Highlights of Vehicles

ISSN 2696-8347, A Peer-Reviewed Open Access Journal by Highlights of Science

https://www.hos.pub/ho/vehicles

Detecting Railway Track Irregularities with Data-driven Uncertainty Quantification

by Andreas Plesner, Allan P. Engsig-Karup and Hans True

Cite this Article

Plesner, A., Engsig-Karup, A. P., & True, H. (2025). Detecting Railway Track Irregularities with Data-driven Uncertainty Quantification. *Highlights of Vehicles*, *3*(1), 1–14. https://doi.org/10.54175/hveh3010001

Highlights of Science

Publisher of Peer-Reviewed Open Access Journals

https://www.hos.pub

Barcelona, Spain

Article

Detecting Railway Track Irregularities with Data-driven Uncertainty Quantification

Andreas Plesner [1,*], **Allan P. Engsig-Karup** [2], **Hans True** [2]

[1] Computer Engineering and Networks Laboratory, Information Technology and Electrical Engineering Department, ETH Zürich, 8092 Zürich, Switzerland
[2] Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), 2800 Kgs. Lyngby, Denmark
* For correspondence: aplesner@ethz.ch

Abstract This study addresses the critical challenge of assessing railway track irregularities using advanced machine learning techniques, specifically convolutional neural networks (CNNs) and conformal prediction. Leveraging high-fidelity sensor data from high-speed trains, we propose a novel CNN model that significantly outperforms state-of-the-art results in predicting track irregularities. Our CNN architecture, optimized through extensive hyperparameter tuning, comprises multiple convolutional layers with batch normalization, Exponential Linear Unit (ELU) activation functions, and dropout regularization. This design enables the model to capture complex spatial and temporal dependencies in the train's dynamic responses, translating them into accurate predictions of track irregularities. The model achieves a mean unsigned error of 0.31 mm on the test set, surpassing the previous state-of-the-art performance and approaching industry-standard benchmarks for track measurement accuracy. This level of precision is crucial for the early detection of track defects that could compromise safety and ride quality. To quantify uncertainty in the model's predictions, we implement conformal prediction techniques, specifically the CV+ and CV-minmax methods. These approaches provide prediction intervals with high reliability, achieving a 97.18% coverage rate for the CV-minmax method. The resulting prediction intervals have an average width of 2.33 mm, offering a balance between precision and confidence in the model's outputs. Notably, our model exhibits impressive computational efficiency, capable of processing over 2000 kilometers of track data per hour. This speed makes it suitable for real-time applications in continuous monitoring systems, potentially revolutionizing the approach to railway maintenance. The integration of CNNs with conformal prediction represents a significant advancement in the field of predictive maintenance for railway infrastructure. By providing both accurate predictions and well-calibrated uncertainty estimates, our approach enables more informed decision-making in track maintenance planning and safety assessments.

Keywords railway track integrity; convolutional neural networks; conformal prediction; predictive maintenance; sensor data analysis; machine learning; uncertainty quantification

Open Access

Received: 21 September 2024
Accepted: 18 February 2025
Published: 7 March 2025

Academic Editor
Krzysztof Zboinski, Warsaw University of Technology, Poland

Copyright: © 2025 Plesner et al. This article is distributed under the terms of the **Creative Commons Attribution License** (CC BY 4.0), which permits unrestricted use and distribution provided that the original work is properly cited.

1. Introduction

In the evolving landscape of global transportation, railways continue to play a pivotal role, offering a unique blend of efficiency, reliability, and environmental sustainability. As rail networks expand and train speeds increase to meet growing passenger expectations, the imperatives of track safety and maintenance have ascended to the forefront of railway operations. The safety, comfort, and operational efficiency of rail services are directly impacted by the integrity of railway tracks, which can develop irregularities caused by wear, environmental factors, and external forces.

The challenge of maintaining railway track quality is multifaceted. Track irregularities, if left unchecked, can lead to a cascade of issues: increased wear on both track and rolling stock, reduced passenger comfort, and in extreme cases, safety risks. Traditional track inspection methodologies, while precise, grapple with significant limitations. High operational costs, limited coverage, and considerable latency between inspections hamper their effectiveness in an era demanding near-real-time monitoring and predictive maintenance.

This paper is an extended version of the work presented at ICANN 2024 [1].

The advent of advanced sensor technology presents a transformative opportunity: the potential to continuously monitor track conditions using data collected from in-service railway vehicles. By equipping regular passenger or freight trains with sophisticated sensors, we can potentially gather vast amounts of data about track conditions during normal operations. This approach promises more frequent and comprehensive track assessments without the need for dedicated inspection vehicles or service interruptions.

However, the sheer volume and complexity of the data generated by such a system pose significant challenges. How can we efficiently process this data to extract meaningful information about track conditions? How can we ensure that our assessments are accurate and reliable enough to inform critical maintenance decisions? And how can we quantify the uncertainty in our predictions to support risk-based decision-making?

Recent advances in machine learning, particularly in deep learning, offer promising solutions to these challenges. Convolutional Neural Networks (CNNs), which have demonstrated remarkable success in processing complex, high-dimensional data in fields such as computer vision, could potentially be adapted to analyze the intricate patterns in vehicle dynamics data that indicate track irregularities.

Moreover, the emerging field of conformal prediction provides a framework for quantifying the uncertainty of machine learning predictions. This is crucial in the context of railway maintenance, where understanding the confidence level of track condition assessments is vital for prioritizing maintenance activities and managing safety risks.

This study aims to explore the potential of combining CNNs and conformal prediction techniques to create a robust, efficient, and uncertainty-aware system for detecting railway track irregularities using in-service vehicle data. Our goal is to develop a methodology that can:

- Accurately detect and quantify track irregularities from vehicle dynamics data.
- Process this data efficiently enough to support real-time or near-real-time monitoring.
- Provide well-calibrated uncertainty estimates to support risk-based decision-making.

By addressing these challenges, we hope to contribute to the development of more proactive, data-driven approaches to railway maintenance. Such approaches have the potential to enhance safety, improve operational efficiency, and reduce maintenance costs across railway networks worldwide.

