Visual analysis of a cold rolling process using data-based modeling

Daniel Pérez[1], Francisco J. García-Fernández[1], Ignacio Díaz[1], Abel A. Cuadrado[1], Daniel G. Ordonez[1], Alberto B. Díez[1], and Manuel Domínguez[2]

[1] Universidad de Oviedo. Área de Ingeniería de Sistemas y Automática
{dperez,fjgarcia,idiaz,dgonzalez,cuadrado,alberto}@isa.uniovi.es
[2] Universidad de León. Instituto de Automática y Fabricación
{diemdg}@unileon.es

Abstract. In this paper, a method to characterize the chatter phenomenon in a cold rolling process is proposed. This approach is based on obtaining a global nonlinear dynamical MISO model, relating four input variables and the exit strip thickness as the output variable. In a second stage, local linear models are obtained for all working points using sensitivity analysis on the nonlinear model to get input/output small signal models. Each local model is characterized by a high dimensional vector containing the frequency response functions (FRF) of the four SISO resulting models. Finally, the FRF's are projected on a 2D space, using the t-SNE algorithm, in order to visualize the dynamical changes of the process. Our results show a clear separation between chatter condition and other vibration states, allowing an early detection of chatter as well as being a visual analysis tool to study the chatter phenomenon.

Key words: dimensionality reduction, cold rolling, data visualization, dynamical systems, data-based models.

1 Introduction

Steel production is usually considered an indicator of economic progress, as it is fairly related to infrastructures and development. After steel production from iron, the material needs to be treated and modified through several mechanical processes, such as the rolling process. The cold rolling of steel is a widely adopted process, in which a steel sheet is passed through a pair of rolls whereby the sheet thickness is reduced. Although this process has been studied for decades [1], many unsolved issues hold. The control of many different parameters is necessary, ranging from those related to the milling itself (force applied, torque,. . .) to those depending on different aspects, such as lubrication or refrigeration. Furthermore, there is an ever increasing demand for higher quality from costumers and, since it is a large and complex process that continuously evolves due to drifts, misadjustments and changes in working conditions, there is a need of continuous improvement in the efficiency of the process. Because of that, the supervision of this process is critical, in order to avoid faults that affect negatively to the material.

One of the most relevant faults in the cold rolling process of steel is called *chatter* [2], an unexpected powerful vibration that affects the quality of the rolled material by causing an unacceptable variation of the final thickness. The real problem of chatter is not only related to the bad quality of the manufactured product, but also to the economic losses suffered. Generally, when chatter appears, it is necessary to lower the rolling speed for a period of time, making the production rate decrease. A practical way to detect chatter is to compute the power spectral density in which this fault appears (normally 100-300Hz). However, although this procedure works well to show up the chatter condition, it fails as an early detector.

A way to predict chatter is to use a model of the rolling process [3]. However, the complexity of the whole process, with several tightly coupled phenomena (such as chemical, mechanical, and thermal) makes it difficult to build an accurate model and moreover to tune its parameters. An approach to enhance the knowledge about complex processes is visualizing their relevant information, using *dimensionality reduction* (DR) techniques [4, 5]. DR techniques allow to project and study the structure of high-dimensional data into a low-dimensional space, typically a 2D/3D for visualization purposes, improving the exploratory data analysis [6].

In the DR field, several techniques have been proposed [7]. One of the first algorithms is Principal Component Analysis (PCA), described by Pearson [8]. After PCA, other DR techniques have been proposed, such as Multidimensional Scaling (MDS) methods, Independent Component Analysis (ICA)[9] or Self-Organizing Maps (SOM)[10]. In the beginning of 21st century, a new trend in DR based on nonlinear models appeared, inspiring a new collection of algorithms. These algorithms –known as *manifold learning*– involve a local optimization by defining local models of the k-nearest neighbours and an alignment in order to obtain the global coordinates of each model, usually implying a singular value decomposition (SVD). Some of the most known techniques are *Isomap* [11], *local linear embedding*(LLE) [12] and *laplacian eigenmaps* (LE) [13]. Similar to these techniques, but based on the probability distribution of data is *t-Stochastic Neighbor Embedding* (*t*-SNE) [14]. This technique, that has attracted attention recently [15, 16], is capable of maintaining the local structure of the data while also revealing some important global structure (such as clusters at different scales), producing better visualizations than the rest.

