# Prediction of mechanical stress in roller leveler based on vibration measurements and steel strip properties

**Riku-Pekka Nikula**

Control Engineering, Faculty of Technology, University of Oulu

P.O. Box 4300, FI-90014 University of Oulu, Finland

e-mail: riku-pekka.nikula@oulu.fi, phone number: +358 294 482475


**Konsta Karioja**

Mechatronics and Machine Diagnostics, Faculty of Technology, University of Oulu

P.O. Box 4200, FI-90014 University of Oulu, Finland


**Kauko Leiviskä**

Control Engineering, Faculty of Technology, University of Oulu

P.O. Box 4300, FI-90014 University of Oulu, Finland


**Esko Juuso**

Control Engineering, Faculty of Technology, University of Oulu

P.O. Box 4300, FI-90014 University of Oulu, Finland

## ABSTRACT

The continuous development of steel products generates new challenges for the maintenance of manufacturing machines in steel mills. Substantial mechanical stress is inflicted on the machines during the processing of modern high-strength steels. This increases the risks of damage and flaws in the processed material may appear if the capability of a machine is exceeded. Therefore, new approaches are needed to prevent the machine condition from deteriorating. This study introduces an approach to the prediction of mechanical stress inflicted on a roller leveler during the processing of cold steel strips. The relative stress level is indicated by features extracted from an acceleration signal. These features are based on the calculation of generalized norms. Steel strip properties are used as explanatory variables in regression models to predict values for the extracted vibration features. The models tested in this study include multiple linear regression, partial least squares regression and generalized regression neural network. The models were tested using an extensive data set from a roller leveler that is in continuous operation in a steel mill. The prediction accuracy of the best models identified indicates that the relative stress level inflicted by each steel strip could be predicted based on its properties.

**Introduction**


Product diversity in steel mills is continuously growing. The development of modern high-strength steels and advanced high-strength steels is driven by demanding market requirements (Silvestre et al. 2015). The production of these modern steels pushes the manufacturing equipment close to its limits of endurance. The manufacturing machines, such as roller levelers, are often designed for steel products with different properties compared with modern steels. Therefore, the machine has to endure harsher conditions than it was originally designed for. The testing and modeling of material behavior and the formability of these new steel materials has received attention recently (Bruschi et al. 2014; Dong et al. 2016; Silvestre et al. 2015; Sriram et al. 2012). The perspective of this study, however, is to evaluate the mechanical stress effect on the roller leveler, inflicted when processing various materials, including high-strength steel. The objective is to improve the safe operation and damage prevention in roller levelers in the light of increasingly challenging production requirements.

Roller leveling is a method for straightening steel plates or strips after final rolling, heat treatment, or cooling operations. Flatness imperfections and uneven stresses can be eliminated by bending the rolled material in alternating directions. The literature on roller leveling is largely dominated by modeling based on finite element models (Huh et al. 2003; Seo et al. 2016; Silvestre et al. 2014) and analytical models (Baumgart et al. 2015; Chen et al. 2015; Cui et al. 2011; Doege et al. 2002; Liu et al. 2012; Silvestre et al. 2014). A vast number of these models are concerned with process simulation and parameter analysis. On the other hand, several studies focus on the analysis and control of material behavior in the leveling process (Dratz et al. 2009; Madej et al. 2011; Morris et al. 2001; Park and Hwang 2002). These approaches have a major significance for the design and the improvement of the leveling process. However, they do not provide direct solutions to the prevention of machine condition deterioration.

The effects of leveling on machine condition have also been investigated by some authors. Sueoka et al. (2002) and Matsuzaki et al. (2008) applied analytical models and actual vibration measurements in the study of polygonal wear on work rolls in a hot leveler. Additionally, Karioja et al. (2015) analyzed vibration measurements to study the stress inflicted on an industrial roller leveler. The effects of leveling parameters on vibration features were analyzed by Nikula and Karioja (2016). Moreover, there is a broader selection of studies related to the vibrations in other steel forming processes and especially rolling processes. Numerical vibration modeling approaches have been used to study non-linear vibrations (Bar and Świątoniowski 2004), mid-frequency vibrations (Bar and Bar 2005), vertical vibrations (Nizioł and Świątoniowski 2005), and the chatter phenomenon (Heidari and Forouzan 2013). Wu et al. (2014) examined the relationships of vibration characteristics with local defects on the roll surface and later the chatter phenomenon (Wu et al. 2015) using numerical approaches and actual vibration measurements. The effect of vibrations on flatness measurements of steel strips has been studied by Usamentiaga et al. (2014) and Usamentiaga et al. (2015). However, the analysis of vibration measurements from roller levelers in industrial cases has rarely been discussed in the literature. This study extends the work presented by Karioja et al. (2015) and Nikula and Karioja (2016) by introducing a modeling approach to the stress evaluation of the industrial leveler.

In this study, a data-driven approach to modeling is used. This is an especially suitable approach for industrial systems that are in continuous use. The model generation is relatively straightforward and detailed information on the physical properties of the modeled system can be partly ignored. On the other hand, comprehensive historical data from typical operation is needed for model training. Data-driven modeling and its application on machine diagnostics and prognostics have been widely studied (Jardine et al. 2006; Lee et al. 2014). The modeling methodologies include linear regression models (Wise and Gallagher 1996), such as multiple linear regression (MLR), principal component regression (PCR), partial least squares regression (PLSR) and non-linear regression methods, such as artificial neural networks (ANN) (Specht 1991) and support vector machines (SVM) (Smola and Schölkopf 2004). Additionally, Bayesian approaches (Mosallam et al. 2016) and stochastic modeling approaches such as Markov, semi-Markov (He et al. 2012), hidden Markov models (Wang 2007) have gained broad interest in machine prognostics recently. Data-driven modeling has been applied in rolling mill applications quite extensively as well. Neural networks have been used in many cases, including plate width set-up value estimation in a hot plate mill (Lee et al. 2000), temperature prediction for steel slabs (Laurinen and Röning 2005), steel hardness prediction (Das and Datta 2007) and the prediction of work roll thermal expansion (Alaei et al. 2016). Furthermore, Faris et al. (2013) used genetic programming to predict rolling force, torque, and slab temperature. Serdio et al. (2014) proposed residual-based fault detection using soft computing techniques for condition monitoring in rolling mills. However, this investigation differs from those examples in the target of the modeling, which is the prediction of the mechanical stress inflicted on the roller leveler.

Acceleration measurements have been used for mechanical stress evaluation recently in some studies. Acceleration is a response to the force applied to the machine, and therefore, it has potential for stress monitoring. In some cases, the acceleration signal provides an even better indication of changes in the stress level compared with the strain gauge signal (Karioja and Lahdelma 2013). Accelerometers are practical sensors for industrial applications and for various fault detection applications as well (Lahdelma and Juuso 2011a). Stress evaluations based on acceleration signals have been previously made for steel mill machines such as the steel cutter (Karioja and Lahdelma 2015) and roller leveler (Karioja et al. 2015). Cumulative stress indices obtained from vibration measurements have also been proposed for a Kaplan water turbine and a load haul dumper (Juuso 2014). Cumulative stress indices were previously introduced for the prediction of roller mill fatigue based on torque measurements (Juuso and Ruusunen 2013) and later extended to the real-time risk analysis of machines and process devices (Juuso and Galar 2016). The stress contributions were obtained using a data-driven non-linear scaling approach (Juuso and Lahdelma 2010). The approach proposed in this study has the following differences with regard to these previous stress evaluation approaches. A data-driven model is proposed for the prediction of the relative stress inflicted by each steel strip. Additionally, the trained model can be used for stress evaluation without real-time measurements, in contrast to the aforementioned approaches. The stress contributions are based on linear evaluation of vibration feature values instead of the non-linear scaling approach.

The research question discussed in this study concerns the identification of the relative stress that is inflicted during the processing of steel strips. The relative stress defines the relative level of the mechanical stress that each leveling event inflicts on the machine. To indicate the relative stress level, features are extracted from an acceleration signal measured from the machine structure. These features are based on the generalized norms, which have been used in stress monitoring (Karioja and Lahdelma 2015) and various industrial condition monitoring applications previously (Lahdelma and Juuso 2011a). The requirement for feature extraction comes from the large amount of data produced by an accelerometer. With generalized norms, this data can be effectively compressed, and at the same time, both long-term stress and impact stress effects can be monitored. Additionally, the automatic computation of large number of such features is practical in condition monitoring approaches.

The working hypothesis is that the correlations of vibration features with steel strip properties could be successfully exploited as the basis for model generation. Additionally, the prediction of the relative stress level is studied using the generated regression models. These models include multiple linear regression, partial least square regression, and generalized regression neural network (GRNN). MLR is used to identify the linear relations between the vibration features and steel strip properties, whereas PLSR is used to reduce the dimensionality and collinearity in the explanatory data. GRNN is used to build a model that is free from the linearity assumption. The applied modeling approach is validated using an extensive data set, which includes data from a wide range of different steel strips. Based on the literature survey, multiple models for roller leveling have been introduced, but these models are mainly used for process simulation or the analysis of material behavior and process parameters. These models are typically complex and their application requires excessive computation. In contrast, this study introduces a straightforward experimental approach that can be applied in an industrial environment to support maintenance planning, for instance.

This paper is organized as follows. The "Materials and methods" section provides a description of the industrial case and the methods used to conduct the study. The results from the industrial stress evaluation case are shown and discussed in the "Results and discussion" section. Finally, the study is summarized in the "Conclusions" section.

## Materials and methods

The principles of roller leveling and the leveler studied here are presented in "Roller leveling" section. The practicalities related to vibration measurements are described in the "Vibration measurements" section. The generalized norms, which are used as the basis for vibration feature generation, are introduced in the "Generalized norms" section. The "Generalized norms in vibration simulation" section demonstrates the effect of the norm order in change detection using simulated vibration signals that imitate signals obtained from the industrial leveler. "Feature generation" introduces the features generated from the signals and steel strip properties. The modeling approach is presented in the "Regression modeling" section.

Roller leveling

The goal of roller leveling is to eliminate shape defects in the material. Steel coils contain flatness defects caused by uneven stresses and defects resulting from thickness variation across the product width (Smith Jr 1997). The stress patterns create longitudinal and transverse curvature. Edge and center waves are caused by a difference in the length of the sheet between the center and the edges (Park and Hwang 2002). Roller leveling is done by subjecting the strip to multiple back and forth bending sequences with increasing roll gaps, as illustrated in Fig. 1. In other words, the strip is exposed to reverse bending. The rolls on the entry side cause more curvature to the strip than the rolls near the exit. Strains in the strip are controlled by the set geometry of the leveler. The principle of roller leveling is based on controlling the plastic deformation through the thickness of the material. Plastic deformation determines the resultant flatness and memory and it also affects the required force. The roll force is a function of material thickness, width, yield strength, roll spacing, and the extent of plastic deformation (Smith Jr 1997). Appropriate control of operational parameters is therefore required for the desirable leveling result.

