# Estimating the Uncertainty of Brake Pad Prognostics for High-Speed Rail with a Neural Network Feature Ensemble

Alexandre Trilla[1], Pierre Dersin[2], and Xavier Cabré[3]

[1,3] *ALSTOM, Santa Perpètua de la Mogoda, Barcelona, 08130, Spain*
*alexandre.trilla@alstomgroup.com*
*francesc-xavier.cabre@alstomgroup.com*

[2] *ALSTOM, Saint Ouen, Paris, 93400, France*
*pierre.dersin@alstomgroup.com*

## ABSTRACT

The friction brake system reduces the speed of the train by transforming the kinematic energy into heat through the abrasion between the carbon pads and the disk. The British Rail Class 390 fleet (Pendolino) features a very high availability, running 1000 miles a day on average, so their wear rate is monotonic and acceptably constant. The prognostics for brake pad degradation are typically conducted with a robust online linear regression technique, which seamlessly accommodates asset-based idiosyncrasies, like the different effort that is exerted on the pad given its location on a motor or a trailer car, on the left or the right hand side of the caliper, etc. This technique is also resilient to abrupt measurement changes due to asset replacements, sensor imprecision, and acquisition failures, while retaining the physical evolution of the wear, which erodes the surface of the pad. This article evaluates the effectiveness of this approach with a dataset of brake pad thickness measurements, at the fleet level (around 12000 asset instances), using a sliding window technique, and refines its performance with a neural network ensemble, which blends physical and location features. The results of the analysis prove that this method meets the requirements of the maintenance staff and thus yields a new avenue for business improvement through the application of the predictive maintenance approach for brake pads.

## 1. INTRODUCTION

There exist many studies that review the advantages of the PHM technology for the industry (Sikorska, J. Z. and Hodkiewicz, M. and Ma, L., 2011). This work is especially concerned with the application of PHM to the maintenance of railway and rolling-stock assets (Atamuradov, V. and Medja-

Figure 1. ALSTOM TrainScanner deployment at the Manchester Traincare Centre.

her, K. and Dersin, P. and Lamoureux, B. and Zerhouni, N., 2017). In this regard, ALSTOM has developed the TrainScanner, see Figure 1, which is a train monitoring system that is aimed at optimising the maintenance of brake pads, pantograph carbon strips, and wheelsets, through the deployment of the PHM methodology and its associated techniques. TrainScanner integrates a series of acquisition subsystems with lasers and 3D cameras that capture the related measures as a train traverses its portal. Then, it automatically conducts the processing and analysis of the collected data, and finally it triggers alarms and issues reports to the maintenance staff. This work is particularly focused on the brake pad prognostics that are attainable with the carbon thickness data provided by TrainScanner over time.

Brake pad prognostics have been initially approached with finite element method simulation (AbuBakar, A. R. and Ouyang, H., 2008), highlighting the importance to consider the braking forces (Malvezzi, M. and Papini, S. and Pugi, L.

and Vettori, G. and Tesi, S. and Rindi, A. and Meli, E., 2013). Other variables have also been incorporated to better estimate the degradation, like the braking energy and the temperature (Antanaitis, D. B. and Riefe, M. T., 2016), the braking action time and the vehicle route (Kreis, C. and Dobberphul, T., 2018), or the brake pad location (Jegadeeshwaran, R. and Sugumaran, V., 2015). Other authors have focused on statistical and histogram information to create a reference wear profile and detect deviations (Chassefeyre, V., 2012) or diagnose brake faults directly (Manghai, T. M. A. and Jegadeeshwaran, R. and Sugumaran, V., 2017).

