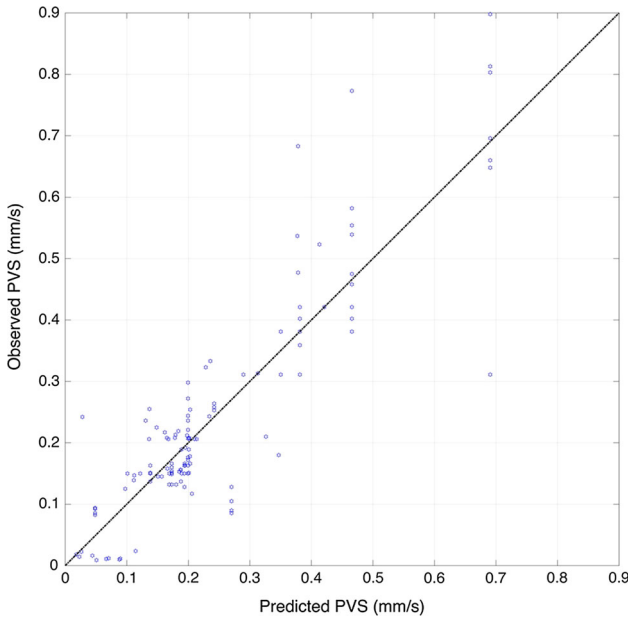

Fig. 7 Scatter plot of observed versus predicted values of PVS for model 6

predicted PVS's as a function of the distance (D) and the train speed (V).

The ANN-based regression model can be written compactly, using the matrix–vector notation, as follows:

$$\text{PVS} = \boldsymbol{\gamma}^{O}\boldsymbol{t}\left(\boldsymbol{\gamma}^{h}\boldsymbol{x} + \boldsymbol{\theta}^{h}\right) + \theta^{O}. \tag{10}$$

Here, $\boldsymbol{x}$ represents the column vector, 16×1, of the input values for each pattern of quantitative and qualitative (categories) predictors, $\boldsymbol{t}\left(\boldsymbol{\gamma}^{h}\boldsymbol{x} + \boldsymbol{\theta}^{h}\right)$ is the column vector, 2×1, of transfer functions (hyperbolic tangent sigmoid) associated to two neurons of the hidden layer, $\boldsymbol{\gamma}^{O}$ is the row vector, 1×2, of the weights for the connections between the two neurons of hidden layer and the single neuron of the output layer, $\boldsymbol{\gamma}^{h}$ is a matrix, 2×16, of the weights associated with the connection between the inputs and the two neurons of hidden layer neurons, $\boldsymbol{\theta}^{h}$ is the column vector, 2×1, of the biases for the two neurons of the hidden layer, and θ^{O} is the bias (scalar) applied in the output layer. For model 6, the values, rounded to 2 decimal places, for the referred matrices and vectors are:

$$\boldsymbol{\gamma}^{h} = \begin{bmatrix} -3.10 & 0.38 & 4.48 & 0.37 & -4.22 & -0.45 & -2,37 & -1,06 & -6,28 & 0,26 & -1,31 & 0,63 & -4,25 & -0,53 & -7,56 & -0,54 \\ -1.06 & -1.40 & 1.03 & 1.39 & -3.56 & -1.02 & -5,33 & -0,88 & -2,06 & -0,11 & -4,00 & -0,20 & -0,33 & -0,17 & -4,43 & 0,08 \end{bmatrix}$$
$$\boldsymbol{\theta}^{I} = \begin{bmatrix} -2,95 \\ -0,65 \end{bmatrix}; \quad \boldsymbol{\gamma}^{O} = [4.17 \quad -4.11]; \quad \theta^{O} = 0.16 \tag{11}$$

coefficient of determination and model complexity are considered together, model 6 is the best of the eight candidate models. So, model 6, with $RR^2 = 0.9972$ and SSE = 0.5293, is selected as the best fitted model, and the scatter plot of the observed versus predicted PVS values is presented in Fig. 7. Figure 8 shows the observed and

Equation (10), together with the adjusted weights and biases, organized as matrices and vectors as presented in Eq. (11), is a full mathematical specification of the adjusted ANN-based regression model.

