Predicting Contact Wire Thickness with Machine Learning Algorithms Based on Historical Data

Jeroen Mulder, Frank Vermeulen, Faridaddin Vahdatikhaki, Xianfei Yin

University of Twente, Civil Engineering and Management, Enschede, The Netherlands
ProRail, Asset Management, Utrecht, The Netherlands

ARTICLE INFO

May 2023

Keywords:
Machine learning
Wear rate
Predictions
End of lifetime

ABSTRACT

For the application of predictive maintenance, a new method is needed which can reliable predict the thickness of contact wires for multiple years into the future. For this, a machine learning model is created using multiple linear regression. The most important features for this model are the train passages, historical wear trend and the thickness of the current wire. A wear rate is predicted for every 10 meters which then is projected into the future to determine the remaining thickness. It is found that this approach performs better than typical machine learning methods whereby the thickness is predicted directly. The predicted thickness is compared with the real thickness for individual predictions and that of a whole wire section. For 95% of the predictions, the average thickness of the whole wire can be predicted with an accuracy ±0.12 mm for a prediction horizon of 4.6 years. The results of this study show that even for noisy data useful predictions can be produced with a novel strategy.

1. Introduction

ProRail is the owner of the Dutch railway infrastructure. It is responsible for the construction, maintenance and safety of all assets related to the transportation of trains including train stations. Train operators, such as NS (Nederlandse Spoorwegen), are using these tracks and are responsible for the transits. In order to maintain a reliable schedule, the availability of the tracks must be as high as possible. This can be challenging as the Dutch railway network is the busiest of Europe [1]. This means that all replacement activities that result in downtime for the tracks must be optimized

Figure 1. Pantograph in contact with contact wire.

One of the assets which are critical for the functionality of electric trains is the contact wire (see Figure 1). Contact wires are a part of the overhead power line which transport the electricity parallel to the track. A train makes contact with the copper contact wire through a pantograph which allows it to draw electricity. To make proper contact with the contact wire, the pantograph exerts a force upwards pressing against the wire. When the train moves this will induce friction and eventually wear down the pantograph and contact wire. Over time, the contact wire decreases in size to the point where it must be replaced.

In 2030, it is expected that the travelled kilometres per train in the Netherlands are increased by 23% ~ 34% compared to 2014 [2]. Increase in utilization will cause faster degradation of the contact wires, which subsequently will lead to more frequent replacements of the contact wire.

The objective of ProRail is to move towards predictive maintenance whereby the maintenance activities can be strategically planned ahead of time, based on predicted asset conditions. By doing this, the predictions are more reliable, mostly automated and can be updated every time new measurements are available. Ideally, different maintenance activities are combined if they are affecting the same section. This will reduce downtime and increase the reliability of transportation schedules. Accurately predicted asset conditions allow for a more optimal replacement strategy.

Currently, it is not clear when the contact wire will be at the end of its lifetime. The estimated year of replacement within ProRail is mainly based on extrapolation of the historical thickness and expert judgement. ProRail has valuable data about the train passages and many parameters about their assets which remains unused with this method. The prediction for future thickness can become more accurate when considering more factors such as the number train passages. Also, prediction models from literature about contact wire degeneration are insufficient as these either do not include multiple features or cannot predict the wear into the future.

In order to generate data-driven predictions which can utilize multiple wear factors, machine learning can be used. This predictive modelling method is able to find complex patterns between multiple features in historical data [3]. The goal of this study is to develop a machine learning model that can predict the thickness of the contact wire for 4-5 years into the future based on the historical data. The results will be assessed by calculating the error between the actual thickness and the predicted thickness for multiple years into the future.

This paper is organized as follows. In section 2, the state of the art for contact wire models is described together with the most important wear factors. In section 3, the methodology for the model is explained. In section 4, the results of the model are presented for the performance of the wear rate prediction and the error when using the model for long term predictions. Finally, the discussion and conclusion close this paper.

2. Literature

2.1. *Previous studies on contact wire wear predictions*

Multiple studies have been conducted about predicting the wear of contact wires. In the studies of Derosa et al. and Bucca & Colina [4][5], the same heuristic model is developed and tested on lab data. This model contains three types of wear: mechanical friction, electrical current and electric arcs. The wear resulting for each wear type has given a weight of its contribution. By combining al three components, the total wear per million pantograph passes is determined. Each component contains variables such as the material hardness of the wire but also coefficients that must be tuned. Tuning is performed by conducting multiple tests in the lab with a test rig. Different parameters have been tested separately to determine the wear contribution. The main parameters for this model are the sliding speed, contact force and electrical current. With this model, future predictions can be made by multiplying the wear rate by the number of expected pantographs that will pass. However, lab tests do not encounter imperfections occurring in the field and may therefore be imprecise for real-world use.