This work conducts a thorough analysis of brake pad wear at the fleet level in order to quantify the uncertainty of the prediction at 40000km into the operating life of the brake pad, which is expected to stretch up to 350000km. The time it takes the trains to run 40000km (around 20 days) is the notice requested by the maintenance team in order to schedule the depot resources effectively. This prognosis evaluation is performed with a sliding window prediction technique, using regression techniques and neural networks (Hota, H. S. and Handa, R. and Shrivas, A. K., 2007). The article is organised as follows: Section 2 describes the analysis procedure that has been followed, including the description of the data, the evaluation technique, and the prognostic methods, along with their preliminary results. Section 3 discusses the overall results and the limitations of the approach, and Section 4 concludes the manuscript and reflects on its impact to the current maintenance plan.

## 2. METHODS AND RESULTS

This section describes the sequential process that has been followed in order to obtain a robust brake pad prognostics procedure. Thus, the development is incremental and preliminary results are provided.

### 2.1. Carbon Pad Data Preprocessing

This article evaluates the effectiveness of brake pad prognostics with a dataset of brake pad thickness measurements at the fleet level, obtained with TrainScanner from November 1, 2016, to March 1, 2017. It comprises the evaluation of 11836 brake pad assets. Each set of carbon pad thickness measurements needs to be preprocessed to add robustness to the prediction. To this end, the following issues are taken into account:

1. Asset replacement: steep positive thickness increments (greater than 20mm) with a final value close to a new asset measure, i.e., 34mm, need to be segmented and treated as different assets

2. Acquisition failures: extreme values need to be regarded as invalid data and discarded from the analysis, such as values out of pad range, zeroes, etc.

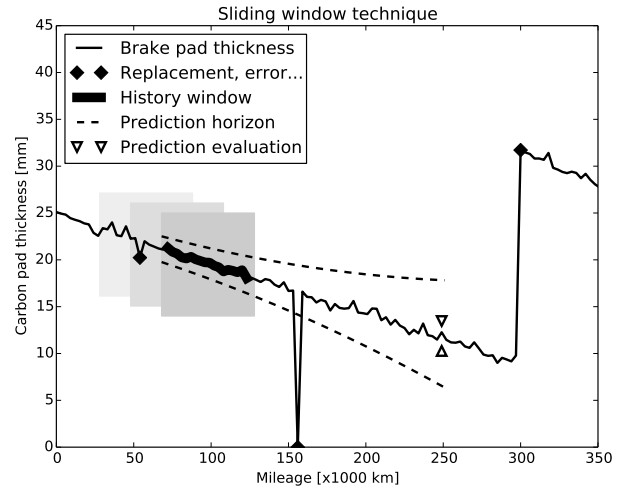3. Sensor precision: TrainScanner's rated measurement



Figure 2. Evaluation of brake pad prognostics with the sliding window prediction technique.

precision is 0.5mm. The prediction needs to be robust to this measurement variability

The resulting set of data is smooth and ready to be subject to further modelling and analysis.

The British Rail Class 390 fleet (Pendolino) is composed of 9-car trainsets, and 11-car trainsets, with 6 or 7 motor cars respectively. Each motor car has two motor axles, and two trailer axles. For a trailer car, all axles are trailer. The most common braking operation combines the electrical braking force of the motor (obviously, this is only available on motor axles), and the friction braking force of the pads, which are available on all axles, but typically they are not used on motor axles (their use is restricted to emergency braking, parking, etc). In addition, the pneumatic pressure applied to the various pads along the train is different, to compensate the contribution of these different technologies and attain a balanced dynamic behaviour for all cars, regardless of the different car weights, service load, speed, etc. The Class 390 Pendolino trains run a steady mission profile (i.e, the West Coast Main Line in the UK), which leads to expect a uniform degradation at the pad level. However, the aforementioned brake system differences also lead to expect differences at the car/axle level.

### 2.2. Sliding Window Prediction Evaluation

A rolling window is used on the continuum of clean carbon thickness measurements in order to provide a history frame that is used to make a prediction, which is then evaluated with the remaining points at a given horizon (Hota, H. S. and Handa, R. and Shrivas, A. K., 2007), see Figure 2. Similar approaches have also been derived using the uncertainty intervals that surround the trend (Greitzer & Ferryman, 2001).