The remainder of this paper is structured as follows. The following subsections establish the benchmark levels and present related work. Section 2 gives an introduction to vehicle modeling and a formal setup of the problem tackled in this project. Section 3 provides a detailed overview of our methodology, including data collection, preprocessing, CNN architecture, and conformal prediction techniques. Section 4 presents our results and analysis. Section 5 discusses the implications of our findings and potential avenues for future research. Including conclusions of the paper with a summary of our key findings and their significance for the field of railway maintenance and safety.

1.1. Benchmark Levels

The EN:13848-5 standard is used to establish a benchmark [2]. This document contains operating limits for the track measurements at various speeds. The strictest limits are deviations of 1 mm for 100-meter running means and standard deviations. Based on this, a 0.1 mm benchmark will be chosen for the mean unsigned error, ME. Furthermore, to ensure that the predictions can be safely used to assess operating limits, further benchmarks are established to say that the maximum of unsigned errors is below 0.5 mm.

Since the limit values in EN:13848-5 are only specified to the nearest millimeter, a maximum unsigned error of less than 0.5 mm would imply that all running means are within 0.5 mm of the actual value [2].

Kawasaki & Youcef-Toumi [3] give error ranges that would accept maximum unsigned errors of less than 4 mm and an ME of around 1 mm, while Hao et al. [4] set a benchmark of 0.25 mm and 0.45 mm of the mean unsigned error in the wavebands [3 m, 42 m] and [42 m, 120 m], respectively. These are less strict than our benchmarks mentioned above. We, therefore, set additional benchmarks of an ME of 0.35 mm (the mean of 0.25 mm and 0.45 mm) and a maximum unsigned error of 4 mm and call these the "satisfying" levels.

1.2. Related Work

This subsection will go over related work to this project. The selected papers are grouped into three categories. Firstly, Vehicle dynamics will look at papers focused on the motivation of continuous track measurements. Secondly, Classic methods will highlight papers that have attacked the problem using traditional signal processing methods. Thirdly, deep learning methods show some papers that have used deep learning methods, as in this paper, to solve the problem.

1.2.1. Vehicle Dynamics

Traditionally, the integrity assessment of railway infrastructure has relied heavily on periodic inspections using specialized measurement vehicles, a process that, while accurate, suffers from limitations such as high costs, limited coverage, and the potential for subjective error. Ravitharan [5] highlights the operational benefits of proactive maintenance strategies, advocating for continuously monitoring track conditions using in-service railway vehicles, a concept explored over the past two decades [3,6]. Lee et al. [7] underscore the direct correlation between vehicle dynamics and track conditions, laying the foundation for the use of vehicle dynamics as a means of assessing track quality.

1.2.2. Classic Methods

Prior works have mainly focused on classical mathematical analysis tools, such as Kalman filters, system identification techniques, digital and analog processing, and other signal processing methods [3,7–14]. These works have prioritized interpretable models over complex data-driven solutions which might have higher accuracy. These traditional methods often encounter mathematical difficulties, such as the issue of double integration of accelerations to obtain positions, which complicates their application in real-world scenarios [6].

1.2.3. Deep Learning Methods

Recent advances in machine learning, particularly in the application of convolutional neural networks, present promising alternatives to traditional methods.

Data-driven machine learning models have begun to shift the paradigm in various domains, showing superior performance in fields such as image analysis [15,16]. In the context of monitoring the condition of the railway track, initiatives have explored the use of cameras on board and binary classification techniques to differentiate between good and bad track conditions [17–20]. However, the adoption of machine learning in this domain is not without its challenges. Despite their promise, these approaches face their own set of limitations, including computational demands and the lack of severity assessment in track irregularities [17].

Research by Hao et al. [4] presents a notable advancement, which highlights the potential of deep learning approaches to predict vertical track irregularities with a high degree of precision. However, this method does not address lateral irregularities. Similarly, the use of autoencoders to compress irregularity data presents innovative solutions, but the evaluation is limited to simulated environments and again to specific types of irregularities [21].

Exploring data-driven methods for road quality monitoring has also yielded encouraging results, suggesting that similar approaches could be beneficial for the maintenance of railway tracks [22].

2. Vehicle Dynamics and Problem Formulation

The following subsections will outline the vehicle description and formally set up the problem to be solved. We need to first understand the vehicle dynamics to see that we can formulate the problem as an inverse problem.

2.1. Vehicle Dynamics

The behavior of a railway vehicle on a track is governed by interactions between various components, with the influences of the track geometry being the main outside factor.

Railway Vehicle Model: We consider a typical four-axle railway vehicle, consisting of:

1. Car body.
2. Two bogies (front and rear).
3. Four wheelsets (two per bogie).

These components are interconnected through primary and secondary suspension systems, which play a critical role in determining the vehicle's response to track irregularities. See Figure 1 for a depiction of the suspension system in a Cooperrider's boogie.

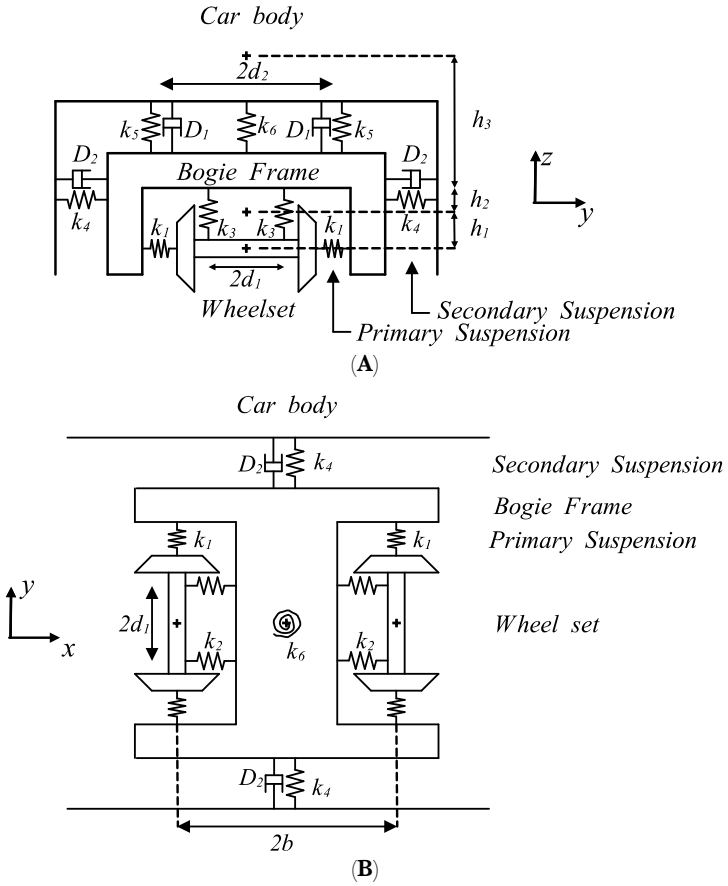

Figure 1. Front-view (**A**) and top-view (**B**) of the primary and secondary suspension system in a Cooperrider's boogie model. Shown to exemplify a suspension system, and how the components can be connected. Figures adapted from [23].