Wei et al. [6] also used a heuristic model to predict the wear rate of contact wires. This model uses electric current, sliding speed, contact force and environmental temperature as variables for the model. The outcome of the model is compared to real-world data for one metro track. The outcome of the model and values from the field are consistent. However, all measured wear rates are in a very small range which makes predictions inherently more accurate. Predictions

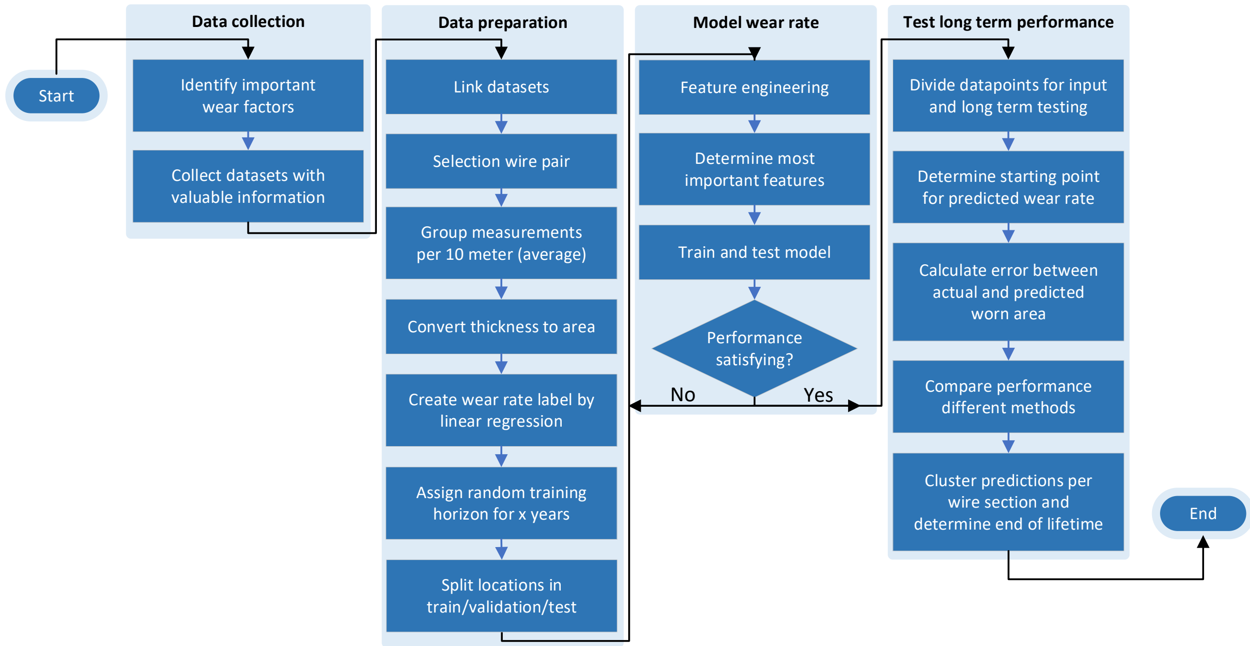

Figure 2. Overview methodology

with this model can be made by assigning the number of expected pantograph passages. Besides this, it is not validated if the results are also accurate for long-term predictions.

Usada [7] made a prediction model to determine the current wire thickness using a neural network. The input for this model solely relies on the signal of the contact force. A window over 27 datapoints of the contact force is used to derive features. The predictions are accurate but cannot be produced for the future with this method.

2.2. *Important wear mechanisms*

Numerous factors are influencing the wear rate of the contact wire. The number of pantograph passages is most dominant in literature as this is linked to most wear mechanisms [7][8]. The electrical current is also often mentioned being a significant contributor as this causes the wire to heat up and causes mechanical stress [9]. The electrical current is the highest during acceleration and deceleration which makes these zones more prone to wear [6]. The contact force is another factor influencing the wear. If the contact force is too low the pantograph will vibrate, bounce and create electrical arcs. If the contact force is too high it will result in excessive friction [10]. Another wear factor that is often mentioned in literature is the train's speed. More heat will be generated at higher speeds which is partly caused by the higher electrical current. Shing [8] found that the wear rate increases when the relative speed is higher. The mentioned wear mechanisms must be considered for model features to predict the wear rate.