It is to note that according to the ISO 13374 standard (ISO, 2003), which is our main PHM development guideline, this prediction effectiveness assessment should be conducted with the Remaining Useful Life figures (i.e., the output of the Prognosis module) instead of the brake pad thickness measurements. However, the actual replacement record is not available due to the uncertainty between the asset replacement actions (which can be done in any depot) and the asset monitoring events (which is only available at Manchester). Therefore, we reframe the objective as a sequence prediction problem.

### 2.2.1. Estimation of Uncertainty

The specific statistical terms of "accuracy" and "precision" are related with the difference between a real magnitude and a calculated value, both in terms of bias and variance error. Its bias, also known as trueness (ISO, 1994), is of little importance in this work to evaluate the effectiveness of a predictive technique, because it can be easily corrected if it is known (or experimentally estimated) in advance, which is a side objective of the evaluation techniques presented in this paper (the main use of bias is for detecting model underfitting). However, the variability of the error has a random nature, and this is the main driver of the prognostics performance: given a predictive system, its uncertainty is assumed to represent the expected maximum variability of the error, for a confidence interval of 95%, i.e., two standard deviations for a Normal distribution.

### 2.3. Weighted Online Linear Regression

In order to cope with the uniform degradation at the carbon level given by the steady mission profile of the fleet, and the different operating idiosyncrasies at the brake level (i.e., motor or trailer axle), this section reviews the online linear regression technique. The optimisation method to fit a linear model $f(\cdot)$ to the brake pad data points $x$ framed under the sliding history window $HW$ is based on the weighted squared-error cost function $C$ (i.e., a least-squares optimisation procedure), see Eq. (1). Note that the data $x_T$ represents the carbon thickness value and $x_M$ represents the train mileage, for one single brake pad asset, following the online approach.

$$C = \sum_{\forall x \in HW} w(x_T - f(x_M))^2 = \sum_{\forall x \in HW} w(x_T - (\alpha x_M + \beta))^2 \tag{1}$$

Also note that the linear model has two variables, the slope $\alpha$, which is commonly referred to as the "wear rate", and the intercept $\beta$, which biases the regression. However, the wear rate is the term that most accurately captures the dynamic behaviour of the carbon degradation (i.e., the speed
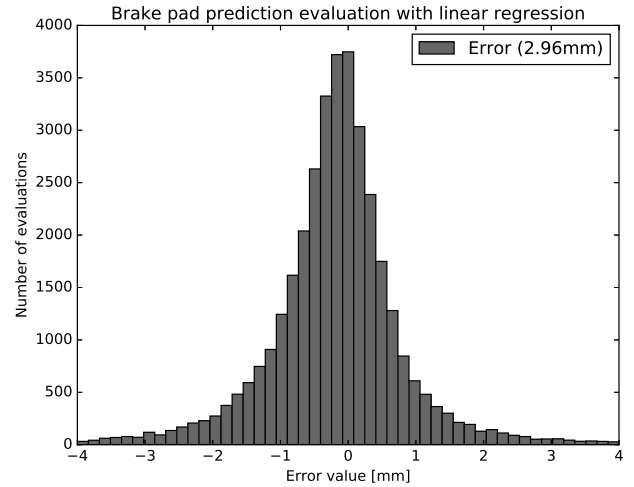


Figure 3. Histogram of the prediction error for a history window of 40000km with weighted online linear regression. In brackets, the estimated uncertainty.

of the degradation), and therefore it is used extensively hereafter. The weighting function $w(\cdot)$ is used to incorporate the robustness considerations described in Section 2.1. Finally, in order to make the prediction $P$, the linear regression function is used to extrapolate the evolution of the degradation over a given prediction horizon $PH$, and the difference with the actual measure is computed as an indicator of effectiveness, which is regarded as the prediction error $PE$, see Eq. (2).

$$PE = x_T^{PH} - P(x_M^{PH}) = x_T^{PH} - f(x_M^{PH}) \tag{2}$$

Figure 3 shows the distribution of the prediction error for a given history window of 40000km. Note that its shape is symmetric around zero, similar to the expected Gaussian, despite being somewhat more peaked. However, the fitting process shows to be correct, supported by more than 36500 instances.