The vehicle's dynamic response can be described by six degrees of freedom for each rigid body:

1. Longitudinal translation (x).
2. Lateral translation (y).
3. Vertical translation (z).
4. Roll rotation (ϕ).
5. Pitch rotation (χ).
6. Yaw rotation (ψ).

In the data, we will not measure the angles directly, but rather opt for, e.g., having two sensors on a wheelset (one at each end), which measure the longitudinal, lateral, and vertical accelerations. From these, we could compute the accelerations for the angles, but there should be no benefit from doing so from the machine learning perspective.

Track irregularities manifest in four primary forms:

1. Vertical profile (longitudinal level).
2. Lateral alignment.
3. Cross level (superelevation).
4. Gauge variation.

These irregularities excite various modes of vehicle motion, with complex interactions between different wavelengths and the vehicle's natural frequencies. And the vehicle's response to track irregularities can be conceptualized as a series of forced vibrations. The frequency and amplitude of these vibrations depend on:

1. The wavelength and amplitude of track irregularities.
2. Vehicle speed.
3. Vehicle suspension characteristics.
4. Mass and inertia properties of vehicle components.

Several non-linear effects complicate the relationship between track irregularities and vehicle dynamics with the primary one being the wheel-rail contact mechanics, including flange contact. The non-linearities are more pronounced in the lateral direction, which makes lateral irregularity prediction more challenging than vertical irregularity prediction [1,24,25].

One could use the physical understanding of the system to enhance the predictive power of the machine learning models [26–30]. However, we leave this for future work, as the physical information needs to be formulated for each vehicle type.

2.2. Problem Formulation

The challenge of predicting railway track irregularities from vehicle dynamics can be formulated as an inverse problem. Given a set of dynamic responses from sensors installed on an in-service railway vehicle, we aim to infer the underlying track geometry irregularities that caused these responses.

Mathematically, we can express this as:

$$y = H(x) + \epsilon, \tag{1}$$

where:

- x represents the track irregularities,
- y represents the measured vehicle dynamics,
- H is the complex, non-linear transfer function that maps track irregularities to vehicle dynamics, and
- ϵ represents measurement noise and model uncertainties.

Our goal is to approximate H^{-1}, the inverse of H, using a convolutional neural network. This inverse function should map the observed dynamics y back to the track irregularities x with high accuracy.

The problem is complicated by several factors:

1. Non-linearity: The relationship between track irregularities and vehicle dynamics is highly non-linear, especially for lateral dynamics.
2. Multi-dimensionality: Both input (dynamics) and output (irregularities) are multi-dimensional time series data.
3. Different frequency domains: Track irregularities are categorized into three wavelength bands (D1, D2, D3), each potentially requiring different modeling approaches.
4. Uncertainty quantification: Beyond point predictions, we need to quantify the uncertainty in our estimates to support risk-based decision-making.

3. Materials and Methods

This section outlines the methodology employed to predict railway track irregularities using CNNs complemented by conformal prediction techniques to estimate the uncertainties of these predictions. The section also includes a bit of background for these methods, how the CNN is designed specifically for the task at hand and training the model. In addition, we will look at the data used for this project.

3.1. Data Collection and Preprocessing

Our dataset comprises high-fidelity sensor readings from an ETR500 high-speed train, capturing various dynamic responses under operating conditions. The data includes:

1. Input features: 38 time series of acceleration measurements from various locations on the train (axle boxes, bogies, car body) in lateral and vertical directions, and vehicle velocity and track curvature. Cf. Table A1.
2. Output features: 12 time series of track irregularities (lateral and vertical for left and right rails, in three wavelength domains D1, D2, D3). Cf. Table 1.
3. Auxiliary data: Vehicle speed and track curvature.

The data was collected over approximately 140 km of track, sampled at ≈1000 Hz, resulting in 287,827 observations. The ETR500 high-speed train only measures the vehicle dynamics, and another inspection vehicle collected the track geometry readings.

Table 1. Features in a sample of geometry dataset—With labels used in this project.

Label	Unit	Description	Notes
Position	km	Position along the track	
Lateral left D1	mm	Lateral irregularities of the left and right rail in the D1 wavelength domain	D1 is the first frequency band with wavelengths in [3 m, 25 m]
Lateral right D1	mm		
Vertical left D1	mm	Vertical irregularities of the left and right rails in the D1 wavelength domain	
Vertical right D1	mm		
Lateral left D2	mm	Lateral irregularities of the left and right rail in the D2 wavelength domain	D2 is the second frequency band with wavelengths in [25 m, 70 m]
Lateral right D2	mm		
Vertical left D2	mm	Vertical irregularities of the left and right rails in the D2 wavelength domain	
Vertical right D2	mm		
Lateral left D3	mm	Lateral irregularities of the left and right rail in the D3 wavelength domain	D3 is the third frequency band with wavelengths in [70 m, 200 m]
Lateral right D3	mm		
Vertical left D3	mm	Vertical irregularities of the left and right rails in the D3 wavelength domain	
Vertical right D3	mm		

The preprocessing steps involved the removal of outliers and normalization and segmentation to ensure compatibility with the CNN architecture. This preprocessing facilitated the transformation of raw sensor data into a structured format conducive to machine learning models.

The data were collected using multiple accelerometers located at various points on the railway vehicle; see Figure 2 for locations. The input data consists of a time series with measurements from each accelerometer. The output data consists of the irregularities of the track in the lateral and vertical directions for the left and right rails. These have further been split into three frequency domains, D1, D2, and D3, with wavelengths of [3 m, 25 m], [25 m, 70 m], and [70 m, 200 m], respectively, thus giving 12 output series. The features in the track geometry (output) data can be seen in Table 1.