3. Proposed method

In Figure 2, the procedure of this study is shown. The process can be divided in four main stages: collecting data, preparation of the data, modelling of the wear rate and testing the long-term performance of the model. In this chapter, the methodology is described according to these steps.

3.1. *Data collection*

3.1.1. *Identify important wear factors*

The most important wear factors are collected via literature and mentioned in the previous chapter. Other available data is also taken into account for generating features. Some measurements could provide additional information about the wear development or can help to derive certain features indirectly if no data is available. The features that are used in this study are explained and motivated later in section 3.3.1 and can also be found in Appendix II.

3.1.2. *Collecting datasets*

For this study, three databases from ProRail are used which hold important information about the wear rate. These datasets consist of information about the measurement train, train passages and speed per segment. The dataset of the measurement train ranges from 2010 to 2018.

The measurement train aims to measure the properties of the contact wires every year for all tracks in the Netherlands. Besides the thickness of the contact wire, the train measures other parameters such as pantograph contact force, horizontal wire position, wire height, cant and. These measurements are combined in one dataset with details about its location. In Table 1, a limited example of the data is shown. To see all relevant columns, see Appendix I.

Table 1. Example data from measurement train (only a few relevant columns are shown)

ID section	Km	Date	Thick avg 1	Thick min 1
078_205BR_64.9	0.00	2015-03-18	9.9	9.9
078_205BR_64.9	0.25	2015-03-18	9.8	9.8
078_205BR_64.9	0.50	2015-03-18	9.9	9.9

At 45 locations in the Netherlands, the deflection of the rail is measured and collected in a database called Quo Vadis. From the rail deflection, the load per axis can be derived. In combination with information about the train formation and its scheduled route, the number of passed trains and the total amounts of tons can be calculated for every rail section. This information is available per month for passenger and goods trains. An example of the data with all used columns can be found in Table 2.

Table 2. Example data from passed trains (only relevant columns are shown)

ID section	Trains travel	Trains goods	Tons travel	Tons goods
005_97V_115.0	1212	1	4065.3	2.5
006_1003L_3.6	1923	40	6132.5	446.6
006_1009AR_4.3	30	2	77.1	18.7

The location and the speed of every train are registered by the control centre. This information is collected for a single day whereby the average speed is calculated for segments of 100 meters for all train passages. It is assumed that the average speed for this particular day is representative for the average railroad traffic throughout the year. In Table 3, an example of the data structure is shown.

Table 3. Example data from train speeds on one specific date (only relevant columns are shown)

ID section	Km from	Km till	Speed avg	Speed max	Speed local
009_131AL_149.4	0	100	103.4	138	140
009_131AL_149.4	100	200	101.4	136	140
009_131AL_149.4	200	300	120.1	139	140

3.2. Data preparation

3.2.1. Linking and filtering dataset

To link the three datasets, the section ID and kilometrage of the tracks are used to create a unique ID. In case data is missing in one of the datasets, the ID is dropped. Also, IDs with less than 8 datapoints per ID are removed.

To make the model more robust, some unique cases are removed. Almost all contact wires in the Netherlands have an initial diameter of 12 mm. Only a few segments are using wires with a diameter of 13 mm and are therefore removed. Also tracks which are used for high-speed rail (HSR) are excluded as these tracks have different overhead wire specifications.

3.2.2. Selection wire pair

Above a single train track, there are two contact wires which form a pair. During the transition between wire sections, there are two overlapping pairs of wires. One wire pair slowly increases in height while the new wire pair slowly decreases its height. This principle can be seen in Figure 3.

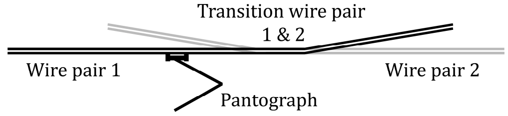

Figure 3. Side view of transition between wire pairs

During this transition, the measurement train measures the two pairs simultaneously. However, it does not indicate which wire is in contact with the pantograph. Because the lowest wire is most likely to be in contact with the pantograph, it is assumed that the two lowest wires form a pair are in contact with the pantograph. If the difference in height between two wires is more than 3 mm, the datapoint is removed as this seems not realistic. Because contact wires are replaced in pairs, the measured parameters of the two wires are combined into one. The wire pair is now presented by an average and minimal value of the wire pair. The average thickness of the two wires is used to create the label. The average and minimum can both be used as a feature, together with the delta thickness between the two wires.