At this point, the brake pad prognosis uncertainty baseline is therefore set to 2.96mm. While the prediction horizon is set by the maintenance staff to 40000km to provide for depot resources, the length of the history window is a degree of freedom that may help to attain better results. This is explored in the next section, among other refinement considerations.

### 2.4. History Window and Location Analysis

Changing the length of the history window turns the linear regression method more conservative and stable (frame increase), or more responsive and sensitive to recent measurements (frame decrease) (Greitzer & Ferryman, 2001). Table 1 shows the impact of this change on the prediction error, and the support is defined as the number of instances per asset.

It can be noted that as the amount of history points increases

Table 1. History window analysis.

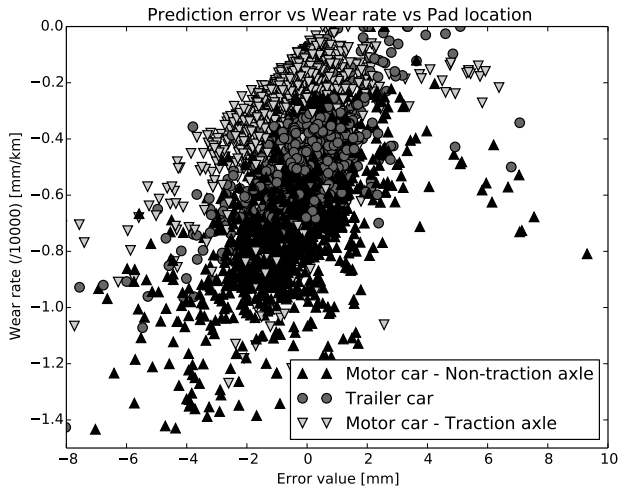| History [km] | Uncertainty [mm] | Support [/asset] |
|---|---|---|
| 20000 | 4.49 | 4.29 |
| 40000 | 2.96 | 3.09 |
| 60000 | 2.13 | 1.94 |
| 80000 | 2.03 | 1.05 |
| 100000 | 1.65 | 0.37 |



Figure 4. Prediction error correlated with wear rate, and with pad location.

(wider frame), the performance results reduce their uncertainty, but so does their support as there are less cases where the predictions can be applied. The decision as to which history length is the optimum may be somewhat arbitrary, but for 100000km, the support value well under one shows that some brake pad assets have not been evaluated, which is unacceptable to our criteria. Therefore, 80000km is regarded as the optimum history length, and the new performance score has been reduced to 2.03mm.

Despite all the former efforts to reduce the uncertainty, there is still a limit to the effectiveness of the prediction. In order to delve into the source of this issue, the brake pad location diversity may be used in order to better understand the nature of the degradation. Figure 4 correlates the pad wear rate with the prediction error, and with the pad location. The error has been centred around zero because its mean is assumed to be a location bias that can be reintroduced in the linear regression intercept term.

The main aspect to be observed is that there is a strong positive relation between the wear rate and the prediction error: the faster the pads wear, the more difficult it becomes to predict their evolution. Therefore, the pads located on motor cars and non-traction axles are generally the most critical ones. There is a small group of predictions that lie out of the con-

fidence interval of 95% (way over 4mm) that are regarded as "inconsistent" because they are shared among all the different locations, not following the group trends. Whenever the system detects that a particular situation is likely to lead to poor prognostic results, it defaults to the average wear rate for the corresponding location. The next section takes advantage of this location information to enhance the value of the prediction output to further push the performance boundary.