The roller leveler under investigation is used for strips of cold steel at the SSAB steel mill in Raahe, Finland. Sheets are cut from a strip on the production line after the leveler using a flying shear. The cutting is performed simultaneously with leveling without the need to stop the strip in the leveler for cutting. The cutting of sheets causes shocks that are conducted to the leveler and emerge as peaks in the monitored vibration signal as described in the "Steel cut effect removal" section later on. The processed steel strips considered in this study showed a large variation of properties. The range of the yield strength was 210–1640 MPa; the length range was 68–1161 m; the thickness range was 1.98–15.21 mm; the weight range was 7400–29280 kg, the width range was 861–1875 mm; the number of cut sheets was 4–465 and 55 different steel grades were processed altogether. Materials with high yield strength and thickness impart substantial force on the rolls, whereas long strips need to be processed for a long duration and potentially inflict a high accumulation of stress.
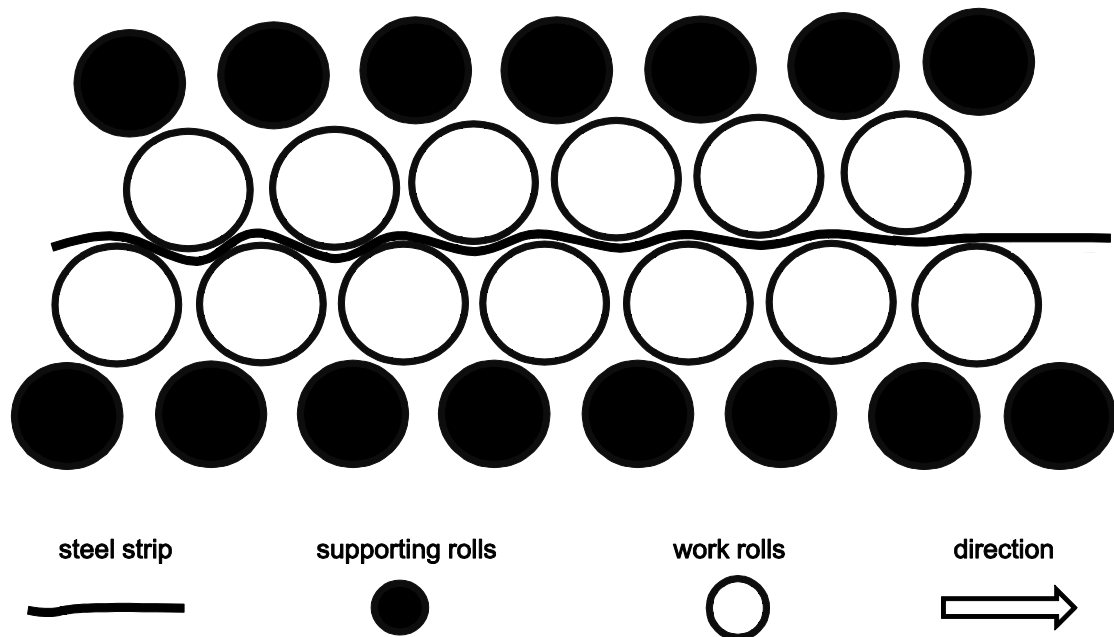


**Fig. 1** The principle of a roller leveler

Vibration measurements

Three accelerometers were stud-mounted on the supporting structure beneath the lower supporting rolls of the leveler. The acceleration was measured horizontally in the cross direction compared with the direction of the roller track. Only the signal from the sensor located in the middle of the roller track was used in the stress estimation, because one signal was considered sufficient for this study. The other signals were used in data preparation, which is explained in the "Data

preparation" section. The accelerometer used was an SKF CMSS 787A-M8, which has a frequency response from 0.7 Hz to 10 kHz with ±3 dB deviation. The measurement hardware included an NI 9234 data acquisition card and an NI CompactRIO for data acquisition. The sampling rate was 25.6 kHz and the only filter used at the hardware level was the built-in antialiasing filter of the data acquisition card. The measurement system was calibrated using a hand-held calibrator.

Generalized norms

The generalized norm introduced by Lahdelma and Juuso (2008) is defined by

$$\left\| x^{(\alpha)} \right\|_p = \left( \frac{1}{N} \sum_{i=1}^{N} \left| x_i^{(\alpha)} \right|^p \right)^{\frac{1}{p}}. \tag{1}$$

This feature is known as the $l_p$ norm of signal $x^{(\alpha)}$ where $p$ is the order of the norm, $\alpha$ is the order of derivation, $x$ stands for displacement, and $N$ is the number of data points. The $l_p$ norm has the same form as the generalized mean, also known as the Hölder mean or power mean (Bullen 2003). The $l_p$ norms are defined in such a way that $1 \le p < \infty$. In the case of $0 < p < 1$, norm (1) is not a proper norm in general, because it violates the triangle inequality $\| x + y \| \le \| x \| + \| y \|$ (Lahdelma and Juuso 2011b). However, in this case, these $p$ values are also valid because $y$ is the null vector. The root mean square (rms) and the peak value, which are special cases of norm (1) when $p = 2$ and $p = \infty$, respectively, are often used as features in condition monitoring (Jantunen and Vaajoensuu 2010; Li et al. 2012). In this study, norms $l_{0.1}$, $l_{0.5}$, $l_1$, $l_2$, $l_4$, and $l_{10}$ were calculated from an acceleration signal ($\alpha = 2$), but other signals could also be used. The features generated from generalized norms are introduced in the "Feature generation" section.

Generalized norms in vibration simulation

The acceleration signal measured from the roller leveler contains varying amplitude levels and peaks with different magnitudes. These features were simulated in order to demonstrate the significance of the norm order in feature extraction. The simulated signals consist of three cosine components with frequencies of 50, 80, and 150 Hz and 0° phases, respectively. The signals were generated by combining 60 samples that had 25600 points, which corresponds to one minute of data using sampling frequency of 25.6 kHz.

In signal 1, the amplitudes ($X$) of each frequency were $X = 0.5$ on samples 1–15, $X = 1$ on samples 16–30, $X = 4$ on samples 31–45, and $X = 10$ on samples 46–60. Gaussian noise with variance $\sigma^2 = 0.5$ was added on each sample. The signal-to-noise ratios (SNR) of these four 15-second segments were SNR = $20 \cdot \log_{10}(l_2^{\text{signal}} / l_2^{\text{noise}})$ = [-1.25 4.77 16.81 24.77], respectively. The complete signal is shown in Fig. 2 on the left.

Signal 2 was generated by combining the second sample segment (samples 16–30) from signal 1 four times in a row. The last three segments were manipulated by adding three events with exceptionally large values on each segment. The magnitudes of these values were ±20, ±40, ±60, also shown in Fig. 2 on the right. One negative and one positive value were added on each event. Signal 3 is the same as signal 2 but without the exceptionally large values.

A single norm value was computed from each one-second sample ($N = 25600$). The norm orders were $p = [0.1\ 0.5\ 1\ 2\ 4\ 10]$. The sixty values of $l_{0.1}$, $l_2$, and $l_{10}$ from signals 1 and 2 are shown in Fig. 2. The increasing amplitude of signal 1 is clearly seen in the norm values, as indicated on the left in Fig. 2. The relative magnitude of change between the segments is shown in Fig. 3. The leftmost graph in Fig. 3 shows the ratios of norm averages from the 2nd, 3rd, and 4th segments to the 1st segment from signal 1. The low order norm (e.g. $p = 0.1$) results in larger relative change compared with the high order norm (e.g. $p = 10$). This can clearly be seen especially in the ratio of the 4th segment to the 1st segment. This behavior shows that the difference between signal amplitude levels is the most distinguishable when norms with low order $p$ are used.

The graphs on the right in Fig. 2 show that exceptionally high signal values have a major influence on the high order norms (e.g. $p = 10$). On the other hand, the effect is small on the low order norms ($p \le 2$). The same effect is illustrated by the ratios of maximum norm values in the middle graph in Fig. 3. This behavior indicates that the effect of a single peak is large on the high order norms and negligible on the low order norms.

The rightmost graph of Fig. 3 illustrates the effect of exceptionally high signal values on the sum values of norms, which were also used in feature generation, as presented in the next section. The sums of complete signals 2 and 3 were compared

by studying their ratios. The influence of exceptionally high values was significant in the sums computed using a norm that had the order $p \geq 4$. When the order of the norm was small, the effects of exceptionally high signal values were negligible and the ratios in Fig. 3 are therefore close to one. This behavior illustrates that large peak values may affect the norm sums if the order of the norm is high and the number of exceptionally high values is large with relation to the number of values summed.
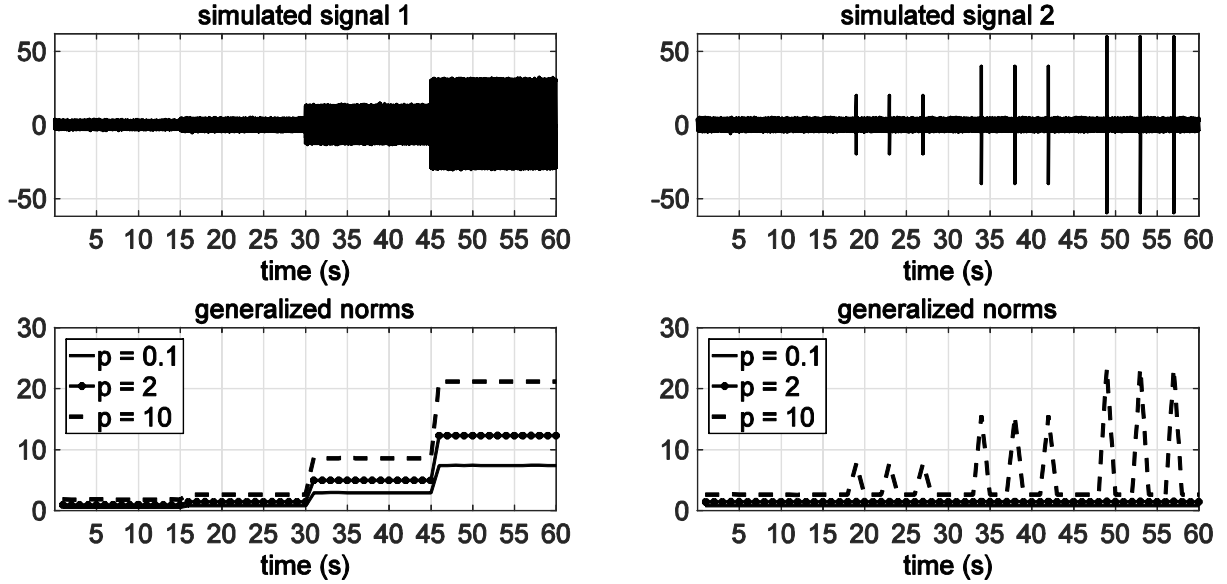