For a full assessment of the statistical quality of the retained model, the analysis of the residuals is mandatory.

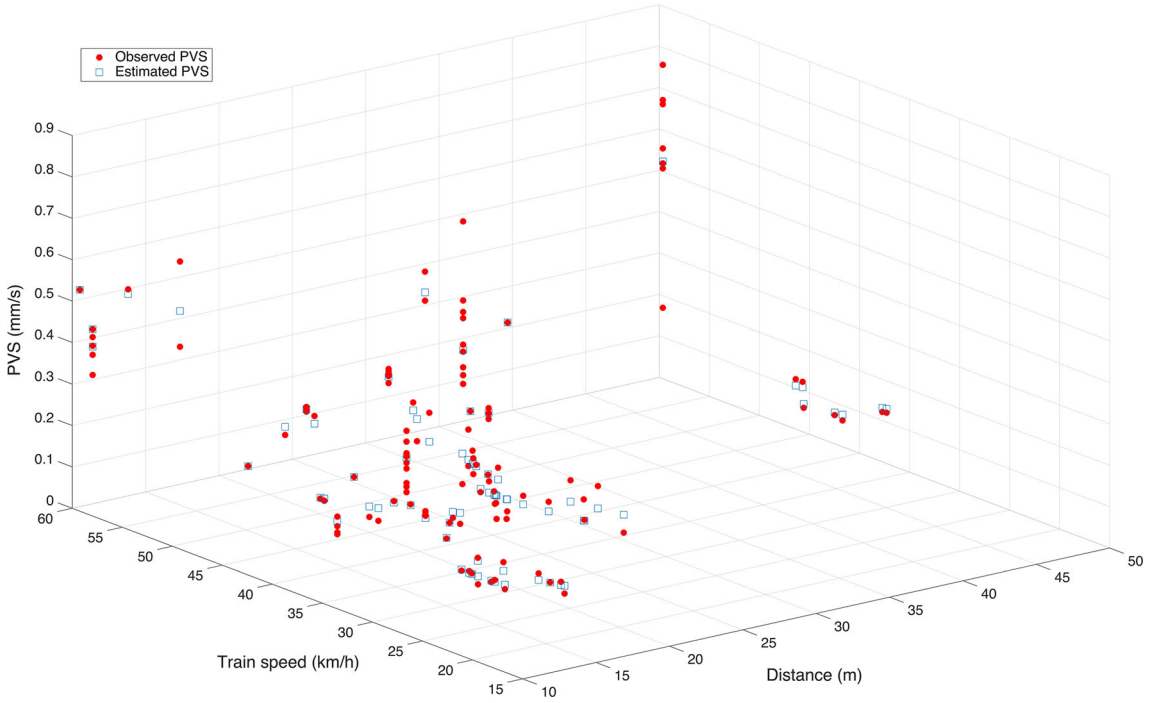

Fig. 8 Observed and predicted PVS as a function of the distance and the train speed

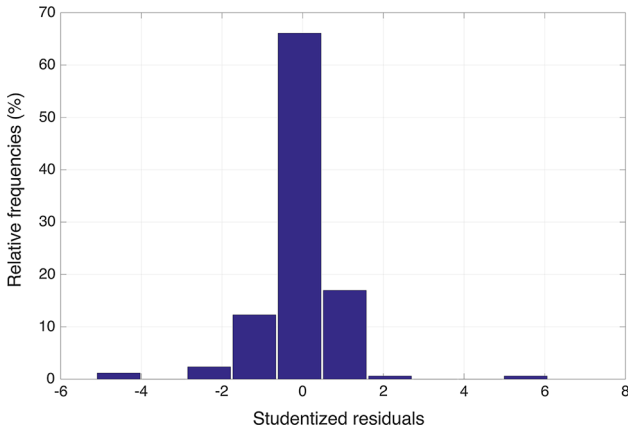

Fig. 9 Studentized residuals for the selected ANN model

The histograms of the residuals (studentized residuals) and a few other residual plots are used as graphical tools to assess standard assumptions about the error component/term of the model.