In this paper, we propose a new approach for the study of chatter, using the DR principle for the analysis of the dynamical behavior of a model of the process. Using a novel feedforward neural network, called *extreme learning machine* (ELM) [17], the proposed approach computes a large feature vector composed of the frequency response functions (FRF) of a set of key physical variables and projects this vector into a 2D space by *t*-SNE algorithm. Thus, the changes in the dynamical behavior of the process are visualized. The paper is organized as follows: in section 2, a description of the method is shown; section 3 describes an experiment and the results of the method proposed and finally section 4 includes the conclusions obtained.

2 Data-based model analysis through manifold learning techniques

2.1 Description of the physical model

Classical cold rolling models try to calculate the force and the torque necessary for a given thickness reduction. As mentioned before, the complexity of an accurate model can be very high because of the assumptions taken [18]. In order to get a simple model to work with, e.g. [19], several assumptions can be done.

The classical form of a rolling force (F) model includes: the tension at the entry and exit side of a rolling stand (σ_{en} and σ_{ex}); the thickness at the entry and exit side (h_{en} and h_{ex}); the width of the strip (w); the friction coefficient (μ) and the hardness level of the material being rolled (S), see Eq. (1).

$$F = f(\sigma_{en}, \sigma_{ex}, h_{en}, h_{ex}, w, \mu, S) \tag{1}$$

A model of the rolling process makes it possible to analyze several defects arising from working operation, such as the chatter phenomenon. This phenomenon is a dynamic process, where variations in the roll force may lead to an unstable state. It is necessary to generate a model where the different factors likely to modify the force equilibrium in the rolling process are taken into account. As explained in [20], the chatter phenomenon comes from a feedback interaction with the variables entry speed, entry tension, the force of the strip on the rolls and exit thickness involved. If a dynamic model of the stand is added to this loop, a proper model to study the chatter phenomenon can be built [21, 22].

2.2 Mathematical estimation of the model

As proposed in [23], data-based models are a practical way to develop a fault detection and prediction mechanism for complex processes. The development of data-based models provides a good feature for their application to industrial processes: fast responses to faults. According to the previous description of the rolling process, we propose a MISO model, defining the exit thickness y_k as the output of the system and force F_k, tension σ_k –used in the classical model Eq. (1)–, entry and exit speed of the strip (V_{en_k} and V_{ex_k} respectively) –due to their relevance in the chatter phenomenon–, as the inputs of the system. We also considered an autoregressive part of the output to account for internal dynamics, resulting in a NARX model.

$$y_k = f(y_{k-1}, \ldots, y_{k-n}, F_k, \sigma_k, V_{en_k}, V_{ex_k}) \tag{2}$$

A simple and fast learning algorithm for single hidden layer feedforward neural networks (SLFN's), called extreme learning machine (ELM) [17], is used to train the NARX model. In order to obtain an optimal order of the model, we apply the Akaike Information Criterion (AIC) [24], obtaining a trade-off solution between the order and the error of the model. A scheme can be seen on the left side of Fig. 1. On the right side of Fig. 1, a comparison between the response
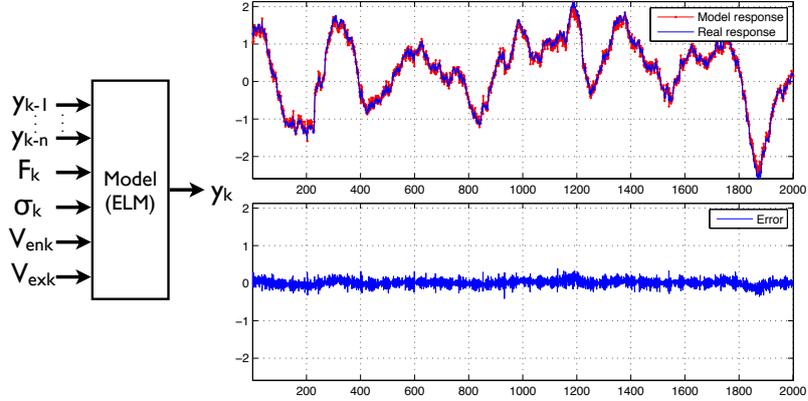
Fig. 1. Scheme of the model developed (left) and an example of its one-step-ahead prediction (right).

of the real process and an example of the one-step-ahead prediction of a trained model for a testing dataset is shown.