### 2.5. Neural Network Feature Ensemble

Neural networks are plausibly renown to be the universal learning system (Hertz, Krogh, & Palmer, 1991). They are interesting models in Artificial Intelligence and Machine Learning because they are powerful enough to succeed at solving many different problems. Historical evidence of their importance can be found as most leading technical books dedicate many pages to cover them comprehensively (Bishop, C. M., 2006; Duda, Hart, & Stork, 2000). Moreover, with the recent advent of Deep Learning, which requires very intricate networks, the neural computation paradigm is leading the state of the art (LeCun, Y. and Bengio, Y. and Hinton, G., 2015).

Neural networks exploit the connectionist learning approach, where a set of non-linear units are interconnected, and their links are weighted in order to accomplish a specific task. One of their greatest advantages is their ability to seamlessly integrate data from different sources. In this regard, Figure 5 shows a neural framework that blends the linear prediction obtained in the former section, with three flags that are indicative of the location of the brake pad asset, thus creating an ensemble of features. The stacking of the neurons into layers and their feed-forward arrangement from left to right is known as multilayer perceptron, and it is a very practical architecture for solving general-purpose problems. Note that at this level, the approach to predict the brake pad thickness is global, as there will be one single model for all assets (the online prediction is treated as an asset feature here).

The range of the input values needs to be normalised around a unitary magnitude to guarantee an effective learning convergence. The pad thickness linear prediction feature is normalised to its maximum (i.e., 34mm), and the location features are treated as binary variables. The non-linear smoothing function for all the units is set to be the logistic sigmoid. Therefore, the maximum output value of the network is normalised and scaled to the 0.79 value to avoid saturation. The network is trained with stochastic gradient descent using backpropagation, with a fast learning rate of 0.2 that is checked to avoid cost overshooting, and a maximum number of dataset iterations of 30 to consider early stopping and improve generalisation.

The challenge with this neural network is to match its expressiveness, which is related to the number of hidden units $H$, with the complexity of the data. The more weights it
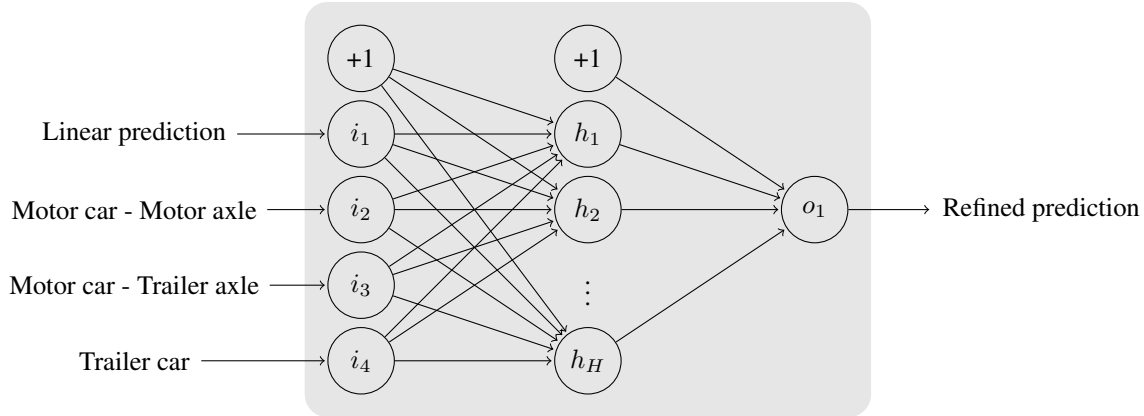
Figure 5. Neural network feature ensemble for brake pad prognostics.

has (for every hidden unit, 6 new weights are added to the network), the more data idiosyncrasies is it able to learn, at the risk of overfitting. In order to determine the optimum size of the hidden layer, a range of values are evaluated with cross-validation ("Encyclopedia of Machine Learning and Data Mining", 2010), applying 3 rounds of random sub-sampling with a train/test split of 95%/5%. This procedure yields over 1800 evaluation points, which is a sufficient sample size to reliably estimate the uncertainty. Figure 6 shows the results of this study. It can be seen that the performance score decreases gradually as the expressiveness of the network grows, especially at the beginning of the process (reduce error bias). At the end, the average score tends to reach a flat spot, but there is also a growing variability for a difference of one single hidden unit (increase error variance). We deem that the optimum size of the hidden layer is of 4 units. The error distribution for this network configuration is shown in Figure 7, where it can be seen that the final system outperforms the previous approaches and reaches a brake pad prognosis uncertainty score of 1.87mm.