**Fig. 2** Simulated signals (above) and generalized norms $l_{0.1}$, $l_2$, and $l_{10}$ from the corresponding signals (below)
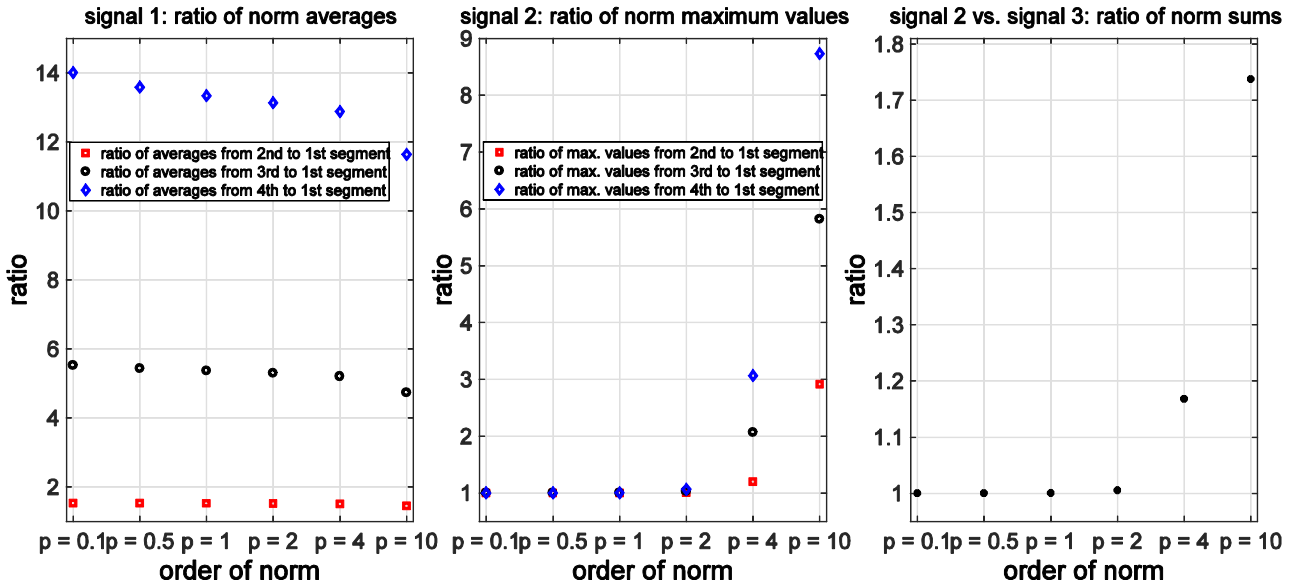


**Fig. 3** Effect of norm order in the detection of signal changes. $1^{st}$, $2^{nd}$, $3^{rd}$, and $4^{th}$ segments correspond to points 1–15, 16–30, 31–45, and 46–60 in the signals, respectively

Feature generation

The features generated from steel strip properties are shown in Table 1. The yield strength, length, weight, width, and thickness of steel strips are features number 1–5, respectively. These features were further transformed using the common logarithm, square, cube, square root, and cube root to produce features 6–30. The features with similar transformation

were multiplied by each other to generate features 31–90. Finally, features 1–5 were multiplied by features 6–30 to generate features 91–190. These features were used as the explanatory variables in the modeling approach presented in the following section.

**Table 1** Steel strip features

| Feature number | Description |
| --- | --- |
| 1, 2, 3, 4, 5 | Yield strength, length, weight, width and thickness, respectively |
| 6–10 | Common logarithm of features 1–5 |
| 11–15 | Square of features 1–5 |
| 16–20 | Cube of features 1–5 |
| 21–25 | Square root of features 1–5 |
| 26–30 | Cube root of features 1–5 |
| 31–40 | Products of features 1–5. Only different features were multiplied by each other. This also concerns features 41–90. |
| 41–50 | Products of features 6–10 |
| 51–60 | Products of features 11–15 |
| 61–70 | Products of features 16–20 |
| 71–80 | Products of features 21–25 |
| 81–90 | Products of features 26–30 |
| 91–110 | Products of features 1–5 and 6–10 (Feature multiplication by its transformation, such as '$\log_{10}$(length)·length' was not included. This also concerns features 111–190) |
| 111–130 | Products of features 1–5 and 11–15 |
| 131–150 | Products of features 1–5 and 16–20 |
| 151–170 | Products of features 1–5 and 21–25 |
| 171–190 | Products of features 1–5 and 26–30 |

In order to obtain the feature values for the relative stress level, the generalized norm sums were computed from the acceleration signal. The sums were computed by adding up the norm values of one-second samples ($N = 25600$) from each leveling event. The leveling of one complete steel strip was considered as a leveling event. The summation was done to include the effect of stress accumulation during the leveling event on the relative stress features. These features were then used as response variables in the models. The response variable set contained $\sum l_{0.1}$, $\sum l_{0.5}$, $\sum l_1$, $\sum l_2$, $\sum l_4$, and $\sum l_{10}$ and the square, square root, and common logarithm of each. Twenty-four variables were included in this part of the set. The same variables were also computed after the removal of data points that correspond to steel cutting events. The removal of steel cut effects is introduced in the "Steel cut effect removal" section. Altogether 48 response variables were included in the complete set. In order to produce comparable modeling results, the values of explanatory variables and response variables were scaled to range 0–1.

Regression modeling

This section presents the applied modeling methods, including the multiple linear regression, partial least squares regression, and generalized regression neural network. Thereafter, the criteria used for model assessment are shown. The last part of this section presents the applied variable selection procedures and the cross-validation approach.

*Multiple linear regression*

Multiple linear regression is a popular and simple regression method, where the response variable is considered a linear combination of certain explanatory variables. MLR models the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. If the model has only one explanatory variable, the model is a simple linear regression model. An MLR model with $N$ observations and $k$ explanatory variables is formally defined by

$$y_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \cdots + \beta_k x_{jk} + \varepsilon_j, \tag{2}$$

where $j = 1, 2,…N$, $y$ denotes the value of the response variable, $x$ is the value of the explanatory variable, $\beta_0$ is the intercept, $\beta_1$–$\beta_k$ are the unknown regression coefficients to be estimated, and $\varepsilon$ is the error term. The model is identified using the least squares fitting.

*Partial least squares regression*

Partial least squares is an extension of principal component analysis and it has the ability to analyze data with several, noisy, collinear and incomplete variables (Wold et al. 2001). The underlying assumption of PLSR is that the observed data is generated by a system or process which is driven by a small number of latent variables, which are not directly observed or measured. The latent variables are linear combinations of the original variables and hold no correlation with each other. The latent variables explain the variation in the explanatory variables $X$ and the variation in $X$ which is the most predictive of the response variables $Y$. That is to say, PLSR maximizes the covariance between matrices $X$ and $Y$. The matrix $X$ is decomposed into a score matrix $T$, loading matrix $P$, and residual $E$. Similarly, the matrix $Y$ is decomposed into a score matrix $U$, loading matrix $Q$, and residual $F$. The matrix decompositions are defined by

$$X = TP' + E ,\tag{3}$$

$$Y = UQ' + F.\tag{4}$$

The latent variables are calculated iteratively extracting informative features one at a time. The number of latent variables is typically smaller than the number of original variables and thus the method is considered as a dimensionality reduction method (May et al. 2011). However, it is also possible to express the PLSR formula in terms of the original variables (Rosipal and Krämer, 2006). There are several methods to determine the number of latent variables in a model. In this study, the number of latent variables was determined based on the cross-validation test result. The function *plsregress* in Matlab was used for model training.

*Generalized regression neural network*

The generalized regression neural network, developed by Specht (1991), is a memory-based network, which includes a one-pass learning algorithm with parallel structure. It approximates any arbitrary function between input and output vectors and draws the function estimate directly from the training data. The method is suitable for regression problems where an assumption of linearity is not justified. A GRNN configuration consists of four layers, which include the input layer, pattern layer, summation layer, and output layer (Kim et al. 2010). Each input unit in the input layer corresponds to individual observed parameters. The input layer is fully connected to the pattern layer, where each neuron represents a training pattern and its output is a measure of the distance of the input from the stored patterns. The pattern layer is connected to the summation layer, which has two different types of summation including $S$-summation neuron and $D$-summation neuron. S-summation neuron determines the sum of the weighted outputs of the pattern layer, whereas the $D$-summation neuron determines the unweighted outputs of the pattern neurons. The connection weight between the $i$th neuron in the pattern layer and the $S$-summation neuron is $y_i$, which is also the target output value corresponding to the $i$th input pattern. The connection weight for $D$-summation neuron is unity. The output layer divides the output of each $S$-summation neuron by that of each $D$-summation neuron. Therefore, a predicted value $\hat{y}(x)$ to an unknown input vector $x$ can be expressed as (Kim et al. 2010)

$$\hat{y}_i(x) = \frac{\sum_{i=1}^{n} y_i \exp[-D(x,x_i)]}{\sum_{i=1}^{n} \exp[-D(x,x_i)]},\tag{5}$$

where $n$ and $x_i$ represent the number of training patterns and the $i$th training input pattern stored between the input and pattern layers, respectively. The Gaussian $D$ function is defined as

$$D(x,x_i) = \sum_{j=1}^{p} \left(\frac{x_j - x_{ij}}{\sigma}\right)^2,\tag{6}$$

where $p$ indicates the number of elements of an input vector. The $x_j$ and $x_{ij}$ represent the $j$th element of $x$ and $x_i$, respectively. The parameter $\sigma$ is referred to as the spread parameter, whose optimal value is often experimentally evaluated. In this study, the spread parameter of the best model was defined based on the cross-validation test result. The tested values were $\sigma = [0.01\ 0.05\ 0.1\ 0.2\ 0.5\ 0.7\ 1\ 1.5]$. The Matlab function *newgrnn* was used for model training.

*Criteria for model performance evaluation*

Four criteria were used to evaluate the models in this study. The predictive performance of the models was evaluated using the root mean squared error of prediction (RMSE), which gives the average prediction error. The RMSE criterion is given by

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2}, \tag{7}$$

where $y_j$, $\hat{y}_j$, and $N$ are the observed value, the corresponding predicted value, and the total number of observations, respectively. The goodness of fit for linear models was evaluated with the coefficient of determination ($R^2$). The general definition of $R^2$ is

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}, \tag{8}$$

where $SS_{res} = \sum(y_j - \hat{y}_j)^2$ is the sum of the squares of residuals and $SS_{tot} = \sum(y_j - \bar{y})^2$ is the total sum of the squares. When the criterion is close to one, the fit of the model is good; when the criterion is close to zero, the fit is poor. The $R^2$ criterion is inappropriate for the evaluation of non-linear regression models (Spiess and Neumeyer, 2010), and therefore it was used only for linear models. Pearson's correlation coefficient was used to evaluate the linear correlation between the model predictions and the observed values in linear and non-linear models. The correlation coefficient for two variables $x$ and $y$ is given by

$$R_{xy} = \frac{\sum_{j=1}^{N}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{N}(x_j - \bar{x})^2 \sum_{j=1}^{N}(y_j - \bar{y})^2}}. \tag{9}$$

Variance inflation factor (VIF) was used to assess the multicollinearity in the models. Multicollinearity is an indication of collinearity between three or more variables even if no pair of variables has a high linear correlation. This situation can be a serious problem for MLR and neural network models with many explanatory variables (May et al. 2011; James et al. 2013). The VIF for each explanatory variable can be computed using the formula

$$VIF = \frac{1}{1 - R_i^2}, \tag{10}$$

where $R_i^2$ is the $R^2$ from a regression of explanatory variable $X_i$ onto all of the other explanatory variables. If $R_i^2$ is close to one, then collinearity is present and the VIF will be large. The smallest value of VIF is one, which indicates the complete absence of collinearity. As a rule of thumb, a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity (James et al. 2013).