Once the model is obtained, a local sensitivity analysis is applied to perform a system identification of the contribution of each input variable to the output. The signals are divided into M overlapped windows of size L. Let's consider $\bar{F}$, $\bar{\sigma}$, $\bar{V}_{en}$, and $\bar{V}_{ex}$, average values for the m-th window, and the delayed samples $y_{k-1}, y_{k-2}, \dots, y_{k-n}$. The m-th local sensitivity analysis ($m = 1, \dots, M$) for the input F is performed adding to $\bar{F}$ a random value $\varepsilon_k \in N(0, \nu)$, being ν a small value.

$$\mathbf{u}_m^{\Delta F} = \left[y_{k-1}, y_{k-2}, \dots, y_{k-n}, \bar{F} + \varepsilon_k, \bar{\sigma}, \bar{V}_{en}, \bar{V}_{ex} \right]^T \tag{3}$$

with an output $\mathbf{y}_m = [y_k]$, for $k = k_m, \dots, k_m + L - 1$, being k_m the first sample of window m. Constructing $\mathbf{U}^{\Delta F} = [\mathbf{u}_m^{\Delta F}]$ and $\mathbf{Y} = [\mathbf{y}_m]$ resulting in an I/O pair $\left\{ \mathbf{U}^{\Delta F}, \mathbf{Y} \right\}$. Similar to Eq. (3), we apply the same method to the other inputs, obtaining $\left\{ \mathbf{U}^{\Delta \sigma}, \mathbf{Y} \right\}$, $\left\{ \mathbf{U}^{\Delta V_{en}}, \mathbf{Y} \right\}$ and $\left\{ \mathbf{U}^{\Delta V_{ex}}, \mathbf{Y} \right\}$ respectively. Each I/O pair defines a local SISO small-signal model of the process.

In order to estimate the dynamical behavior of each small-signal model, we compute FRF of the model. Let $P_y(m, j)$ and $P_{u_i}(m, j)$ be the power densities of the j-th frequency in the m-th small-signal model of the output and the i-th input, respectively. The SISO FRF of input i for all windows can be computed as $\mathbf{G}_i$ where

$$\mathbf{G}_i(m, j) = 10 \cdot log_{10} \left(\left| \frac{P_y(m, j)}{P_{u_i}(m, j)} \right| \right) \tag{4}$$

describes the gain of the j-th frequency bin for the m-th small-signal model, of the i-th input expressed in dB.

Finally, in order to project all data using a DR technique, all the FRF's of each input, $\mathbf{G}_i$, are joined into an augmented matrix $\mathbf{G}$, as expressed in Eq. (5).

$$\mathbf{G} = [\mathbf{G}_1 \ \mathbf{G}_2 \ \mathbf{G}_3 \ \mathbf{G}_4] \qquad \mathbf{G} \in \mathbb{R}^{M \times D} \tag{5}$$

Each of the M rows of $\mathbf{G}$ is a large D-dimensional feature vector which describes the local dynamical behavior on a given window.

2.3 Dimensionality reduction

The visualization of the dynamic behavior of the model is made by computing a dimensionality reduction using t-SNE [14]. This technique is based on similarities of the data by computing probabilities for original and projection space, defining neighborhoods with a value called *perplexity*. The computation of the two joint-probability distributions p_{ij} and q_{ij} corresponds to Gaussian and Student's t-distribution respectively, see Eq. (6) and (7).

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|^2/2\sigma^2)} \tag{6}$$

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{7}$$

With the aim of projecting new data points on the 2D map obtained with the training dataset, an out-of-sample extension of the t-SNE was developed. The addition of each new point was computed by ensuring that the sum of probabilities is equal to 1. The variance σ of a Gaussian centered at the new point is searched keeping the perplexity value fixed. The differences of the two probability distributions are measured by Kullback-Leibler divergences. Hence, the cost function is

$$C = KL(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{8}$$

The coordinates of the new projections are obtained minimizing this cost function, and maintaining previous coordinates of the training dataset fixed. This minimization of the cost function is made using a gradient-descent method for optimization. Finally, as a summary of the method, a flowchart of the different stages used is shown in Fig. 2.