Concerning the studentized residuals, as shown in Fig. 9, approximately 95 % of the studentized residuals are between -2 and 2 limits.

The residual plots in Fig. 10a, b shows clear random pattern (or at least absence of any pattern) of the studentized residuals versus the two quantitative variables, distance and train speed. The absence of a systematic pattern in the residuals as function of the quantitative explanatory variables characterizes the retained model as an adequate one, that is, the specification of the functional form of the model is correct and that no important explanatory variable has been omitted.

The histogram in Fig. 9, together with the horizontal bands at ± 2 drawn (in red) in Fig. 10a, b, supports the additional assumption that the distribution of the studentized residuals are approximately normal with mean zero and variance one.

From Fig. 10a, b also is apparent that, for a fixed value of the corresponding predictor, the distribution of the residuals is almost symmetric around the value zero. Residuals with absolute value greater than 3 or 4 may be regarded as indicative of possible outliers in the dependent variable PVS.

5 Conclusions

A field data analysis, obtained during monitoring works of ground vibration due to railway traffic in urban areas, based on ANN approach, was performed. Limitations of using only field quantitative data, particularly those used for characterization of the vibration propagation media, were overcome by including qualitative data available about of the propagation media.

The ANN approach presented in the paper, involved the specification, calibration and validation of 8 candidate ANN-based models. The used criteria were based on the results of the training performance (sum of squared errors) and information criteria such as AIC and the relation

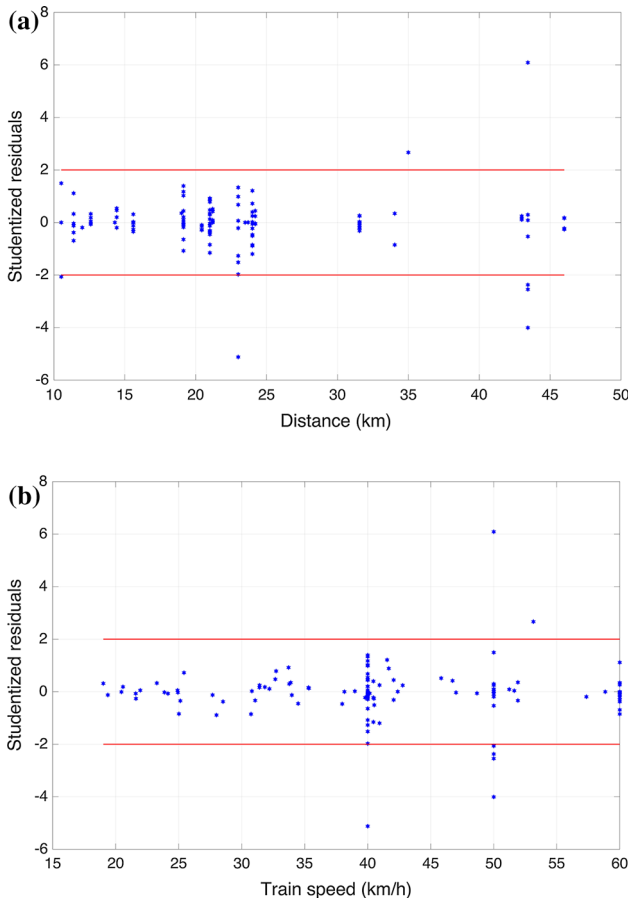

Fig. 10 Studentized residuals as a function of the distance D (**a**) and the train speed V (**b**)

between the coefficient of determination and its maximum value, in order to evaluate the fit and overfitting possibility. This methodology was successfully applied in Lisbon area, where an ANN-based low-complexity model was obtained (Eqs. 10, 11).

According to the obtained results, the best fitted ANN-based model (model 6) consists of 16 input elements, 2 neurons in the hidden layer with hyperbolic tangent sigmoid transfer function, and 1 element in the output layer, with a linear identity transfer function. The correspondent SSE value is 0.5293, the AIC value is −920.00. Also, the regression between the observed and predicted PVS values (Fig. 7) shows a coefficient of determination $R2$ of 0.9720 and the ratio of coefficients of determination (RR^2) of 0.9972.