*Variable selection and model validation*

Performing the variable selection using an exhaustive subset selection approach requires the evaluation of a very large number of subsets. Suboptimal search procedures can significantly reduce the number to be evaluated (Whitney, 1971). Therefore, forward selection was applied to variable selection in this study. The exhaustive search, which tests all variable combinations in the models, was tested with the MLR models as an alternative variable selection approach. The approach was tested to investigate if the optimal subsets are clearly better than the suboptimal subsets defined using the forward selection. However, this approach was not tested with the PLSR and GRNN models due to their computationally more burdensome training procedure.

In forward selection, the variables are included in progressively larger subsets so that the prediction performance of the model is maximized. First, $P$ models each consisting of only one explanatory variable are built. $P$ is the number of candidate variables. The variable that gives the best value for the observed performance criterion is selected. Then, $P$-1 models are built each including the already selected variable and each of the remaining variables one at a time. The performance of the models is evaluated and the variable leading to the best model performance is added to the model. The addition of one variable at a time is repeated until the desired number of variables have been selected.

Models with one to four explanatory variables were generated using the exhaustive search approach. The whole set of steel property features, presented in Table 1, was used in the exhaustive search for models with one and two explanatory variables. Features 1–90 were used in models with three explanatory variables and features 1–30 in models with four explanatory variables. These reduced sets were selected due to the high computational requirement of exhaustive search.

The variables for the linear models were selected based on the model performance on the test sets using the cross-validated average of $R^2$. For GRNN, the variables were selected based on the cross-validated average of $R_{xy}$ on the test sets. The selection was done based on the testing results in order to obtain models with the best prediction ability. The collinearity of explanatory variables in the best models was estimated using the maximum value of VIF from the particular models. A VIF value higher than five was considered as an indication of collinearity.

Repeated random sub-sampling validation, also known as Monte Carlo cross-validation (Picard and Cook 1984; Shao 1993), was used to validate the model performance. The data set was split one hundred times into training and test sets. Each test set included 20% of the points selected randomly and the remaining 80% of the points were included in the training set. All the models with different variable configurations were tested using the same random sets in order to enable an equal comparison of the models.

**Results and discussion**

Measurements in an industrial environment are susceptible to complications that need to be considered in data analysis. Signal pre-processing is an important prerequisite for credible results. The applied data preparation is therefore presented in the "Data preparation" section in a detailed manner and the removal of the steel cut effect from the signal is introduced in the "Steel cut effect removal" section. The main results are presented in the "Modeling results" section. The correlations between steel strip properties and vibration features are analyzed thereafter. The concluding discussion on roller leveler stress estimation based on the observations is presented in the "Discussion" section.

Data preparation

Acceleration was measured during a period of 37 days. The measurement was continuous and consequently the data included the leveling events as well as other irrelevant events. One minute of data from three sensors was saved in a single file. The elimination of unnecessary data was initiated by deleting the individual files that had an $l_2$ value smaller than $0.045$ m/s$^2$ in all three signals. The norm was computed from the data points of each complete signal separately, producing three values for each file. Crossing this limit value was considered an indication of action on the leveler.

The data acquisition hardware was automatically restarted once a day to prevent the measurement system from crashing due to an unknown problem. An interruption in sensor power supply caused a large deviation in signal values. High values were then obtained for a few seconds as a consequence of the system settling time. Therefore, files that contained a signal value higher than 200 m/s$^2$ were completely removed.

The acceleration signals were then connected to specific steel strips based on the time stamps that indicated the start of the processing and the end of the coil tail drive. Expertise and inference were needed to address the inaccuracy in the time stamps. Leveling durations that were shorter than five minutes based on the number of files were considered incomplete and the corresponding events were thus removed. If the previous leveled strip had more than a 15-minute overlap with the strip under review based on the time stamps, the previous strip was also rejected in this case.

During data analysis, it was noticed that a periodic disturbance was present in some of the signals when the leveler was idle. This disturbance increased the $l_2$ value computed from the data points of a complete signal in a file. The files with the disturbance were removed by checking the ratio of $l_{10}$ to $l_2$. However, this was done only for files with an $l_2$ maximum over $0.2$ m/s$^2$, computed in one-second segments ($N = 25600$). Files with a lower maximum were automatically accepted because the ratio check was not appropriate in these cases. In the ratio check, files with $l_{10}/l_2 \geq 3.75$ computed from the data points of the complete signal were accepted if 10% of the ratios computed in three-second segments ($N = 76800$) from the complete signal were also higher than 2.1. The objective of the three-second examination was to eliminate the influence of single ratio peaks that were not caused by the leveler operation.

After these stages, the leveling events that then contained less than six files or more than 60 files were rejected. These events were considered incomplete events or events that still included several minutes of idle machine state. Finally, 752 steel strip leveling events were accepted for data analysis. The signal average was subtracted from the signals in order to ensure they had a zero mean.

Steel cut effect removal

The flying shear next to the steel leveler causes notable shocks, which can also be seen as high peaks in some of the measurements. Fig. 4 shows an example of the acceleration signal from the leveling of a single steel strip. The duration of this signal was 11 minutes. The steel strip was relatively thick (15.2 mm) and clearly distinguishable peaks emerged in the measured signal. The influence of cutting can also be seen in the $l_{10}$ values, as shown in Fig. 4. 51 sheets were cut from this particular strip. However, it is difficult to define which shocks are definitely caused by cutting based on the acceleration signal alone.

To remove the effect of the flying shear, a steel cut effect removal approach was applied to half of the 48 response variables. The $l_{10}$ value was assumed to represent the steel cut effect if it was larger than the average of $l_{10}$ computed in one-second segments from the whole leveling event and more than 1.5 times larger than the mean of four previous $l_{10}$ values. The corresponding data points were removed from all the generalized norms ($l_{0.1}$, $l_{0.5}$, $l_1$, $l_2$, $l_4$, $l_{10}$) based on the steel cut effects found in $l_{10}$. The check was performed starting from the end of the event moving towards the start of the event. An example of removing the steel cut effect for one signal is shown in Fig. 4.
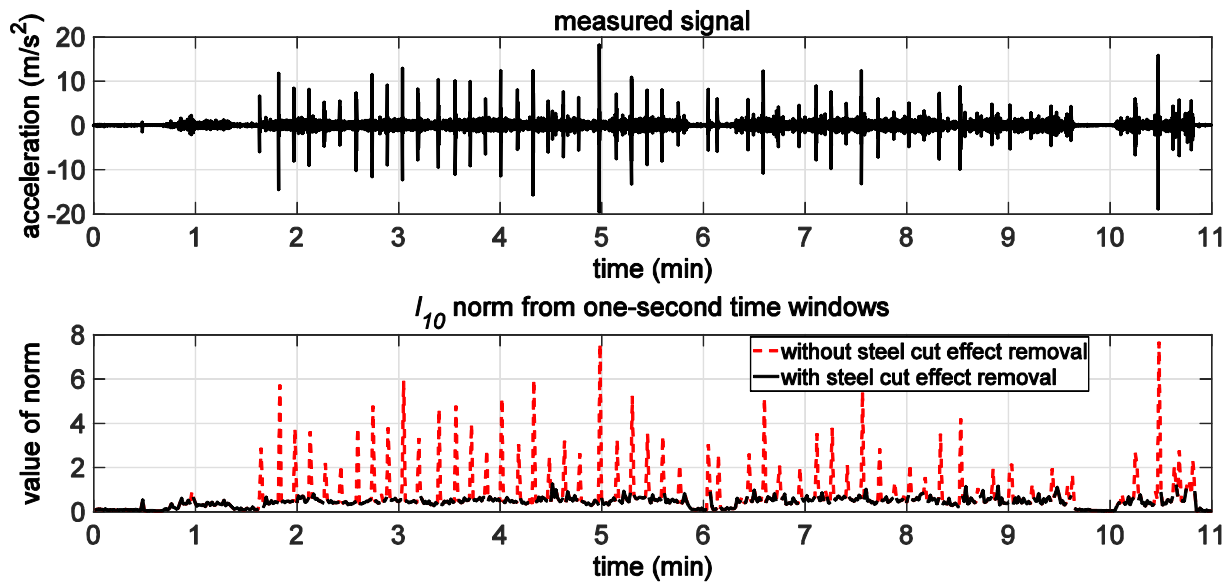


**Fig. 4** Acceleration signal and $l_{10}$ norm during leveling of a steel strip

Modeling results

Regression models were generated to predict the relative stress on the machine. Forty-eight vibration features, which were introduced in the "Feature generation" section, were used as response variables in the models to indicate the stress level. All of the steel strip features shown in Table 1 were used as explanatory variable candidates in forward selection. MLR and PLSR were tested using one to ten explanatory variables, whereas GRNN was tested using one to five explanatory variables. MLR was also tested using one to four explanatory variables selected using the exhaustive search approach. The features tested as explanatory variables are given in Table 2. The rightmost column shows the number of variable combinations in each case. It can be seen that the number of possible combinations becomes high when the number of explanatory variables in a model increases.

**Table 2** Data sets in exhaustive search

| Number of explanatory variables | Steel strip features | Number of variable combinations |
|---|---|---|
| 1 | 1–190 | 9120 |
| 2 | 1–190 | 861840 |
| 3 | 1–90 | 5639040 |
| 4 | 1–30 | 1315440 |

Fig. 5 shows the modeling results of the best models with different number of explanatory variables. The results indicate that the performance of linear regression models reached the highest level when three to four explanatory variables were used. The use of additional variables did not improve the result. When the MLR models built using the exhaustive search and forward selection are compared, it is obvious that the exhaustive search approach resulted in better prediction accuracy only when two explanatory variables were used. This clearly demonstrates the impracticality of the exhaustive search approach for more than two explanatory variables in this case. The result also implies that the reduced candidate sets in exhaustive search had weaker predictive power compared with the full set of candidates even when suboptimal variable combinations from the full set were used. However, the differences in the model performances are small as shown in Fig. 5.