3   Experiment and Results

To validate this method, it was applied to data from a cold rolling facility. As explained in section 2.2, the variables used for modeling were: rolling force, entry tension, entry and exit speeds of the strip, and exit thickness. These variables were acquired using a data acquisition system with a sampling rate of 2000 Hz.
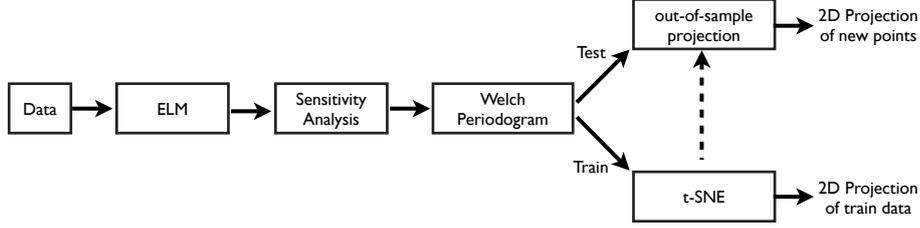
Fig. 2. Flowchart of the method.

The training dataset was composed of signals of $N = 54300$ samples, including normal operating conditions and a chatter episode from a unique coil. In Fig. 3, the reduction of the rolling speed to avoid chatter effect can be seen, as well as the powerful variation in the other variables, especially in the thickness signal, which is closely related to the quality of the resulting product. As for testing the method, different chatter episodes of several coils were used.

To obtain the MISO model, ELM was trained using 1000 neurons in the hidden layer, with the signals normalized to zero mean and $\sigma = 1$. In Table 1, a comparison of the RMS errors of the linear regression and the EML models is shown, as well as the AIC and the Theil's Index (U) [25], for several model orders.

Model Order (n)	Linear RMSE	EML RMSE	U	AIC
5	0.1856	0.1744	0.7377	$-1.5445 \cdot 10^5$
10	0.1353	0.1307	0.5529	$-1.7998 \cdot 10^5$
15	0.1252	**0.1159**	**0.4903**	$\mathbf{-1.9058 \cdot 10^5}$
20	0.1277	0.1192	0.5042	$-1.8809 \cdot 10^5$
25	**0.1251**	0.1167	0.4937	$-1.8993 \cdot 10^5$
30	0.1273	0.1194	0.5051	$-1.8787 \cdot 10^5$

Table 1. Value of the RMS errors, AIC, and Theil's Index (U) for each order of the NARX model.

According to the AIC criterion the order selected is $n = 15$. This model describes accurately the nonlinear dynamics around the working points of the process.

For the sensitivity analysis the signals were windowed into segments of $L = 1000$ elements with an overlapping of 90%. After this analysis, the Welch Periodogram was applied to each SISO model, using a Hamming window of size 125 and an overlapping of 50%. The final size of matrix $\mathbf{G}$ is 543×516. Using this matrix, the t-SNE algorithm was applied using a value of perplexity of 30 to project data into a 2D space.
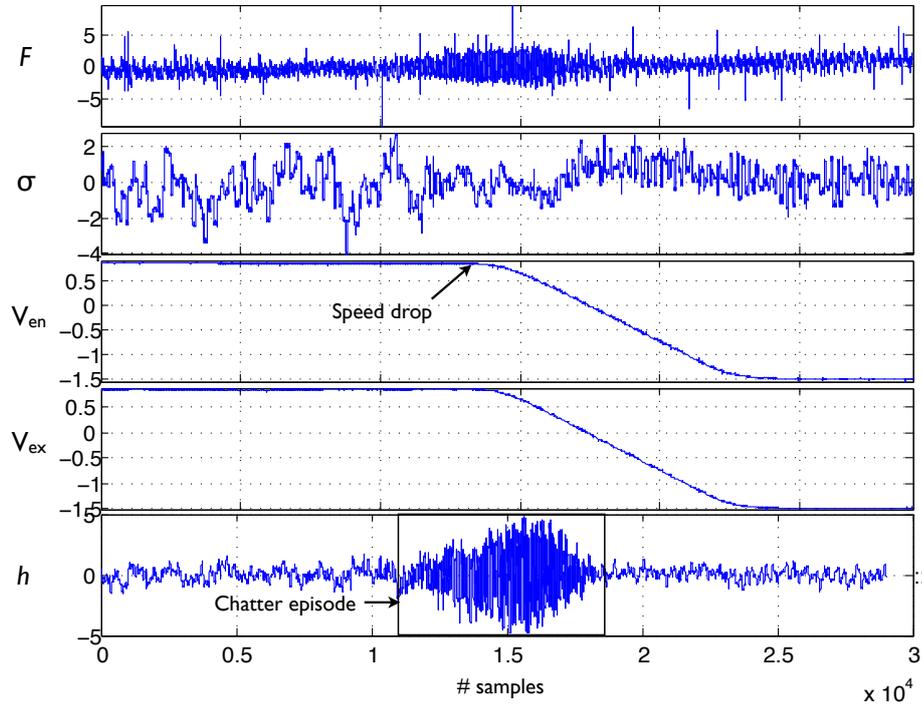
Fig. 3. Training dataset to model the process. When chatter occurs, a drop in the speed in order to reduce its effect can be seen. After achieving a proper speed level, the effect is mitigated.