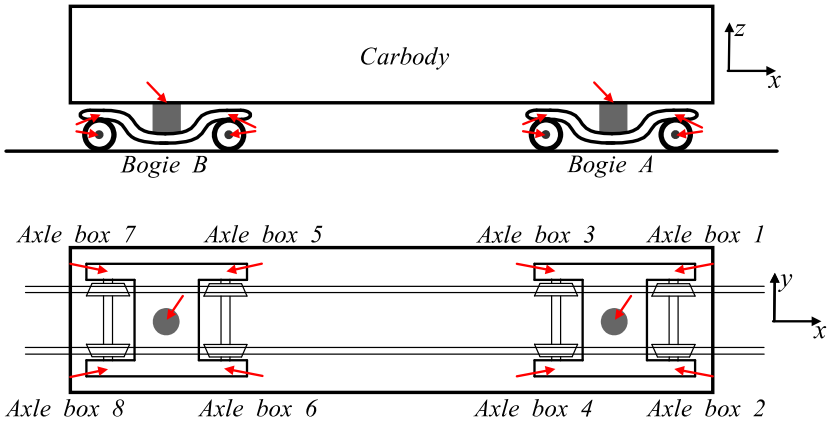

Figure 2. Data measurement locations of the vehicle dynamics. The red arrows indicate the placement of the accelerometers on the axle boxes, bogies, and car body.

An outlier analysis found that five of the accelerometers had regions where they were faulty. We zeroed the data from the faulty sensor data in these regions but kept the data from the non-faulty sensors as they were. The outliers were found by plotting the data and noticing that some sensors had segments with values that were constant and many times larger than all other values collected by the same sensor.

The data were collected with a sampling frequency of ≈1000 Hz, but as the railway vehicle moved at a non-constant speed, this gave a non-uniform sampling in space. We, therefore, interpolated to have a constant sample spacing of 0.167 m.

We then split the data into training and testing regions. The data have 287,827 observations, and the 1st to 23,827th, the 94,001st to 117,827th, and the 188,001st to 211,827th samples are used as the test data. The training data consists of the remaining data.

These regions are shown in Figure 3. The training data are further divided into training and validation segments by splitting it into nine regions and using one for validation and the remaining eight for training. Six of the nine regions are used for validation; a separate model is trained for each of the six validation regions, and the results are the mean across the six models.

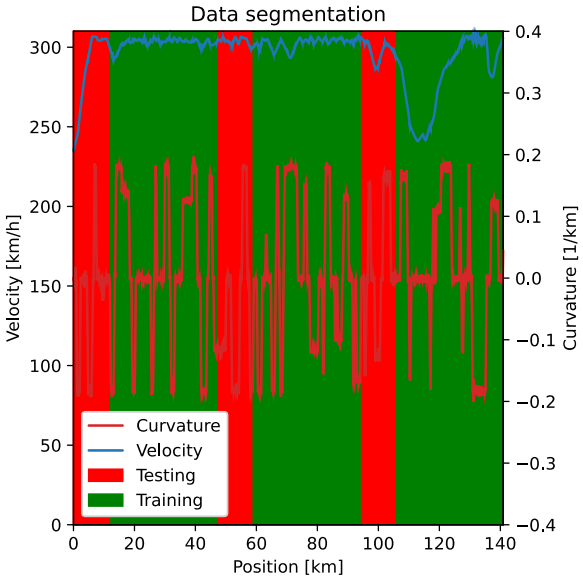

Figure 3. Training and testing regions of the data. During the training of the models, the training data is divided into training and validation segments by splitting it into nine regions and using one for validation and the remaining eight for training.

To summarize, our data processing is as follows:

1. Data measurements: Measure vehicle components' accelerations and rails' irregularities.
2. Outlier Detection and Handling: Identify regions with faulty sensor data in five input channels. These faulty readings were zeroed out rather than removed to maintain the time series structure.
3. Interpolation: Ensure consistent spatial sampling by interpolating the data to a constant spacing of 0.167 m along the track. This is done using cubic spline interpolation. This step was necessary due to variations in train speed during data collection.
4. Normalization: Input features were standardized to zero mean and unit variance to facilitate model training.
5. Segmentation: The data was split into training (70%), validation (10%), and test (20%) sets. The split was done spatially to maintain the temporal structure of the data:
   (a) Test set: the 1–23,827, 94,001–117,827, and the 188,001–211,827 samples.
   (b) Training and validation: Remaining data, which were further divided into nine segments for cross-validation.
6. Windowing: For the CNN input, we use each of the nine segments as a sample with a batch size of one for each input channel. The output is cropped to remove the first and last samples the models do not make predictions for.

This preprocessing pipeline ensures that our data is clean, consistently formatted, and structured appropriately for our CNN model while preserving the spatial and temporal relationships crucial for accurate irregularity prediction.

3.2. Convolutional Neural Network Architecture

The CNN architecture was designed to process time-series data, capturing spatial and

temporal dependencies inherent in the train's dynamic responses. The model comprises multiple convolutional layers, each followed by pooling layers to reduce dimensionality and enhance feature extraction. Dropout layers were incorporated to mitigate overfitting, ensuring the model's generalizability across different track conditions. The model consists of batch normalization of the input and then three hidden CNN layers using batch normalization, the Exponential Linear Unit (ELU) activation function, and dropout of 60%, with a final CNN layer to obtain the output [31–33]. The first convolutional layer uses very large kernels to ensure that features with 300-m wavelengths can be captured. This is relevant as the irregularities can exhibit wavelengths up to 200 m. A diagram of the final network has been shown in Figure 4.

Hyperparameters, including the learning rate, number of convolutional layers, kernel size, and dropout rate, were tuned using a combination of grid search and cross-validation to find the optimal model configuration by comparing the validation losses. We employed an iterative approach for the grid search, tuning only one or a few variables simultaneously.