The number of latent variables in PLSR was in the range from one up to the number of explanatory variables in each model. The best modeling results according to $R^2$ with one to four explanatory variables were obtained using the same number of latent variables as there were explanatory variables. Therefore, these models were effectively MLR models. This indicates that the identification of latent variables from the original data did not improve the models. The best performing models with 5–10 explanatory variables had four latent variables each, but the model performance did not increase in comparison with the models that included fewer explanatory variables.

Fig. 5 reveals that GRNN had better performance compared with linear regression models in model training. This implies that the neural network learned the training data effectively. However, the performance in model testing was almost the same with linear models when one to three explanatory variables were used. With four and five explanatory variables, the RMSE criterion for GRNN was clearly lower. This can be partly explained by the use of different response variable in the model. The response variable for the best linear regression models with 1–10 explanatory variables and GRNN with 1–3 explanatory variables was '$\log_{10}(\sum l_{0.1})$' with steel cut effect removal. The response variable for the best performing GRNN with four explanatory variables was '$(\sum l_{0.1})^{1/2}$'. With five explanatory variables, it was '$(\sum l_{0.5})^{1/2}$'. However, the correlation coefficient of GRNN models was only slightly higher in comparison with the linear models.
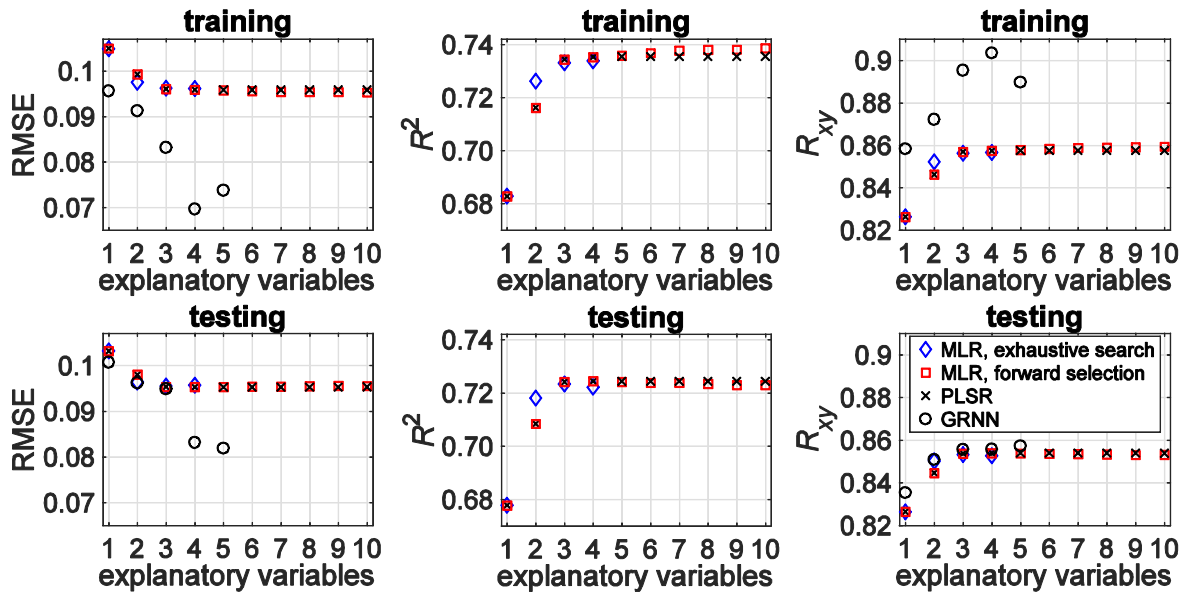


**Fig. 5** Cross-validation results from training and testing for the models with the best performance. The number of explanatory variables is shown on the horizontal axes and the values of performance criteria RMSE, $R^2$, and $R_{xy}$ are shown on the vertical axes

Fig. 6 shows the maximum VIF values for the models with two to five explanatory variables. The VIF values were computed using data from all 752 steel strips. The results reveal that all the models had VIF value over 10 when five explanatory variables were used. All the MLR models built using the exhaustive search approach and the MLR models built using forward selection with 1–3 explanatory variables had VIF value smaller than five indicating the absence of collinearity. Fig. 6 also shows that VIF values increased together with the number of explanatory variables, when variables were selected using forward selection and the same response variable was used in the models. In contrast, the maximum VIF may also decrease as variables are added in models built using exhaustive search, as shown in Fig. 6. The collinearity has to be taken into consideration in the case of PLSR, because the models with 1–4 explanatory variables were MLR models in practice. Based on the maximum VIF values, the PLSR models with fewer than four explanatory variables had an acceptable level of collinearity. Fig. 6 reveals that the GRNN model had multicollinearity when five explanatory variables were used. The rightmost graph shows the maximum VIF values when '$\log_{10}(\sum l_{0.1})$' with steel cut effect removal was used as the response variable in GRNN. In that case, VIF values were on the acceptable level when fewer than four explanatory variables were used.
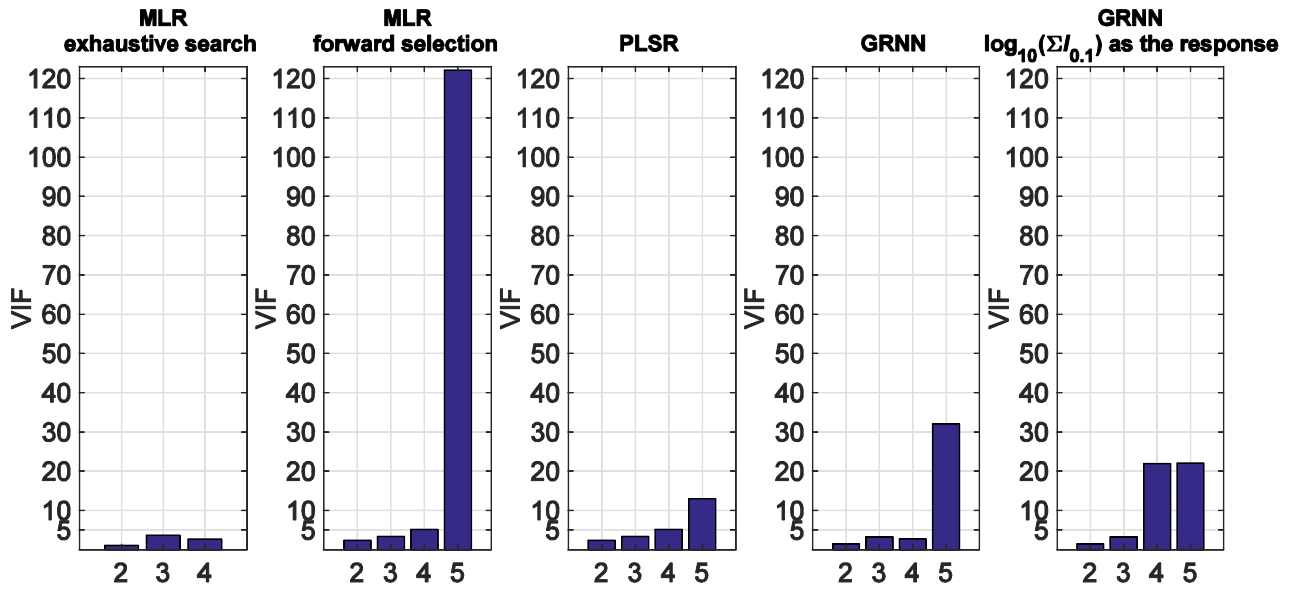


**Fig. 6** Maximum variance inflation factor (VIF) in the best performing models with two to five explanatory variables. The horizontal axes show the number of explanatory variables and the bars show the corresponding maximum VIF value in the particular model

Table 3 summarizes the modeling results for the best models while the VIF value of explanatory variables was allowed to be less than five. The results of the linear models show that the training and testing results were consistent. Over-fitted models in training and overly optimistic testing results were avoided with the applied procedure. The results of GRNN models show that the neural network learned the data slightly better, but the prediction accuracy for the test data was similar to the linear models. This becomes clearly evident by comparing models I, II, and IV, which had the same response variable '$\log_{10}(\sum l_{0.1})$' with steel cut effect removal. The GRNN (model III) had the best performance based on the RMSE, but the $R_{xy}$ of testing was almost the same in comparison with other models.

The explanatory variables selected for the presented models are given in Table 4. The first variable that was selected for the linear regression models by using forward selection was '$\log_{10}(\text{strength}) \cdot \log_{10}(\text{length})$'. This indicates that it had the strongest linear correlation with the response variable '$\log_{10}(\sum l_{0.1})$' with steel cut effect removal. This result indicates that the application of a mathematical transformation of the features improved their correlation. In fact, all the variables in Table 4 are mathematical transformations or interaction terms of steel strip features, such as length, yield strength, thickness, and weight. The width of steel strip or its transformations were not included in the selected variables.

The performances of the best MLR model (II) and the best GRNN model (III) are illustrated in Fig. 7. These models were trained using all 752 steel strips. The observed values of the response variable in the case of MLR are mainly scattered in the range 0.1–0.9. In the case of GRNN, the majority of the observed points lies in the range 0–0.6 and only a small

sample of points is above that. However, the observed values above 0.8 and the corresponding model predictions agree quite well. The vast majority of predicted values lie in the range 0–0.54. In the case of MLR, all of the predicted values lie roughly in the range 0.16–0.82. The linear model was unable to predict the values outside this range correctly.

Fig. 7 shows that the relative stress inflicted by steel grade B, which was the most common steel grade leveled, was broadly scattered when using either of the response variables selected for models II and III. In the case of MLR model on the left, the residuals of prediction were larger. The observations from steel grades C and D, which are relatively strong, seem to agree with the predictions quite well considering both models. The observed values for these steel grades were in the range 0.59–0.83 on the left and in the range 0.25–0.54 on the right in Fig. 7. Considering the distributions of observed stress values, these ranges indicate that the relative stress was relatively high when these grades were processed. Steel grade A, which has the lowest yield strength of the presented grades seems to inflict different stress compared with the prediction especially when the MLR predictions on the left are considered.