The resulting map using t-SNE technique, can be seen in Fig. 4 (a), where the entry speed is used as color and size encodings of the points. The 2D map shows two main zones of points from normal behavior of the process and a subset of points, revealing a chatter condition.

The method was applied to two subsets of data from different strips, denoted as x and +, including a chatter episode. The developed out-of-sample extension allows to project this novel test data over the trained map. These new points are placed in parts of the map that corresponds to their operating conditions (Fig. 4 (b), (c) and (d)). Their positions reveal a similar behavior to the nearest neighbors of the training points. Thus, different operating conditions are mapped in different coordinates, allowing to detect the dynamical differences in the process behavior.
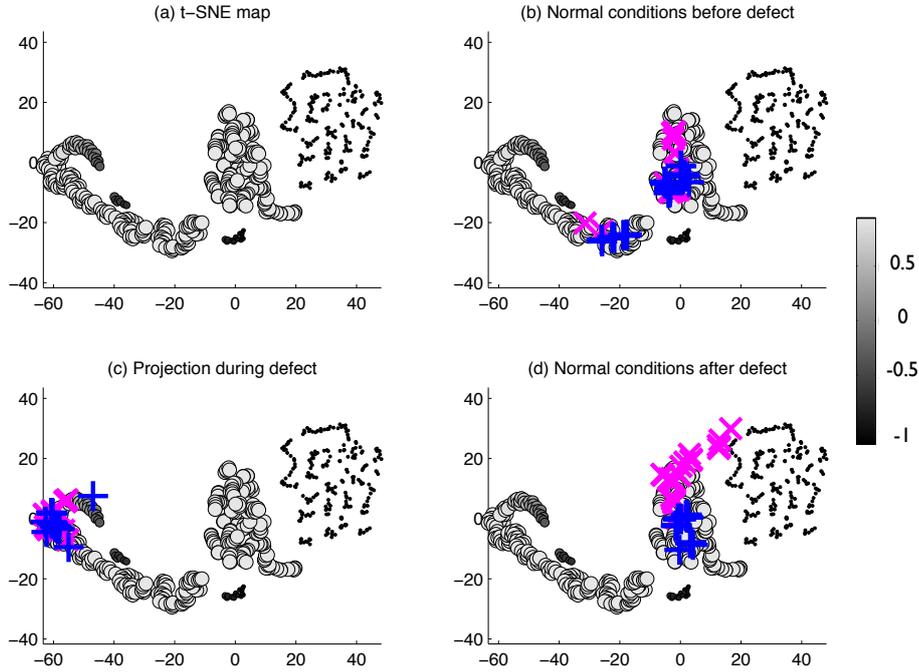
Fig. 4. (a) *t*-SNE projection, (b), (c), and (d) new points projected depending on operating condition.

4   Conclusions

A spectral density estimation was applied to a model of a cold rolling process in order to analyze the chatter defect, which causes an unexpected vibration in the process. The analysis yields information about frequency response functions (FRF's), which can be considered as high-dimensionality data in order to apply a DR technique to visualize them. The algorithm applied (*t*-SNE) to a training dataset corresponding to fault situation perfectly unfolds the structure of data in the high-dimensional space, giving clear insights of this sort of faults. The method accurately defines different zones of the process, distinguishing between normal operating and chatter conditions.