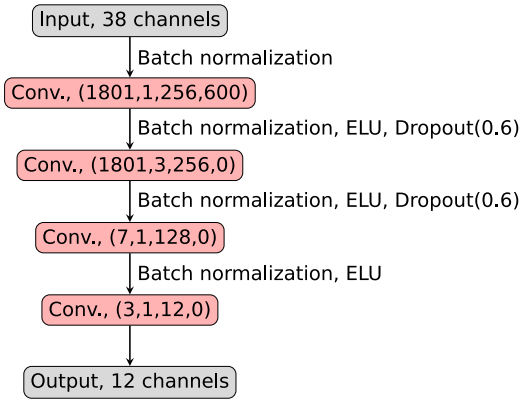

Figure 4. Depiction of the CNN after tuning the hyperparameters. The parameters for convolution layers (Conv.) are given by a tuple with (kernel size, stride, # of channels, and padding).

3.3. Conformal Prediction Framework

To quantify the uncertainty of CNN predictions, we applied conformal prediction methods. These methods use residuals from the training dataset to construct prediction intervals for new observations. Different variants of conformal prediction, Naïve, Holdout, and Cross-Validation, were evaluated to determine the most effective approach for this application. The Cross-Validation variant can further be split into three versions, CV, CV+, and CV-minmax [34–37].

The best intervals were produced by CV+ and CV-minmax, so the results will include only these. These methods have assumption-free theoretical guarantees that the α level interval contains $> 1 - \alpha$ of the samples [35]. For our results, we will use $\alpha = 0.05$ intervals. The focus on CV+ and CV-minmax is due to the better theoretical guarantees of these methods [35].

3.4. Evaluation Metrics

The performance of the CNN model and the effectiveness of the conformal prediction intervals were evaluated using a few metrics. For CNN, the metrics were the mean and maximum of unsigned errors and the compute time. For conformal prediction, the focus was on the accuracy of the prediction intervals measured through the coverage probability (how often the true values were inside the interval) and the width of the interval assessed through the mean and maximum width.

4. Results

This section will present the results of this project for the best CNN model constructed, the use of conformal predictions, and, lastly, the compute time required to evaluate the model and produce prediction intervals.

4.1. CNN Predictions

The CNN model displayed proficiency in predicting track irregularities from dynamic responses of in-service railway vehicles. The model, after rigorous tuning, achieved a satisfactory

ME by beating the "satisfying" benchmark for the ME. This significant achievement is depicted in Figure 5, illustrating the training and validation mean errors across epochs, where the model's performance is notably highlighted by its capacity to maintain errors below the "satisfying" benchmark level.

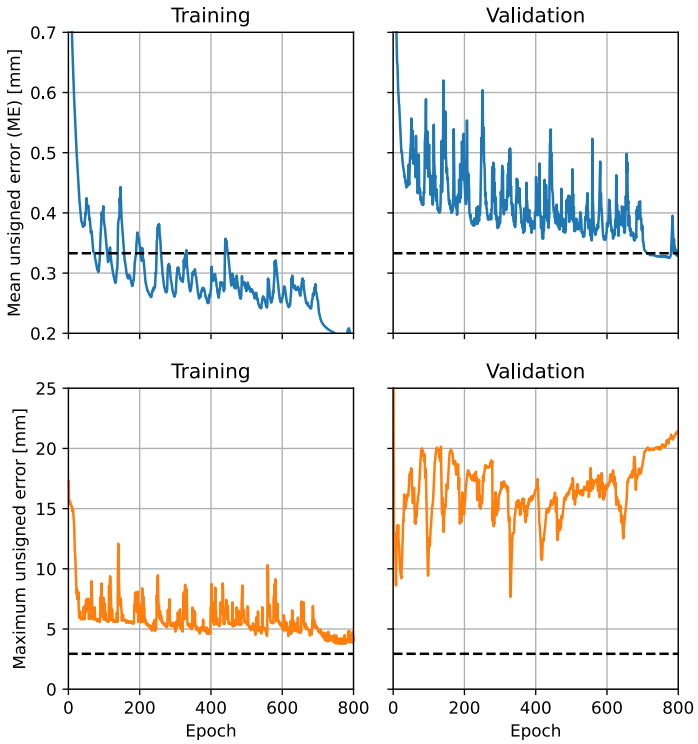

Figure 5. The training and validation mean and maximum unsigned error during training for the best-performing CNN model. The black dashed lines are the "satisfying" benchmark levels. We can see that the model gets a satisfactory mean unsigned error, but the maximum is still off.

Architecture and hyperparameter optimization played a pivotal role in enhancing the model's accuracy. The final CNN model utilized a sophisticated arrangement of convolutional layers coupled with dropout regularization and batch normalization techniques. These elements collectively contributed to a robust model capable of discerning the intricate patterns associated with track irregularities from the vast and complex data derived from railway dynamics.

An extensive error analysis was conducted to dive into the predictive capabilities of the model and areas of improvement. This analysis was crucial in understanding the nuances of the model's performance, including the instances where it deviated from expected outcomes. Despite achieving high accuracy, the model faced challenges with maximum errors, especially in the validation data, prompting a detailed examination of error characteristics to identify potential model enhancements. The result of this analysis showed that the model makes the largest errors in regions with faulty sensor data. This is highlighted in Table 2, which shows key statistics for the model evaluated on the test data. The test data did not contain faulty sensor data. From the table, we see that the model also gets a satisfactory mean, but not a satisfactory maximum, on the test data. However, the aggregated maximum unsigned error is much smaller in the test data compared to the validation data errors seen in Figure 5.

Additionally, the model beats the state-of-the-art results from [4] for short wavelengths and the aggregated mean unsigned errors. However, the model falls short of the 0.1- and 0.5-mm benchmarks for, respectively, the mean and maximum of unsigned errors.

Thus, to further improve the model, the focus should be on the data used to train the model. This might then eliminate the issues with missing sensor data.

4.2. Conformal Prediction

Integration of conformal prediction methods notably enhanced the predictive capabilities of the CNN model. The CV+ and CV-minmax methods were used to calculate the prediction intervals for the test data, which would improve the confidence in the model outputs for new predictions.

Table 2. Mean and maximum of the unsigned test errors for each of the 12 output features. Values highlighted in red and in bold are those that exceed satisfactory levels. Recall that D1, D2, and D3 correspond to wavelengths of [3 m, 25 m], [25 m, 70 m], and [70 m, 200 m], respectively. From this, the mean unsigned errors for the three wavelength regions are 0.205 mm, 0.248 mm, and 0.486 mm, respectively.