3.2.3. Group measurements

The measurement train gives a datapoint for every 25 centimetres along the contact wire. One problem with this data is that the accuracy of the measurements is low due to deviations in the registered location and the precision of the thickness measurement itself [11]. To reduce the noise of these measurements and limit computing time, multiple datapoints are clustered. The number of datapoints within this group can be adjusted and is for this research set to 10 meters. This means that if all datapoints are valid, 40 datapoints are included per grouped datapoint. The way these 40 datapoints are processed varies per feature. For most features the average of the values is used. Also other transformations are made such as the minimum value and the standard deviation. In Appendix II, the applied transformation for every feature is given.

3.2.4. Conversion from thickness to worn area

Due to the round shape of the wire, the decrease in thickness is not linear. The contact area for a new wire is smaller which makes it decrease faster in thickness when it is new. If the wire wears down, the contact area will increase which results in a slower wear rate for thickness. An example of this phenomenon can be seen in Figure 4. With a wire of 12 mm in diameter, the first worn millimetre is equivalent to 4.5 mm^2. Once the remaining thickness is 8 mm, 11.6 mm^2 is removed to decrease the thickness by another millimetre.

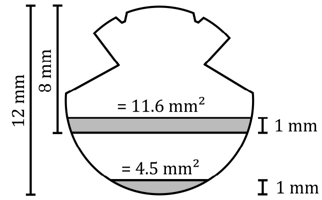

Figure 4: Increased surface area for decreasing thickness

Archard's law states that the removed material due to wear is proportional to the work applied by friction forces [12]. This means that with the same frictional force, a fixed volume of debris will be removed. As can be seen in Equation 1, the contact area is not relevant for the worn volume.

$$V = \frac{k\,F\,s}{H}$$

$V = volume\ of\ wear$
$k = \ wear\ coefficient$
$F = pushing\ force$
$s = sliding\ distance$
$H = hardness\ of\ material$

Equation 1. Archard's wear equation

In the case of contact wires, the frictional force will be the pantograph that slides along the wire. The occurring wear in relation to the passed pantographs will be linear in terms of volume according to Archard's law. If possible, linear relations are preferred when analysing data. Therefore, the thickness must be translated into a volumetric degradation. This can be accomplished by using Equation 2. This formula gives the worn area based on the radius and thickness of the wire. In Figure 5, the variables of the formula are visualised whereby the worn area is shown in grey.

$$a = R^2\ cos^{-1}\ \frac{r}{R} - r\ \sqrt{R^2 - r^2}$$

Equation 2. Conversion of wire thickness to worn area

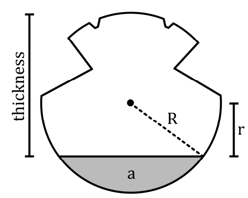

Figure 5. Visualization of variables for conversion wire thickness to worn area

3.2.5. Wear rate label

To train the model, a wear rate is needed which can be linked to the features. This wear rate is the desired output for the model and is called the 'label'. The wear rate is expressed in square millimetres of copper that wears down per year (mm^2/year).

Since the measurement data is noisy, it is impossible to know the exact yearly wear rate. Therefore, an approximation has been made based on historical data. As train schedules and other factors are rather constant for each year, no disrupting changes are expected for the wear rate. The wear rate label is determined using linear regression for all years for which data is available. By doing this, the data of all years are utilized to give a reliable wear rate that is constant for the whole period. In Figure 6, it can be seen that the data from multiple years is transformed into a single yearly wear rate.

In some cases, the deviations in the data are too large to give a reliable wear rate. These cases can be detected by looking at the R^2 score which is given for the fitted regression line. If the R^2 score is low, the line does not fit the data well which is in most cases caused by extreme noise. The slope of the fitted regression line is a second indicator to detect unreliable results. A line might be fitted with a slope that suggests a negative wear rate. This is physically impossible as the worn area can only increase. One possible scenario for such cases is a replacement of the wire. Wire replacements are not registered in a structured way and can thus only be detected properly by analysing the data throughout the years. Another reason for an unreliable wear rate is the number of available datapoints. The more datapoints, the higher the reliability of the wear rate generated by linear regression. For reliable wear rates, the data must thus be filtered on reasonable R^2 scores, positive slopes and the minimum number of available datapoints. For this research, R^2 scores below 0.4 are removed. The same applies for negative wear rates and locations with less than 8 datapoints.