The developed out-of-sample approach provides the possibility to project new data points in order to supervise the existence of problems in the process. The visualization of new projected points in the map accelerates the fault detection and helps to predict this type of faults. This procedure can be time consuming depending on the size of the dataset because once computed the map for training data the new projections are computed for each point. An alternative which reduces computational burden is to use a Nadaraya-Watson regression model

[26], which estimates new projections of data based on observed values from training data.

Although it is applied to data from a cold rolling facility, this method is suitable for application to other industrial processes, whose main defects could be detected by analyzing their dynamic behavior.

Acknowledgments.

References

1. Roberts, W.L.: Cold rolling of steel. Marcel Dekker, Inc., New York (1978)
2. Yun, I.S., Wilson, W.R.D., Ehmann, K.F.: Review of chatter studies in cold rolling. International Journal of Machine Tools and Manufacture **38**(12) (1998) 1499–1530
3. Hu, P.H., Ehmann, K.F.: A dynamic model of the rolling process. part I: homogeneous model. International Journal of Machine Tools and Manufacture **40**(1) (2000) 1 – 19
4. Cuadrado, A.A., Diaz, I., Diez, A.B., Obeso, F., Gonzalez, J.A.: Visual data mining and monitoring in steel processes. In: Industry Applications Conference, 2002. 37th IAS Annual Meeting. Conference Record of the. Volume 1. (2002) 493–500 vol.1
5. Díaz, I., Domínguez, M., Cuadrado, A., Fuertes, J.: A new approach to exploratory analysis of system dynamics using som. Applications to industrial processes. Expert Systems with Applications **34**(4) (2008) 2953–2965
6. Kourti, T., MacGregor, J.: Process analysis, monitoring and diagnosis, using multivariate projection methods. Chemometrics and intelligent laboratory systems **28**(1) (1995) 3–21
7. Lee, J.A., Verleysen, M.: Nonlinear dimensionality reduction (2007)
8. Pearson, K.: LIII. On lines and planes of closest fit to systems of points in space. Philosophical Magazine Series 6 **2**(11) (1901) 559–572
9. Hyvärinen, A., Karhunen, J.: Independent component analysis (2001)
10. Kohonen, T.: Self Organizing Maps. Springer (1995)
11. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290** (2000) 2319–2323
12. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290** (2000) 2323–2326
13. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation **15**(6) (2003) 1373–1396
14. Van Der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(2579-2605) (2008) 2579–2605
15. Bushati, N., Smith, J., Briscoe, J., Watkins, C.: An intuitive graphical visualization technique for the interrogation of transcriptome data. Nucleic Acids Research **39**(17) (2011) 7380–7389

16. Jamieson, A.R., Giger, M.L., Drukker, K., Li, H., Yuan, Y., Bhooshan, N.: Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE. MEDICAL PHYSICS **37**(1) (2010) 339–351
17. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: Theory and applications. Neurocomputing **70**(1–3) (2006) 489 – 501
18. Venter, R., Abd-Rabbo, A.: Modelling of the rolling process–I: Inhomogeneous deformation model. International Journal of Mechanical Sciences **22**(2) (1980) 83 – 92
19. Freshwater, I.: Simplified theories of flat rolling–I: the calculation of roll pressure, roll force and roll torque. International Journal of Mechanical Sciences **38**(6) (1996) 633–648
20. Paton, D.L., Critchley, S.: Tandem mill vibration: Its cause and control. In: Mechanical Working; Steel Processing XXII, Proceedings of the 26th Mechanical Working; Steel Processing Conference., Chicago, IL, USA, Iron and Steel Soc Inc (1985) 247–255
21. Meehan, P.A.: Vibration instability in rolling mills: Modeling and experimental results. Journal of Vibration and Acoustics **124**(2) (2002) 221–228
22. Kimura, Y., Sodani, Y., Nishimura, N., Ikeuchi, N., Mihara, Y.: Analysis of chatter in tandem cold rolling mills. ISIJ International **43**(1) (2003) 77–84
23. Venkatasubramanian, V.: A review of process fault detection and diagnosis Part III: Process history based methods. Computers & Chemical Engineering **27**(3) (2003) 327–346
24. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**(6) (1974) 716 – 723
25. Theil, H.: Economics and information theory. North-Holland Pub. Co.; Rand McNally, Amsterdam, Chicago, (1967)
26. Simon, H.: Neural networks: a comprehensive foundation. Prentice Hall (1999)