	Mean [mm]	**Maximum [mm]**
Lateral left D1	0.14	2.54
Lateral right D1	0.13	2.62
Vertical left D1	0.27	2.23
Vertical right D1	0.28	2.56
Lateral left D2	0.2	3.55
Lateral right D2	0.18	3.49
Vertical left D2	0.3	2.28
Vertical right D2	0.31	2.63
Lateral left D3	0.35	**7.33**
Lateral right D3	0.33	**7.61**
Vertical left D3	**0.63**	**6.58**
Vertical right D3	**0.64**	**6.19**
Aggregates	0.31	**7.61**

The predictions are made as a mean aggregate of the six model instances trained for each validation segment. As depicted in Figure 6, the CV+ and CV-minmax methods achieved high true value coverage rates, illustrating their effectiveness in encompassing data variability. The CV+ intervals are slightly narrower than those for CV-minmax but are overall very similar.

Table 3 presents the aggregate statistics for the $\alpha = 0.05$ intervals. It reveals that although CV-minmax offers higher coverage at 97.18% compared to CV+'s 95.76%, it produces wider intervals on average (2.33 mm for CV-minmax versus 1.78 mm for CV+). Notably, as will be shown later, the CV-minmax method's prediction intervals are computed faster than those of CV+, an advantage for real-time applications.

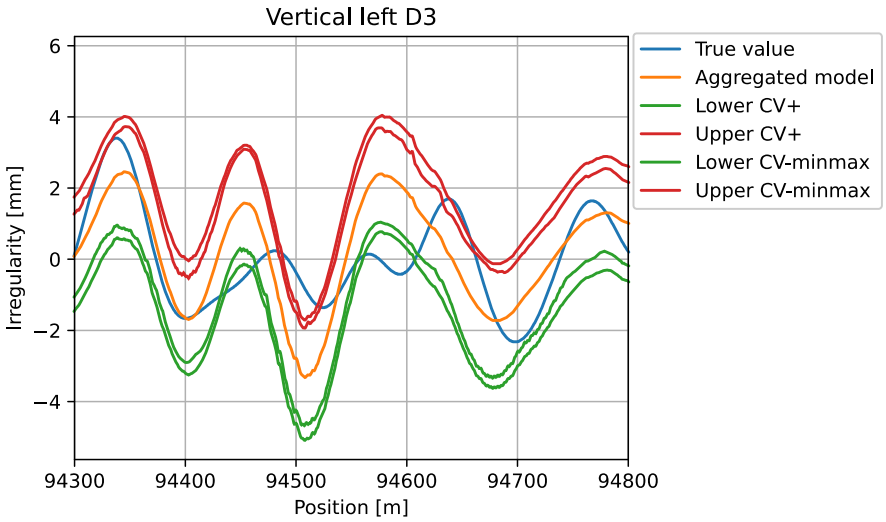

Figure 6. Comparison of prediction intervals from CV+ and CV-minmax methods against the true values of track irregularities for the vertical D3 (long wavelengths) irregularities of the left rail. We see that the intervals often capture the true value, but there are instances, where they fail.

Table 3. Aggregate statistics for conformal prediction intervals using the CV+ and CV-minmax methods. CV+ produces narrower intervals, but they have a slightly lower coverage.

Measure	**CV+**	**CV-minmax**
True value coverage (%)	95.76	97.18
Average interval width (mm)	1.78	2.33
Maximum interval width (mm)	4.54	5.25

4.3. Compute Time

A crucial aspect of the CNN model's design was its ability to process and make predictions at a rate that exceeds the operational speeds of high-speed railway vehicles, which can reach speeds over 300 km per hour.

The CNN model could process 35.30 km of test data in 60.72 seconds using a GTX 970. Thus, the model can process track data at a rate of 2093 km per hour. Meanwhile, for conformal predictions, the CV-minmax and CV+ methods require 0.12 seconds and 12 minutes, respectively, to process the test data. This means they can process track data at rates of over 1,000,000 km and 176.5 km per hour, respectively.

Thus, the CV+ method cannot process sufficiently fast. However, the CNN model and CV-minmax demonstrated exceptional efficiency, capable of evaluating substantial lengths of track data within a constrained timeframe, thereby ensuring its applicability in real-time monitoring systems.

We propose a CNN-based model with uncertainty quantified via conformal predictions as a solution for continuous real-time monitoring of railway track conditions.

5. Conclusion

This research ventured into the domain of using data-driven machine learning methods, with a focus on convolutional neural networks and conformal prediction, to predict railway track irregularities from the observed dynamics of in-service railway vehicles. The core achievement was the development of a predictive model that not only delivered satisfactory accuracy in detecting track irregularities but also incorporated conformal prediction to estimate the uncertainty of these predictions reliably. Satisfactory results are set as a mean unsigned error of 0.35 mm based on state-of-the-art results from related work.

Our model has a mean unsigned error of 0.31 mm on the test set, thus improving the state-of-the-art results of [4].

Interestingly, the conformal prediction methodology achieved a high coverage of 97.18% of the true values, with prediction intervals of an average width of 2.33 mm, thus ensuring a robust and reliable predictive framework.

However, it was noted that, while the prediction coverage was impressively high, the width of the intervals, though relatively small, indicates room for optimization to refine the precision further. These intervals were derived using the CV-minmax method, highlighting the potential for real-time application of this approach, given its ability to evaluate more than 1M km of track data per hour. Additionally, the CNN model could process track data at a rate of more than 2000 km per hour. This efficiency underscores the feasibility of deploying this methodology in real-world settings, where it can serve as a cornerstone for continuous real-time monitoring of railway track conditions using in-service high-speed vehicles.

The journey to improve the accuracy and reliability of track irregularity detection through machine learning is far from over. Future endeavors can pivot around several key areas to push the boundaries of current achievements. Primarily, addressing the identified data issues will be crucial. This includes refining sensor data quality by removing or correcting data from faulty sensors and handling outliers more effectively. The model could, for instance, be made more robust so that it can allow for faulty sensors.

Further exploration of vehicle modeling offers a promising avenue for advancement. Transitioning the codebase to Julia has opened up new possibilities for using scientific computing methods. For example, delving into the domain of scientific machine learning, specifically through the lens of Neural Ordinary Differential Equations (NODEs), presents an exciting frontier. This approach could fundamentally change the way we model vehicle dynamics by integrating data-driven insights directly into the differential equations governing these dynamics. Future research could analyze how the kernels in the model learn to predict the vertical vs. the lateral irregularities.