## 3. DISCUSSION

This work exposes the gradual performance enhancement of a brake pad prognosis technique based on linear regression, which also intends to emulate the uniform physical degradation of carbon pads subject to a steady operational regime. The adjustment of the history window yields a prediction uncertainty of 2.03mm, and its refinement with location information through a neural network ensemble drops this figure to 1.87mm, which may add value to the current alarm threshold set at 7mm. However, the final asymmetric distribution of the error seems to indicate that this procedure cannot be pushed any further. The observed error overlap among pads with different working conditions also points toward this conclusion. Maybe the occasional use of a variable history length could be of help to reduce the higher uncertainty of pads showing a faster wear rate (Greitzer & Ferryman, 2001). Finally, given
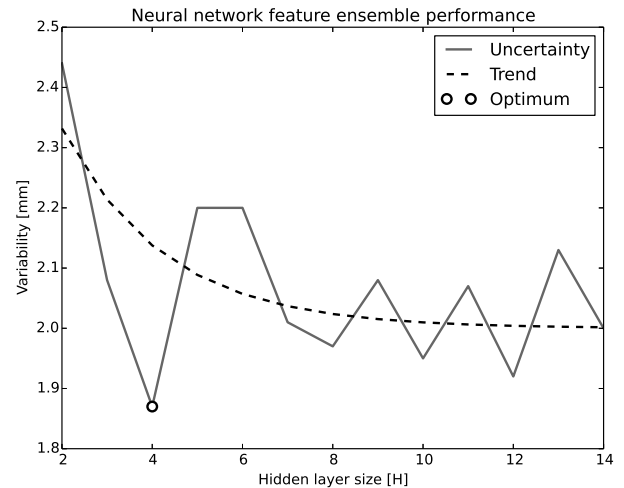


Figure 6. Neural network performance with respect to the size of its hidden layer.

that the dimensionality of the hidden layer of the neural net maintains that of the input layer, we are inclined to believe that the obtained solution is a feature space transformation that is more suitable for making better predictions, rather than a direct multivariate regression.

In addition to the techniques presented in this work, we have also informally evaluated other possible methods. We have tested non-linear regression with a higher order polynomial, but the results show no advantage, which reinforces the linear degradation behaviour. This also discards using more compact autoregressive techniques like ARMA and ARIMA. We have also tried to map the dynamic character of the pad thickness evolution into a spatial pattern, looking forward to applying a sequence recognition technique based on a Time-Delay Neural Network (Peddinti, Povey, & Khudanpur, 2015). For reference, a distance window of 40000km can be encoded in 8-10 shifts (i.e., acquisition delays) with our data. The net-
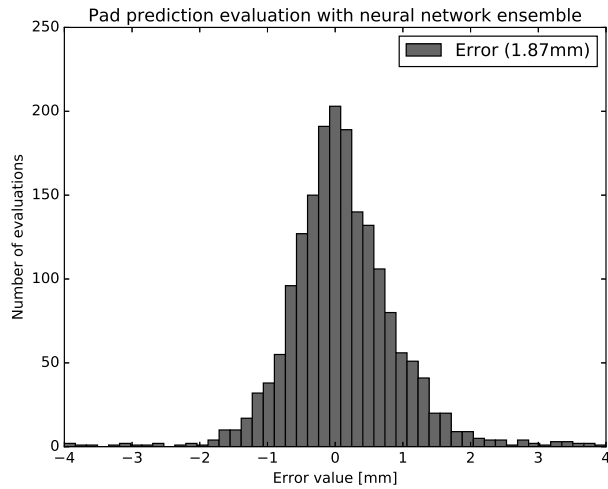
Figure 7. Histogram of the refined prediction error with a neural network feature ensemble using 4 hidden units. In brackets, the estimated uncertainty.