**Table 3** Performance of the best models

|  |  | RMSE | | $R_{xy}$ | | $R^2$ | | maximum VIF |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Model no. | Training | Testing | Training | Testing | Training | Testing | |
| MLR (exhaustive search) | I | 0.0962 | 0.0955 | 0.8562 | 0.8532 | 0.7331 | 0.7233 | 3.6794 |
| MLR and PLSR (forward selection) | II | 0.0960 | 0.0953 | 0.8569 | 0.8537 | 0.7343 | 0.7241 | 3.3213 |
| GRNN | III | 0.0696 | 0.0831 | 0.9034 | 0.8557 | | | 2.7433 |
| GRNN (logarithmic response variable) | IV | 0.0831 | 0.0949 | 0.8952 | 0.8556 | | | 3.2413 |

**Table 4** Explanatory variables in the models

| Model no. | 1st variable | 2nd variable | 3rd variable | 4th variable |
| --- | --- | --- | --- | --- |
| I | $\log_{10}(\text{length})$ | $\text{thickness}^2$ | $\log_{10}(\text{strength}) \cdot \log_{10}(\text{weight})$ | |
| II | $\log_{10}(\text{strength}) \cdot \log_{10}(\text{length})$ | $\text{thickness}^3$ | $\text{strength}^{1/2} \cdot \text{weight}$ | |
| III | $\log_{10}(\text{strength}) \cdot \log_{10}(\text{length})$ | $\text{strength}^3 \cdot \text{thickness}^3$ | $\text{thickness}^{1/3} \cdot \text{weight}$ | $\text{weight}^3 \cdot \text{thickness}^3$ |
| IV | $\log_{10}(\text{strength}) \cdot \log_{10}(\text{length})$ | $\text{thickness}^2 \cdot \text{strength}$ | $\text{weight}^2 \cdot \text{strength}$ | |
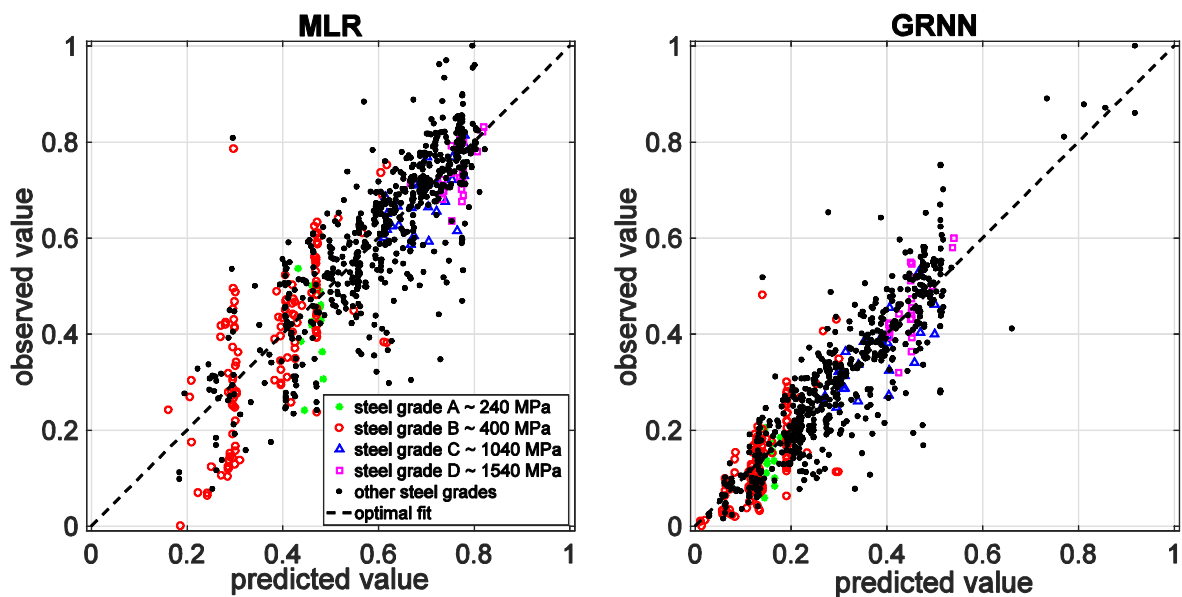


**Fig. 7** Performance of models II (left) and III (right) on training data consisting of 752 steel strips

Table 5 gives the parameters for model II trained using all 752 steel strips. The parameter significance is assessed using the *p*-value of *F*-test. The results indicate that each parameter is significant for the model. The spread parameter $\sigma$ in the best GRNN model (III) was 0.05.

**Table 5** Parameters in MLR trained using all 752 steel strips. The observed values of explanatory variables and the predicted values are defined on the range 0–1

| Model II | | | | |
|---|---|---|---|---|
| Response variable | $y = \log_{10}(\sum l_{0.1})$ with steel cut effect removal | | | |
| Observed range | $0.5296 \leq y \leq 2.9382$ | | | |
| | | | | |
| Explanatory variables | Observed range | Regression coefficients | *p*-value | VIF |
| | | $\beta_0 = 0.3525$ | $3.3674 \cdot 10^{-86}$ | |
| $x_1 = \log_{10}(\text{strength}) \cdot \log_{10}(\text{length})$ | $4.6368 \leq x_1 \leq 8.8844$ | $\beta_1 = 0.3659$ | $3.7458 \cdot 10^{-40}$ | 3.3213 |
| $x_2 = \text{thickness}^3$ | $7.7624 \leq x_2 \leq 3.5187 \cdot 10^3$ | $\beta_2 = -0.2069$ | $1.1597 \cdot 10^{-27}$ | 2.5736 |
| $x_3 = \text{strength}^{1/2} \cdot \text{weight}$ | $1.5061 \cdot 10^5 \leq x_3 \leq 1.1579 \cdot 10^6$ | $\beta_3 = 0.1464$ | $3.7936 \cdot 10^{-12}$ | 1.4951 |

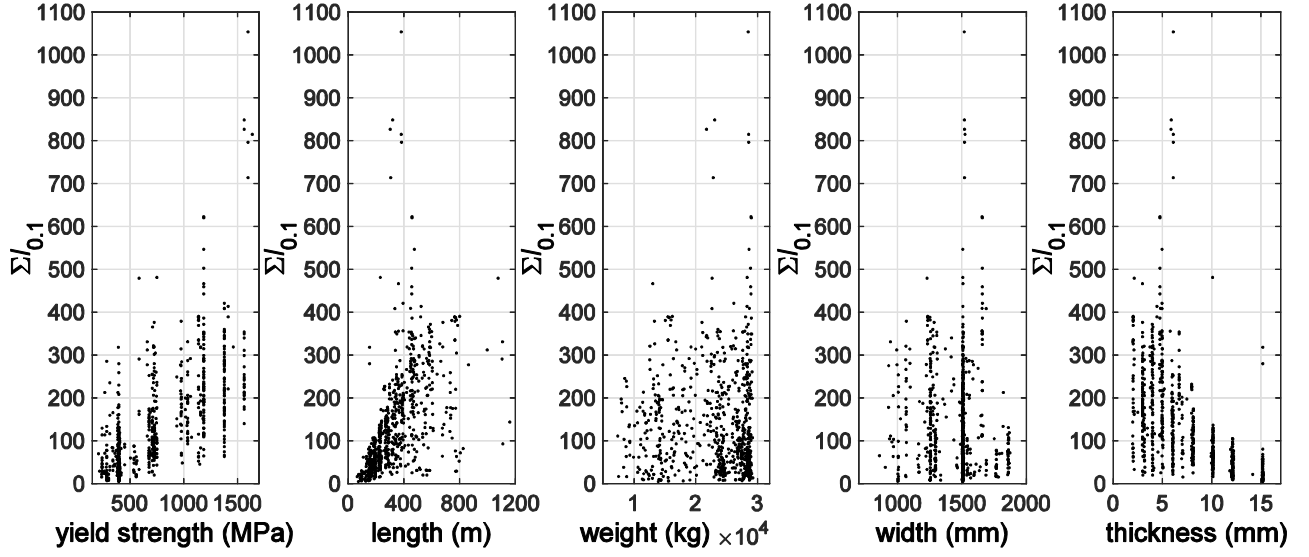Correlations between steel strip properties and vibration features

Fig. 8 illustrates the correlations between the steel strip properties and vibration feature '$\sum l_{0.1}$' without scaling. The positive correlations of the vibration feature with yield strength and strip length and negative correlation with strip thickness are evident. Weight and width have vague correlations with the vibration feature. These observations indicate that the norm sum is influenced by the joint effect of steel strip properties rather than the effect of a single property such as the length or thickness.

Fig. 9 illustrates the effect of steel cutting on vibration feature '$\sum l_{10}$' without scaling. The feature values without removing the cut effect are presented on the left and the values with steel cut effect removal are shown on the right. The feature is presented as a function of yield strength (above) and steel thickness (below). Fig. 9 clearly shows that the processing of certain types of steel strips inflicts shocks that have a major effect on the norm sums. It seems that the steel cut effect is pronounced on signals measured during the leveling of 5–8 mm thick strips and strips with a yield strength from 900 MPa upwards. Otherwise, the removal of the steel cut effect has a relatively minor influence on the sums. These observations demonstrate the importance of removing the steel cut effect when analyzing the relative stress inflicted by the leveler operation alone. However, the shocks caused by steel cutting probably stress the roller leveler as well.

Pearson's linear correlation coefficients between the steel strip properties and vibration feature '$\sum l_{0.1}$' for four example steel grades are presented in Table 6. The correlations for different steel grades vary significantly. For instance, the correlation of length is close to zero for grade A, but seems to increase together with increasing yield strength, while grade B is around 400 MPa, grade C is around 1040 MPa, and grade D is around 1540 MPa. The correlation of yield strength, on the other hand, is around -0.5 for grades A and D, whereas it is close to zero for grades B and C. Steel grades A and D had slightly more varying yield strength than grades B and C, and this resulted in higher correlation. The non-consistent correlations in Table 6 indicate that tailored stress models for different steel grades could improve the modeling performance for particular steel grades. However, the development of such models requires a larger data set, because most of the 55 steel grades in this study included only a few strips in the data set analyzed.

**Table 6** Linear correlations between steel strip properties and $\sum l_{0.1}$ for four different steel grades

| | grade A (12 strips) | grade B (123 strips) | grade C (27 strips) | grade D (23 strips) |
|---|---|---|---|---|
| yield strength | -0.5050 | 0.0072 | -0.0849 | -0.5224 |
| length | -0.0772 | 0.5378 | 0.6736 | 0.7092 |
| weight | 0.0345 | 0.0707 | 0.4162 | 0.1453 |
| width | 0.0773 | -0.0676 | -0.0601 | 0.0749 |
| thickness | 0.0855 | -0.5281 | -0.2483 | -0.5556 |