We tried using transfer learning to simulate vehicle dynamics and pre-training the model on these simulated data. However, this did not produce the expected benefits in this study, suggesting a potential misalignment in data formatting or a lack of representation in the ODE system. Future research could aim to refine these aspects, potentially leading to breakthroughs in model performance and generalizability. Similarly, future work could try using physics-informed neural networks (PINNs) to enhance the model by incorporating physical laws directly into the learning process.

In sum, the groundwork laid by this project not only contributes to the current body of knowledge but also charted a course for future research to explore uncharted territories in railway track maintenance and safety through the lens of advanced machine learning techniques.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The data used in this project is not publicly available. The data are provided by a research group that wishes to stay anonymous, and the authors were not given permission to share the data nor the data source.

Author Contributions

Hans True and Allan Engsig-Karup were responsible for project administration, supervision, and review and editing. Hans True was further responsible for data curation. Andreas Plesner was responsible for all remaining tasks.

Conflicts of Interest

The authors have no conflict of interest to declare.

References

1. Plesner, A., Engsig-Karup, A. P., & True, H. (2024). Detecting railway track irregularities using conformal prediction. In *Proceedings of the 33rd International Conference on Artificial Neural Networks*. Springer, Cham.
2. European Committee for Standardization. (2017). *En CEN 13848-5, railway applications–track–track geometry quality–part 5: geometric quality levels–plain line, switches and crossings*.
3. Kawasaki, J., & Youcef-Toumi, K. (2002). Estimation of rail irregularities. *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, 5, 3650–3660. https://doi.org/10.1109/acc.2002.1024495
4. Hao, X., Yang, J., Yang, F., Sun, X., Hou, Y., & Wang, J. (2023). Track geometry estimation from vehicle–body acceleration for high-speed railway using deep learning technique. *Vehicle System Dynamics*, *61*(1), 239–259. https://doi.org/10.1080/00423114.2022.2037669
5. Ravitharan, R. (2019). Safer rail operations: Reactive to proactive maintenance using state-of-the-art automated in-service vehicle-track condition monitoring. In *Proceedings of the 2018 International Conference on Intelligent Rail Transportation*. IEEE. https://doi.org/10.1109/ICIRT.2018.8641587
6. Weston, P., Roberts, C., Yeo, G., & Stewart, E. (2015). Perspectives on railway track geometry condition monitoring from in-service railway vehicles. *Vehicle System Dynamics*, *53*(7), 1063–1091. https://doi.org/10.1080/00423114.2015.1034730
7. Lee, J. S., Choi, S., Kim, S. S., Kim, Y. G., Kim, S. W., & Park, C. (2012). Waveband analysis of track irregularities in high-speed railway from on-board acceleration measurement. *Journal of Solid Mechanics and Materials Engineering*, *6*(6), 750–759. https://doi.org/10.1299/jmWWmp.6.750
8. Balouchi, F., Bevan, A., & Formston, R. (2021). Development of railway track condition monitoring from multi-train in-service vehicles. *Vehicle System Dynamics*, *59*(9), 1397–1417. https://doi.org/10.1080/00423114.2020.1755045
9. Chudzikiewicz, A., Bogacz, R., Kostrzewski, M., & Konowrocki, R. (2018). Condition monitoring of railway track systems by using acceleration signals on wheelset axle-boxes. *Transport*, *33*(2), 555–566. https://doi.org/10.3846/16484142.2017.1342101
10. Muñoz, S., Ros, J., Urda, P., & Escalona, J. L. (2021). Estimation of lateral track irregularity through kalman filtering techniques. *IEEE Access*, *9*, 60010–60025. https://doi.org/10.1109/ACCESS.2021.3073606
11. Muñoz, S., Ros, J., Urda, P., & Escalona, J. L. (2021). Estimation of lateral track irregularity using a kalman filter. experimental validation. *Journal of Sound and Vibration*, *504*, 116122. https://doi.org/10.1016/j.jsv.2021.116122
12. Naganuma, Y., Kobayashi, M., & Okumura, T. (2010). Inertial measurement processing techniques for track condition monitoring on shinkansen commercial trains. *Journal of Mechanical Systems for Transportation and Logistics*, *3*(1), 315–325. https://doi.org/10.1299/jmtl.3.315
13. Tsunashima, H., & Hirose, R. (2022). Condition monitoring of railway track from car-body vibration using time–frequency analysis. *Vehicle System Dynamics*, *60*(4),1170–1187. https://doi.org/10.1080/00423114.2020.1850808
14. Wei, X., Liu, F., & Jia, L. (2016). Urban rail track condition monitoring based on in-service vehicle acceleration measurements. *Measurement*, *80*, 217–228. https://doi.org/10.1016/j.measurement.2015.11.033