work has effectively converged, but not improving the proposed approach with the feature ensemble. In this line, we have also tried to approach a deep learning architecture with an additional hidden layer, but the results have been somewhat disappointing. We attribute this to the curse of dimensionality and the lack of sufficient training instances (this could be intuited based on the reported expressiveness analysis for the multilayer perceptron). Furthermore, given the high degree of overlap among the asset degradation characteristics, we consider it would we worth looking into a similarity-based prognostic technique as an alternative to the aforementioned parametric approaches. Similarly, the inclusion of other features related to the passenger weight distribution, weather conditions, and the like, may reasonably impact the degradation of the carbon pads, but so far we have not found any evident behaviour associated with the small air pressure difference at the car level. However, exploring it further is out of the scope of this article.

## 4. CONCLUSION

At present, the replacement maintenance criterion for the Class 390 brake pads is based on a single thickness threshold value. This is an evident ineffective approach because it does not take into account the rate of wear that the different pads have, and thus, the same thickness value can lead to very different operating mileages before the asset reaches its actual end of life (say, when there is no carbon left on the pad).

This article presents the most sophisticated technique for TrainScanner brake pad prognostics, which is based on a neural network ensemble that blends a robust linear regression with brake location features. It yields an uncertainty performance around 1.87mm at the asset level and for a prediction

horizon of 40000km (25000mi), which is related to the time that is necessary for planning maintenance resources at the depot. Therefore, if the expected mileage until the next visit is under this distance frame, the pad limit can be safely extended to the aforementioned thickness value.

The future work that is currently envisaged may further deal with data idiosyncrasies in order to add more robustness to the method, dealing with the data that lies out of the confidence interval. Alternatively, we also expect to explore other learning paradigms and seek the complementary characteristics that may help the current approach thrive and further push the effectiveness boundary.

## REFERENCES

AbuBakar, A. R. and Ouyang, H. (2008). Wear Prediction of Friction Material and Brake Squeal Using the Finite Element Method. *Wear*, *264*(11), 1069–1076.

Antanaitis, D. B. and Riefe, M. T. (2016). *Method of estimating brake pad wear and vehicle having a controller that implements the method* (Tech. Rep. No. US9416835B2). GM Global Technology Operations LLC.

Atamuradov, V. and Medjaher, K. and Dersin, P. and Lamoureux, B. and Zerhouni, N. (2017). Prognostics and Health Management for Maintenance Practitioners - Review, Implementation and Tools Evaluation. *International Journal of Prognostics and Health Management*, *8*(60), 1–31.

Bishop, C. M. (Ed.). (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc.

Chassefeyre, V. (2012). *System for controlling wear of brake pads of e.g. front left wheel in vehicle, has control unit including measuring unit for measuring wear of pads, and detection unit detecting evolution intervals with respect to typical wear of pads* (Tech. Rep. No. FR2969240A1). Peugeot Citroen Automobiles SA.

Duda, R. O., Hart, P. E., & Stork, D. G. (Eds.). (2000). *Pattern Classification*. Wiley-Interscience.

Encyclopedia of Machine Learning and Data Mining. (2010). In Sammut, C. and Webb, G. I. (Ed.), . Springer US.

Greitzer, F. L., & Ferryman, T. A. (2001). Predicting remaining life of mechanical systems. In *Intelligent ship symposium iv.*

Hertz, J., Krogh, A., & Palmer, R. G. (Eds.). (1991). *Introduction to the theory of neural computation*. Addison-Wesley Longman Publishing Co., Inc.