**Fig. 8** Scatter plots of steel strip properties with $\sum l_{0.1}$
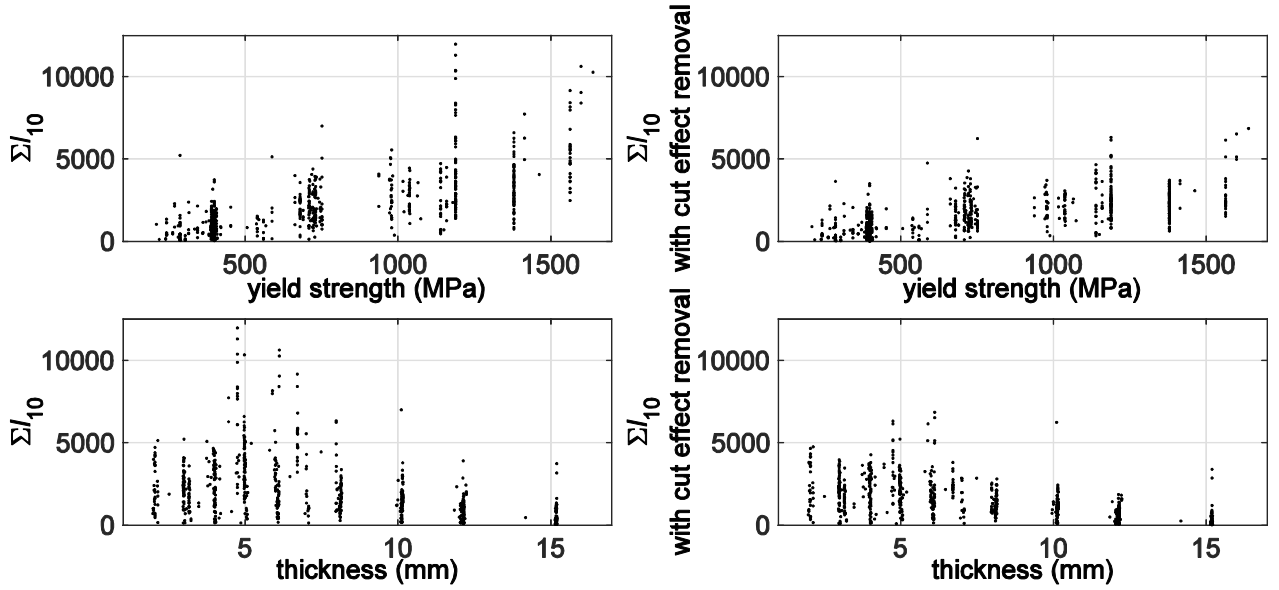


**Fig. 9** The effect of shocks caused by steel cutting on the feature $\sum l_{10}$. The effect is included in the feature values in the plots on the left and removed on the right

## Discussion

The modeling results indicate that combinations of steel strip properties could be used to predict the mechanical stress inflicted on the roller leveler. The results also show that the application of mathematical transformations of the features increased the linear correlations between the steel strip properties and vibration features. In this case, the vibration feature '$\log_{10}(\sum l_{0.1})$' with steel cut effect removal was the best response variable for MLR. The best response variable for GRNN was '$(\sum l_{0.1})^{1/2}$'. These results indicate that a low-order norm was generally more appropriate than a high-order norm in the applied vibration features. The sorted values of the selected features are shown in Fig. 10, which demonstrates the effects of mathematical transformations on the features. The use of logarithm on '$\sum l_{0.1}$' made the relative stress values spread more evenly in the relative stress range 0–1, while the values of '$\sum l_{0.1}$' had a distribution with strongly positive skew. A steep deviation can be seen at both ends of the '$\log_{10}(\sum l_{0.1})$' curve. MLR was not able to estimate these extreme values correctly, which was also evident in Fig. 7. The use of the square root in '$\sum l_{0.1}$' changed the distribution of relative

stress so that the majority of points were in the range 0–0.6. As shown in Fig. 7, GRNN was also able to learn the values above this range quite well.
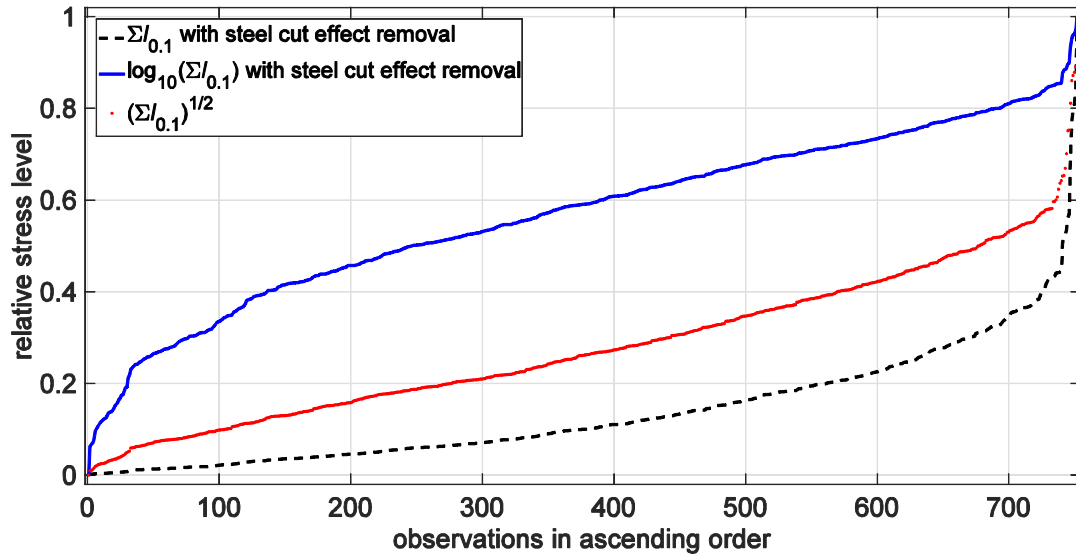


**Fig. 10** The observed relative stress during the processing of 752 steel strips in ascending order illustrating the distribution of the values in the stress range 0–1

The usability and reliability of prediction models in industrial practice are significant matters. Linear models often have advantage over the non-linear models in terms of interpretability. As shown in Table 5, the best MLR model can be used with four model parameters, which means it can be applied by using standard office software in a straightforward manner. In contrast, the best GRNN model has to remember the hundreds of training patterns and a multiple number of connection weights defined during model training. Consequently, the explicit analysis of model parameters and the transferability to standard office software bring challenges. However, there are ways to reduce the amount of training patterns, such as clustering (Specht 1991), but that is a topic for another study. Another weakness in the non-linear data-driven models may be their performance when new data are introduced. Steel mills often manufacture a large range of products and new products are continuously developed. The trained models may become repeatedly outdated as new products arrive. The transparency of MLR enables reasonable testing for new data, because the effect of explanatory variable manipulation on the model response can be easily interpreted. Such testing is considerably more difficult with complex models. Moreover, the GRNN predictions had a clear difference between the training and testing results, while MLR produced consistent results. These considerations indicate that the MLR is the most reasonable model option from the tested models for practical stress prediction. Other prediction algorithms presented in the literature could be tested in future investigations. The testing of alternative variable selection methods could potentially lead into improved models as well.

The strong steel grades C and D resulted in higher observed values of relative stress compared with the values from steel grades A and B which was also demonstrated in Fig. 7. The MLR predicted the stress inflicted by these strong steel grades more accurately in comparison with the most common steel grade B. This can be explained by the effect of the logarithm on high values. In the case of GRNN, differences in the prediction accuracy for different steel grades were smaller. The generation of steel grade specific stress models is a potential topic for future development. The data set should be more extensive than the one used in this study. The studied data set had 55 steel grades, but the number of strips representing each grade varied and consequently the data were dominated by certain steel grades. However, the general correlations between steel strip properties and vibration features were discovered using this approach. Nonetheless, the effects of relative stress values still need to be investigated in future studies. This could be done by estimating the stress of leveling events in process history and then investigating the relation of the estimates with fault and maintenance history. However, the use of vibration data as the response of machine operation can already be considered as an improved approach to stress evaluation compared with assessments relying solely on production data.

The combination of work roll movement or motor power with the vibration signal, which was not possible in this study, could improve the reliability of the proposed approach. It would then be certain that the effects of the idle state and possible measurement disturbances during the idle state could be avoided in the norm sums. The operational parameters of the roller leveler and steel strip properties correlate strongly (Nikula and Karioja 2016). Therefore, based on the results, it is assumed that the steel strip properties alone could be used for stress prediction. On the other hand, the prediction accuracy shows variation that cannot be explained based solely on the effects of steel strip properties. The combination of instantaneous operational parameter values with the vibration signal could provide additional possibilities for stress estimation.

The steel cut effect was removed from the response variable of the best linear model. Therefore, the values of this variable mainly represent the effects incited by the leveling operation and the influence of steel cutting is mostly avoided. This also means that the presented observations could be useful in the development of monitoring approaches to other roller levelers that are not influenced by steel cutting.

The relative stress proposed in this study could be used for the estimation of condition deterioration in roller levelers. Plenty of recent studies focus on the prediction of remaining useful life and prognosis of deteriorating systems (Benkedjouh et al. 2015; He et al. 2012; Mosallam et al. 2016; Ragab et al. 2016; Shi and Zheng 2016; Son et al. 2016). The proposed relative stress features could be used as the monitored indicators in these kind of approaches. These features could be utilized in risk assessments that are based on cumulative stress.

**Conclusions**

In this paper, an approach to the prediction of the relative stress inflicted on a roller leveler was introduced and validated using measurements from an industrial case. The stress estimates were based on the sums of generalized norms computed from an acceleration signal acquired during the leveling of steel strips. Regression models were used to identify the steel strip properties and combinations of them that could be used to explain the values of the norm sums. The mathematical transformations of steel strip properties and norm sums improved their linear correlation, which consequently improved the performance of the linear regression models. The generalized regression neural network had the best prediction accuracy from the tested models, but the superiority over other models was remote. In addition, the neural network structure is complicated in comparison with the tested linear regression models, and therefore its application potential may be limited in practice. The results indicate that the stress effects seen during the processing of various steel grades were diverse, and consequently, more elaborate modeling approaches could improve stress predictions in some cases. However, a regression model trained based on extensive measurement data is an advanced approach to stress prediction when compared with assessments made only on the basis of production data. The use of relative stress for long-term risk assessment is a topic for future research.

**References**

Alaei, H., Salimi, M., & Nourani, A. (2016). Online prediction of work roll thermal expansion in a hot rolling process by a neural network. *International Journal of Advanced Manufacturing Technology*, 85(5), 1769-1777.

Bar, A., & Bar, O. (2005). Types of mid-frequency vibrations appearing during the rolling mill operation. *Journal of Materials Processing Technology*, 162-163, 461-464.

Bar, A., & Świątoniowski, A. (2004). Interdependence between the rolling speed and non-linear vibrations of the mill system. *Journal of Materials Processing Technology*, 155-156, 2116-2121.

Baumgart, M., Steinboeck, A., Kiefer, T., & Kugi, A. (2015). Modelling and experimental validation of the deflection of a leveller for hot heavy plates. *Mathematical and Computer Modelling of Dynamical Systems*, 21(3), 202-227.

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2015). Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing*, 26(2), 213-223.

Bruschi, S., Altan, T., Banabic, D., Bariani, P. F., Brosius, A., Cao, J., Ghiotti, A., Khraisheh, M., Merklein, M., & Tekkaya, A. E. (2014). Testing and modelling of material behaviour and formability in sheet metal forming. *CIRP Annals – Manufacturing Technology*, 63(2), 727-749.

Bullen, P. S. (2003). *Handbook of Means and their Inequalities*, 2nd edition. Dordrecht: Kluwer Academic Publishers.

Chen, W., Liu, J., Cui, Z., Wang, Y., & Wang, Y. (2015). A 2.5-dimensional analytical model of cold leveling for plates with transverse wave defects. *International Journal of Iron and Steel Research*, 22(8), 664-671.

Cui, L., Hu, X., & Liu, X. (2011). Analysis of leveling strategy for a plate mill. *Advanced Materials Research*, 145, 424-428.