15. Aslam, Y., & Santhi, N. (2019). A review of deep learning approaches for image analysis. In *Proceedings of the 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 709–714). IEEE. https://doi.org/10.1109/ICSSIT46314.2019.8987922
16. O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L, et al. (2020). Deep learning vs. traditional computer vision. In *Advances in Computer Vision, Proceedings of the 2019 Computer Vision Conference (CVC)* (pp. 128–144). Springer, Cham. https://doi.org/10.1007/978-3-030-17795-9_10
17. Mittal, S., & Rao, D. (2017). *Vision based railway track monitoring using deep learning*. arXiv. https://doi.org/10.48550/arXiv.1711.06423
18. De Rosa, A., Kulkarni, R., Qazizadeh, A., Berg, M., Gialleonardo, E. D., Facchinetti, A., et al. (2021). Monitoring of lateral and cross level track geometry irregularities through onboard vehicle dynamics measurements using machine learning classification algorithms. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *235*(1), 107–120. https://doi.org/10.1177/0954409720906649
19. Tsunashima, H. (2019). Condition Monitoring of Railway Tracks from Car-Body Vibration Using a Machine Learning Technique. *Applied Sciences*, *9*(13), 2734. https://doi.org/10.3390/app9132734
20. Yang, C., Sun, Y., Ladubec, C., & Liu, Y. (2021). Developing Machine Learning-Based Models for Railway Inspection. *Applied Sciences*, *11*(1), 13. https://doi.org/10.3390/app11010013
21. Li, C., He, Q., & Wang, P. (2021). Estimation of railway track longitudinal irregularity using vehicle response with information compression and Bayesian deep learning. *Computer-aided Civil and Infrastructure Engineering*, *37*(10), 1260–1276. https://doi.org/10.1111/mice.12802
22. Varona, B., Monteserin, A., & Teyseyre, A. (2020). A deep learning approach to automatic road surface monitoring and pothole detection. *Personal and Ubiquitous Computing*, *24*, 519–534. https://doi.org/10.1007/s00779-019-01234-z
23. Plesner, A. (2022). *Using data-driven state-of-the-art machine learning and conformal prediction for track irregularities from observed dynamics of inservice railway vehicles* [Master's Thesis]. Technical University of Denmark.
24. True, H., Christiansen, L. E., Plesner, A. L., Ammitzbøll, A. L., & Dahl, B. J. (2020). Why is it so difficult to determine the lateral position of the rails by a measurement of the motion of an axle on a moving vehicle? In *Second International Conference on Rail Transportation*. Southwest Jiaotong University.
25. True, H., Christiansen, L. E., Plesner, A. L., Ammitzbøll, A. L., & Dahl, B. J. (2023). On the problem of the dynamical reactions of a rolling wheelset to real track irregularities. *Railway Engineering Science*, *31*(1), 1–19. https://doi.org/10.1007/s40534-022-00288-9
26. Cai, S., Mao, Z., Wang, Z., Yin, M., & Karniadakis, G. E. (2021). Physics-informed neural networks (PINNs) for fluid mechanics: A review. *Acta Mechanica Sinica*, *37*(12), 1727–1738. https://doi.org/10.1007/s10409-021-01148-1
27. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5
28. Markidis, S. (2021). *Physics-informed deep-learning for scientific computing*. arXiv. https://doi.org/10.48550/arXiv.2103.09655
29. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). *Physics informed deep learning (Part I): Data-driven solutions of nonlinear partial differential equations*. arXiv. https://doi.org/10.48550/ARXIV.1711.10561
30. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045
31. Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2020). *Fast and accurate deep network learning by exponential linear units (ELUs)*. arXiv. https://doi.org/10.48550/arXiv.1511.07289
32. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning* (Vol. 37, pp. 448–456). JMLR.
33. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.
34. Balasubramanian, V., Ho, S. S., & Vovk, V. (2014). *Conformal prediction for reliable machine learning: theory, adaptations and applications*. Elsevier. https://doi.org/10.1016/C2012-0-00234-7
35. Barber, R. F., Candes, E. J., Ramdas, A., & Tibshirani, R. J. (2021). Predictive inference with the jackknife+. *The Annals of Statistics*, *49*(1), 486–507. https://doi.org/10.1214/20-AOS1965
36. Shafer, G., & Vovk, V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research*, *9*, 371–421.
37. Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer. https://doi.org/10.1007/b106715

Appendix A

Table A1. Features in a sample of dynamics dataset—With provided labels and labels used in this project. We use all except the position as the input features.

ID	Label	Unit	Description
0	Position	km	Position along the track
1	Velocity	km/h	Velocity of the railway vehicle
2	AccB1Y	m/s^2	Lateral acceleration of axle box 1
3	AccB1Z	m/s^2	Vertical acceleration of axle box 1
4	AccCR1Y	m/s^2	Lateral acceleration of bogie A at axle box 1
5	AccCR1Z	m/s^2	Vertical acceleration of bogie A at axle box 1
6	AccB2Y	m/s^2	Lateral acceleration of axle box 2
7	AccB2Z	m/s^2	Vertical acceleration of axle box 2
8	AccCR2Y	m/s^2	Lateral acceleration of bogie A at axle box 2
9	AccCR2Z	m/s^2	Vertical acceleration of bogie A at axle box 2
10	AccB3Y	m/s^2	Lateral acceleration of axle box 3
11	AccB3Z	m/s^2	Vertical acceleration of axle box 3
12	AccCR3Y	m/s^2	Lateral acceleration of bogie A at axle box 3
13	AccCR3Z	m/s^2	Vertical acceleration of bogie A at axle box 3
14	AccB4Y	m/s^2	Lateral acceleration of axle box 4
15	AccB4Z	m/s^2	Vertical acceleration of axle box 4
16	AccCR4Y	m/s^2	Lateral acceleration of bogie A at axle box 4
17	AccCR4Z	m/s^2	Vertical acceleration of bogie A at axle box 4
18	AccB5Y	m/s^2	Lateral acceleration of axle box 5
19	AccB5Z	m/s^2	Vertical acceleration of axle box 5
20	AccCR5Y	m/s^2	Lateral acceleration of bogie B at axle box 5
21	AccCR5Z	m/s^2	Vertical acceleration of bogie B at axle box 5
22	AccB6Y	m/s^2	Lateral acceleration of axle box 6
23	AccB6Z	m/s^2	Vertical acceleration of axle box 6
24	AccCR6Y	m/s^2	Lateral acceleration of bogie B at axle box 6
25	AccCR6Z	m/s^2	Vertical acceleration of bogie B at axle box 6
26	AccB7Y	m/s^2	Lateral acceleration of axle box 7
27	AccB7Z	m/s^2	Vertical acceleration of axle box 7
28	AccCR7Y	m/s^2	Lateral acceleration of bogie B at axle box 7
29	AccCR7Z	m/s^2	Vertical acceleration of bogie B at axle box 7
30	AccB8Y	m/s^2	Lateral acceleration of axle box 8
31	AccB8Z	m/s^2	Vertical acceleration of axle box 8
32	AccCR8Y	m/s^2	Lateral acceleration of bogie B at axle box 8
33	AccCR8Z	m/s^2	Vertical acceleration of bogie B at axle box 8
34	AccCSAY	m/s^2	Lateral acceleration of car body at bogie A
35	AccCSAZ	m/s^2	Vertical acceleration of car body at bogie A
36	AccCSBY	m/s^2	Lateral acceleration of car body at bogie B
37	AccCSBZ	m/s^2	Vertical acceleration of car body at bogie B
38	Curvatura	1/m	Curvature of the circle the track is forming