Besides the good results revealed by the considered criteria, the residual analysis of the selected model shows that, for lower values of PVS (those in the limits of human discomfort), the model offers good prediction results. However, for higher values of PVS, there is an intrinsic random variability of the data.

Nevertheless, the histogram of the studentized residuals for the best fitted ANN-based model (Fig. 9) demonstrates good adequacy of the model to the observed data.

The presented results also show that the consideration of qualitative data as additional inputs of ANN-based regression model can be easily and successfully made and leads to better predictions compared to multiple linear regression methods [21] where the obtained sum of squares of the error is 0.951 and AIC of −855.844.

Given the data considered in the present study, which states the present problem domain, the application of this model can be limited for further locations (where one or more conditions can be different from those considered in this study). To overcome this issue, new monitoring campaigns should be made, in order to update the model.

Moreover, as it is well known, the use of these ANN approaches can be easily updated and integrated in vibration real-time control systems of railway transport systems in urban areas, in order to minimize the consequent environmental impacts.

References

1. Connolly DP, Kouroussis G, Laghrouche O, Ho CL, Forde MC (2014) Benchmarking railway vibrations—track, vehicle, ground and building effects. Constr Build Mater 92:64–81. doi:10.1016/j.conbuildmat.2014.07.042
2. Connolly DP, Marecki GP, Kouroussis G, Thalassinakis I, Woodward P (2015) The growth of railway ground vibrations problems—a review. Sci Total Environ. doi:10.1016/j.trgeo.2015.09.002
3. Kouroussis G, Connolly DP, Verlinden O (2014) Railway-induced ground vibrations—a review of vehicle effects. International journal of rail transportation 2(2):69–110. doi:10.1080/23248378.2014.897791
4. Wilson GP, Saurenman HJ, Nelson JT (1983) Control of ground-borne noise and vibration. J Sound Vib 87(2):339–350. doi:10.1016/0022-460X(83)90573-4
5. Nelson JT (1996) Recent developments in ground-borne noise and vibration control. J Sound Vib 193(1):367–376. doi:10.1006/jsvi.1996.0277
6. Hanson CE, Towers DA, Meister LD (2005) High-speed ground transportation noise and vibration impact assessment. HMMH report 293630-4. U.S. Department of Transportation, Federal Railroad Administration, Office of Railroad Development, October
7. Hanson CE, Towers DA, Meister LD (2006) Transit noise and vibration impact assessment. Report FTA-VA-90-1003-06. U.S. Department of Transportation, Federal Transit Administration, Office of Planning and Environment, May
8. International Organization for Standardization, ISO 14837-1:2005. Mechanical vibration—ground-borne noise and vibration arising from rail systems, part 1: general guidance
9. Grundmann H, Lieb M, Trommer E (1999) The response of a layered half-space to traffic loads moving along its surface. Arch Appl Mech 69(1):55–67. doi:10.1007/s004190050204
10. Sheng X, Jones CJC, Petyt M (1999) Ground vibration generated by a load moving along a railway track. J Sound Vib 228(1):129–156. doi:10.1006/jsvi.1999.2406
11. Picoux B, Le Houédec D (2005) Diagnosis and prediction of vibration from railway trains. Soil Dyn Earthq Eng 25:905–921. doi:10.1016/j.soildyn.2005.07.002