Das, P., & Datta, S. (2007). Exploring the non-linearity in empirical modelling of a steel system using statistical and neural network models. *International Journal of Production Research*, 45(3), 699-717.

Doege, E., Menz, R., & Huinink, S. (2002). Analysis of the levelling process based upon an analytic forming model. *CIRP Annals – Manufacturing Technology*, 51(1), 191-194.

Dong, L., Li, S., He, J., & Cui, R. (2016). Investigation on shearing and local formability of hot-rolled high-strength plates. *ASME Journal of Manufacturing Science and Engineering*, 138(9).

Dratz, B., Nalewajk, V., Bikard, J., & Chastel, Y. (2009). Testing and modelling the behaviour of steel sheets for roll levelling applications. *International Journal of Material Forming*, 2, 519-522.

Faris, H., Sheta, A., & Öznergiz, E. (2013). Modelling hot rolling manufacturing process using soft computing techniques. *International Journal of Computer Integrated Manufacturing*, 26(8), 762-771.

He, D., Li, R., & Bechhoefer, E. (2012). Stochastic modeling of damage physics for mechanical component prognostics using condition indicators. *Journal of Intelligent Manufacturing*, 23(2), 221-226.

Heidari, A., & Forouzan, M. R. (2013). Optimization of cold rolling process parameters in order to increasing rolling speed limited by chatter vibrations. *Journal of Advanced Research*, 4(1), 27-34.

Huh, H., Heo, J. H., & Lee, H. W. (2003). Optimization of a roller levelling process for A17001T9 pipes with finite element analysis and Taguchi method. *International Journal of Machine Tools and Manufacture*, 43, 345-350.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer-Verlag.

Jantunen, E., & Vaajoensuu, E. (2010). Self adaptive diagnosis of tool wear with a microcontroller. *Journal of Intelligent Manufacturing*, 21(2), 223-230.

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20, 1483-1510.

Juuso, E. K. (2014). Intelligent indices for online monitoring of stress and condition. In *The 11th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, Manchester, UK (pp. 637-648).

Juuso, E. K., & Galar, D. (2016). Intelligent real-time risk analysis for machines and process devices. In U. Kumar et al. (Ed.), *Current Trends in Reliability, Availability, Maintainability and Safety: an Industrial Perspective* (pp. 229-240). Switzerland: Springer.

Juuso, E., & Lahdelma, S. (2010). Intelligent scaling of features in fault diagnosis. In *The Seventh International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, Stratford-upon-Avon, UK (pp. 1358-1372).

Juuso, E., & Ruusunen, M. (2013). Fatigue prediction with intelligent stress indices based on torque measurements in a rolling mill. In *The Tenth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, Kraków, Poland (pp. 460-471).

Karioja, K., & Lahdelma, S. (2013). Applying acceleration and strain signals for the stress evaluation of a steel cutter. In *The Tenth International Conference on Condition Monitoring and Machine Failure Prevention Technologies*, Kraków, Poland (pp. 246-258).

Karioja, K., & Lahdelma, S. (2015). Applying weighted $l_p$ norms and MIT measurement indices in the stress evaluation of a steel cutter. *International Journal of Condition Monitoring*, 5(1), 20-27.

Karioja, K., Nikula, R.-P., & Liedes, T. (2015). Vibration measurements and signal processing in stress monitoring of a steel leveler. In *Maintenance, Condition Monitoring and Diagnostics & Maintenance Performance Measurement and Management: MCMD 2015 and MPMM 2015*, Oulu, Finland.

Kim, B., Kwon, S., & Kim D. H. (2010). Optimization of optical lens-controlled scanning electron microscopic resolution using generalized regression neural network and genetic algorithm. *Expert Systems with Applications*, 37, 182-186.

Lahdelma, S., & Juuso, E. K. (2008). Signal processing in vibration analysis. In *The Fifth International Conference on Condition Monitoring and Machine Failure Prevention Technologies*, Edinburgh, UK (pp. 867-878).

Lahdelma, S., & Juuso, E. K. (2011a). Signal processing and feature extraction by using real order derivatives and generalised norms. Part 2: Applications. *International Journal of Condition Monitoring*, 1(2), 54-66.

Lahdelma S., & Juuso, E. K. (2011b). Signal processing and feature extraction by using real order derivatives and generalised norms. Part 1: Methodology. *International Journal of Condition Monitoring*, 1(2), 46-53.

Laurinen, P., & Röning, J. (2005). An adaptive neural network model for predicting the post roughing mill temperature of steel slabs in the reheating furnace. *Journal of Materials Processing Technology*, 168(3), 423-430.

Lee, D. Y., Cho, H. S., & Cho, D. Y. (2000). A neural network model to determine the plate width set-up value in a hot plate mill. *Journal of Intelligent Manufacturing*, 11(6), 547-557.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems – Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 42, 314-334.

Li, R., Sopon, P., & He, D. (2012). Fault features extraction for bearing prognostics. *Journal of Intelligent Manufacturing*, 23(2), 313-321.

Liu, Z., Wang, Y., & Xingchun, Y. (2012). A new model for the plate leveling process based on curvature integration method. *International Journal of Mechanical Sciences*, 54(1), 213-224.

Madej, L., Muszka, K., Perzyński, K., Majta, J., & Pietrzyk, M. (2011). Computer aided development of the levelling technology for flat products. *CIRP Annals – Manufacturing Technology*, 60(1), 291-294.

Matsuzaki, K., Sueoka, A., Ryu, T., & Morita, H. (2008). Generation mechanism of polygonal wear of work rolls in a hot leveler and a countermeasure by dynamic absorbers. *International Journal of Machine Tools and Manufacture*, 48(9), 983-993.

May, R. Dandy, G., & Maier, H. (2011). Review of input variable selection methods for artificial neural networks. In K. Suzuki (Ed.), *Artificial Neural Networks – Methodological Advances and Biomedical Applications* (pp.19-44). InTech.

Morris, J. W., Hardy, S. J., Lees, A. W., & Thomas, J. T. (2001). Formation of residual stresses owing to tension levelling of cold rolled strip. *Ironmaking and Steelmaking*, 28(1), 44-52.

Mosallam, A., Medjaher, K., & Zerhouni, N. (2016). Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *Journal of Intelligent Manufacturing*, 27(5), 1037-1048.

Nikula, R.-P., & Karioja, K. (2016). The effect of steel leveler parameters on vibration features. In *Proceedings of the 9th EUROSIM Congress on Modelling and Simulation*, Oulu, Finland (pp. 392-397).

Nizioł, J., & Świątoniowski, A. (2005). Numerical analysis of the vertical vibrations of rolling mills and their negative effect on the sheet quality. *Journal of Materials Processing Technology*, 162-163, 546-550.

Park, K., & Hwang, S.-M. (2002). Development of a finite element analysis program for roller leveling and application for removing blanking bow defects of thin steel sheet. *ISIJ International*, 42(9), 990-999.

Picard, R. R., & Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387), 575-583.

Ragab, A., Ouali, M.-S., Yacout, S., & Osman, H. (2016). Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan-Meier estimation. *Journal of Intelligent Manufacturing*, 27(5), 943-958.

Rosipal, R., & Krämer, N. (2006). Overview and Recent Advances in Partial Least Squares. In: C. Saunders et al. (Ed.), *Subspace, Latent Structure and Feature Selection. Lecture Notes in Computer Science*, 3940. Berlin, Heidelberg: Springer.

Seo, J. H., Van Tyne, C. J., & Moon, Y. H. (2016). Effect of roll configuration on the leveling effectiveness of tail-up bent plate using finite-element analysis. *ASME Journal of Manufacturing Science and Engineering*, 138(7).

Serdio, F., Lughofer, E., Pichler, K., Buchegger, T., & Efendic H. (2014). Residual-based fault detection using soft computing techniques for condition monitoring at rolling mills. *Information Sciences*, 259, 304-320.

Shao, J. (1993). Model selection by cross-validation. *Journal of the American Statistical Association*, 88(422), 486-494.

Shi, H., & Zeng, J. (2016). Real-time prediction of remaining useful life and preventive opportunistic maintenance strategy for multi-component systems considering stochastic dependence. *Computers & Industrial Engineering*, 93, 192-204.

Silvestre, E., Sáenz de Argandoña, E., Galdos, L., & Mendiguren, J. (2014). Testing and modeling of roll levelling process. *Key Engineering Materials*, 611-612, 1753-1762.

Silvestre, E., Mendiguren, J., Galdos, L., & Sáenz de Argandoña, E. (2015). Comparison of the hardening behaviour of different steel families: from mild and stainless steel to advanced high strength steels. *International Journal of Mechanical Sciences*, 101-102, 10-20.

Smith Jr., R. P. (1997). Flatness control in coiled plates: Lukens' wide, cut to length line. *Iron and Steel Engineer*, May 1997, 29-34.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Journal of Statistics and Computing*, 14(3), 199-222.

Son, J., Zhou, S., Sankavaram, C., Du, X., & Zhang, Y. (2016). Remaining useful life prediction based on noisy condition monitoring signals using constrained Kalman filter. *Reliability Engineering and System Safety*, 152, 38-50.

Specht, D. F. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6), 568-576.

Spiess, A.-J., & Neumeyer, N. (2010). An evaluation of $R^2$ as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacology*, 10(6).

Sriram, S., Yao, H., & Ramisetti, N. (2012). Development of an empirical model to characterize fracture behavior during forming of advanced high strength steels under bending dominated conditions. *ASME Journal of Manufacturing Science and Engineering*, 134(3).

Sueoka, A., Matsuzaki, K., & Ryu, T. (2002). Polygonal wear of work rolls in a hot leveler of steel making machine. *JSME International Journal Series C*, 45(3), 673-681.

Usamentiaga, R., García, D. F., Molleda, J., Bulnes, F. G., & Braña, G. B. (2014). Vibrations in steel strips: effects on flatness measurement and filtering. *IEEE Transactions on Industry Applications*, 50(5), 3103-3112.

Usamentiaga, R., Molleda, J., Garcia, D. F., Bulnes, F. G., Entrialgo, J., & Suárez Alvarez, C. M. (2015). Flatness measurement using two laser stripes to remove the effects of vibrations. *IEEE Transactions on Industry Applications*, 51(5), 4297-4304.

Wang, W. (2007). A prognosis model for wear prediction based on oil-based monitoring. *Journal of the Operational Research Society*, 58, 887–893.

Whitney, A. W. (1971). A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, C-20, 1100-1103.

Wise, B. M., & Gallagher, N. B. (1996). The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*, 6(6), 329-348.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109-130.

Wu, S., Wang, L., Shao, Y., & Yuan, Y. (2014). Vibration characteristic analysis of twenty-high rolling mill with local defect on roll surface based on time-varying contact stiffness. *Engineering Failure Analysis*, 42, 297-310.

Wu, S., Shao, Y., Wang, L., Yuan, Y., & Mechefske, C. K. (2015). Relationship between chatter marks and rolling force fluctuation for twenty-high roll mill. *Eng