Hota, H. S. and Handa, R. and Shrivas, A. K. (2007). Time Series Data Prediction Using Sliding Window Based RBF Neural Network. *International Journal of Computational Intelligence Research*, *13*(5), 1145-1156.

ISO. (1994). *Accuracy (trueness and precision) of measure-*

*ment methods and results – Part 1: General principles and definitions* (Tech. Rep. No. 5725-1:1994). International Organization for Standardization.

ISO. (2003). *Condition monitoring and diagnostics of machine systems: Data processing, communication and presentation* (Tech. Rep. No. 13374-1:2003). International Organization for Standardization.

Jegadeeshwaran, R. and Sugumaran, V. (2015). Health Monitoring of a Hydraulic Brake System Using Nested Dichotomy Classifier – A Machine Learning Approach. *International Journal of Prognostics and Health Management*, *6*(14), 1–10.

Kreis, C. and Dobberphul, T. (2018). *A method for predictive wear detection in automotive vehicle brake systems* (Tech. Rep. No. DE102016215725A1). Volkswagen AG.

LeCun, Y. and Bengio, Y. and Hinton, G. (2015). Deep Learning. *Nature*, *521*, 436-444.

Malvezzi, M. and Papini, S. and Pugi, L. and Vettori, G. and Tesi, S. and Rindi, A. and Meli, E. (2013). Prognostic and Simulation Tools for the Preliminary Design and Verifications of Braking Performance of Railway Vehicles . *Chemical Engineering Transactions*, *33*, 751–756.

Manghai, T. M. A. and Jegadeeshwaran, R. and Sugumaran, V. (2017). Brake Fault Diagnosis Through Machine Learning Approaches – A Review. *Structural Durability and Health Monitoring*, *12*(1), 43–67.

Peddinti, V., Povey, D., & Khudanpur, S. (2015). A time delay neural network architecture for efficient modeling of long temporal contexts. In *Interspeech.*

Sikorska, J. Z. and Hodkiewicz, M. and Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, *25*, 1803-1836.

## BIOGRAPHIES

**Alexandre Trilla** graduated from LaSalle University of Barcelona with a M.Sc. in Electronics and Telecommunications Engineering in 2008, and a M.Sc. in IT Management in 2010. He has an academic research background in spoken language processing, and an industrial research background in PHM. He has authored several publications in scientific conferences and journals (IEEE Transactions on Audio, Speech, and Language Processing, Chemical Engineering Transactions, and the Journal of Rail and Rapid Transit). At present, he is a data scientist working on the deployment of PHM to the railway environment.

**Pierre Dersin** graduated from the Massachusetts Institute of Technology (MIT) with a Ph.D. in Electrical Engineering after receiving a Master's degree in Operations Research also from MIT. With ALSTOM Transport since 1990, he founded the "RAM Center of Excellence" . He is currently RAM (Reliability-Availability-Maintainability) Director and PHM (Prognostics & Health Management) Director of ALSTOM Digital Mobility. He has authored numerous publications in scientific conferences and journals in RAMS, PHM, automatic control and electric power systems (including IEEE Transactions on Automatic Control, Power Apparatus & Systems, ESREL, RAMS Symposia French Lambda-Mu conferences, the 2012 IEEE-PHM Conference) . He served on the IEEE Reliability Society AdCom from 2012 to 2017 and as VP, Technical Activities, in 2017. He is also a member of the IEEE Future Directions Committee. He is a contributor of 4 chapters in the "Handbook of RAMS in Railways: Theory & Practice" ( CRC Press, 2018).

**Xavier Cabré** was born in Barcelona. He graduated from LaSalle University of Barcelona with a M.Sc. in Electronics and Telecommunications Engineering in 1998. He has an industrial research background in high-technology environments. In 2015, he co-authored a patent in image processing and computer vision for the detection of railway components. At present he is the Project/Program manager of TrainScanner.