12. Lombaert G, Degrande G, Kogut J, François S (2006) The experimental validation of a numerical model for the prediction of railway induced vibrations. J Sound Vib 297(3–5):521–535. doi:10.1016/j.jsv.2006.03.048
13. Lombaert G, Degrande G, Vanhauwere B, Vandeborght B, François S (2006) The control of ground-borne vibrations from railway traffic by means of continuous floating slabs. J Sound Vib 297(3–5):946–961. doi:10.1016/j.jsv.2006.05.013
14. Galvín P, Domínguez J (2007) Analysis of ground motion due to moving surface loads induced by high-speed trains. Eng Anal Bound Elem 31:931–941. doi:10.1016/j.enganabound.2007.03.003
15. Metrikine AV, Vrouwenvelder T (2000) Surface ground vibration due to a moving train in a tunnel: two-dimensional model. J Sound Vib 234(1):43–66. doi:10.1006/jsvi.1999.2853
16. Andersen L, Jones CJC (2006) Coupled boundary and finite element analysis of vibration from railway tunnes—a comparison of two and three-dimensional models. J Sound Vib 293(3–5):611–625. doi:10.1016/j.jsv.2005.08.044
17. Degrande G, Clouteau D, Othman R, Arnst M, Chebli H, Klein R, Chatterjee P, Janssens B (2006) A numerical model for ground-borne vibrations from underground railway traffic based on a periodic finite element-boundary element formulation. J Sound Vib 293(3–5):645–666. doi:10.1016/j.jsv.2005.12.023
18. Verbraken H, Lombaert G, Degrande G (2011) Verification of an empirical prediction method for railway induced vibrations by means of numerical simulations. J Sound Vib 330:1692–1703. doi:10.1016/j.jsv.2010.10.026
19. Connolly DP, Kouroussis G, Giannopoulos A, Verlinden O, Woodward PK, Forde MC (2014) Assessment of railway vibrations using an efficient scoping model. Soil Dyn Earthq Eng. 58:37–47. doi:10.1016/j.soildyn.2013.12.003
20. Connolly DP, Kouroussis G, Woodward PK, Giannopoulos A, Verlinden O, Forde MC (2014) Scoping prediction of re-radiated ground-borne noise and vibration near high speed rail lines with variable soils. Soil Dyn Earthq Eng 66:78–88. doi:10.1016/j.soildyn.2014.06.021
21. Paneiro G, Durão FO, Costa e Silva M, Falcão Neves P (2015) Prediction of ground vibration amplitudes due to urban railway traffic using quantitative and qualitative field data. Transp Res D Transp Environ 40(1):1–13. doi:10.1016/j.trd.2015.07.006
22. Dinis da Gama C (2013) Unifying energetic approach for the control of ground vibration. In: Kwasniewsky M, Lydzba D (eds) Rock mechanics for resources, energy and environment. CRC Press, Boca Raton, pp 727–732
23. Gardien W, Stuit HG (2003) Modelling of soil vibrations from railway tunnels. J Sound Vib 267(3):605–619. doi:10.1016/S0022-460X(03)00727-2
24. Jaeger JC, Cook N, Zimmerman RW (2007) Fundamentals of rock mechanics. Blackwell, London
25. Ju SH (2007) Finite element analysis of structure-borne vibration from high-speed train. Soil Dyn Earthq Eng. 27(3):259–273. doi:10.1016/j.soildyn.2006.06.006
26. Kutner MH, Nachtsheim CJ, Neter J, Li W (2005) Applied linear statistical models. McGraw-Hill Irwin, New York
27. Draper NR, Smith H (1998) Applied regression analysis. Wiley, New York
28. Strang G (1988) Linear algebra and its applications. Thomson, Belmont
29. Beale MH, Hagen MT, Demuth HB (2015) Neural network toolbox™ user's guide. The MathWorks Inc, Natick
30. MacKay DJC (1991) Bayesian methods for adaptive models. Ph.D. thesis, California Institute of Technology
31. Luke S (2013) Essentials of metaheuristics. Lulu. http://cs.gmu.edu/∼sean/book/metaheuristics/. ISBN 978-0-557-14859-2
32. Thomas FE, Himmelblau DM, Lasdon LS (2001) Optimization of chemical processes. McGraw-Hill, New York
33. Locatelli M, Schoen F (1999) Random Linkage: a family of acceptance/rejection algorithms for global optimization. Math Prog 85:379–396. doi:10.1007/s101070050062
34. Rinnooy AHGK, Timmer GT (1989) Global optimization. In: Nemhauser GL et al (eds) Handbooks in OR and MS, vol 1. Elsevier, Amsterdam
35. Rinnooy AHGK, Timmer GT (1987) Stochastic global optimization methods, part 2: multi level methods. Math Prog 39:57–78. doi:10.1007/BF02592071
36. Abraham B, Ledolter J (2006) Introduction to regression modeling. Thomson, Belmont