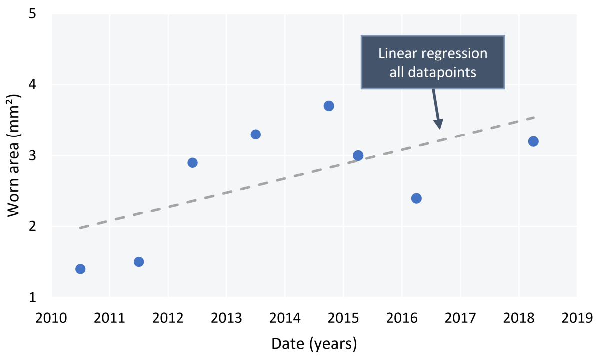

Figure 6. Creation of wear rate label by linear regression over all datapoints

3.2.6. Random training horizon

The training horizon determines how many of the datapoints are included for training. The individual datapoints within the training horizon are combined to a grouped datapoint per feature. For this, the average for each feature is calculated for all included datapoints. The period of the training horizon is assigned randomly. In the example of Figure 7 this period is ranging from 2013 to 2016. Randomly shifting the training horizon avoids biases induced by trends in the data. Also measurement offsets for specific years are neutralized by this method.

By limiting the training horizon, not all datapoints are included. An advantage of using only a few datapoints for training can be the responsiveness of the model. In this way, changes in features will influence the prediction to a greater extent, which is desired if for example the state of the asset has changed suddenly. The more datapoints are used for the training horizon, the more consistent the output will be. Especially if the data contains a lot of noise, smoothing is necessary to avoid outliers influencing the output too much. However, the downside of including many datapoints for training is that a sudden change in the state of the asset will not be immediately detected. With noisy signals it is hard to determine if an outlier is a change in asset condition or just noise. Therefore, an optimum needs to be found between the stability and responsiveness of the model. This will be done in an experiment which will be explained in section 3.4.3.

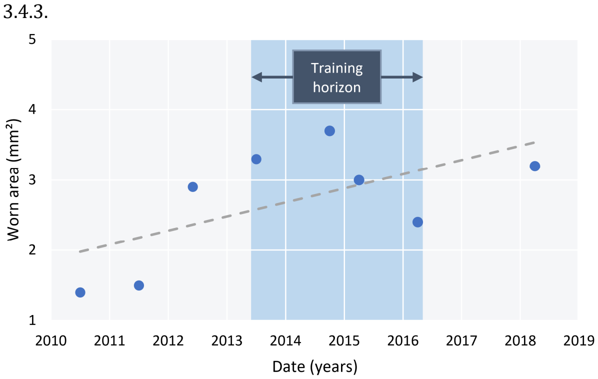

Figure 7. Random assigned training horizon

3.2.7. Split train/validation/test

The ratio between training, validation and testing is 70/20/10. This means that 70% of the data is used to train the model and 20% is used for validation and optimization of the model. After the optimization, the test set of 10% is used to see how the model performs without performance-boosting optimizations. With this ratio most data can be used for training while still remaining a proper validation and testing set.

3.3. Model development

3.3.1. Feature engineering

As mentioned before, the measurements of the measurement train are grouped into blocks of 10 meters whereby the measurements are converted to one datapoint. This conversion is also a part of feature engineering as one measurement value can now be transformed into multiple features. This conversion is done for all dates, which means that a list of features is created for each date. For the input of the model, a single list of features is desired. Therefore, the features from multiple dates are combined by taking the average of all dates. Now every 10 meter block has one single list of features that can be used as input for the model. An overview of these steps is given in Figure 8. Besides that the conversion from a timeseries into a single point allows for a more straightforward machine learning method, the features are also less sensitive to noise.

The features used in the model can be divided into seven categories. The features will be described below per category. For an overview of all features and a brief explanation, see Appendix II.

Contact wire properties

The measurement train has two measurements for each wire; the average and minimum values are measured over a distance of 25 centimetres. For both measurements, a feature is created by taking the average and minimum value for all measurements in the 10 meter block. Also the standard deviation is derived from these measurements. This feature indicates the roughness of the wire within the block of 10 meters.

Another feature based on the contact wire is the delta thickness. This is the difference in thickness between the left and the right wire. This is delta thickness is calculated for the average and minimum measured thickness.

Position wire

The measurement train measures the position of the wire in relation to the centre of the train. The height is one feature that is created based on the vertical position of the wire. The height is measured for both wires which allows to create a feature that indicates the delta height between the two wires. Also the horizontal position is measured, which generates a feature that indicates the distance between the wire and the centre of the track.

Cant

The cant of the track is measured by the measurement train. The tilt of the track indicates indirectly if the 10-meter block is located in a curved section of the track. This is because cant is applied to compensate for the centrifugal forces during a turn.

Speed

The measurement train logs its speed while measuring. This speed is used as a feature to indicate the relative speed of trains. Each section of a track has also a maximum allowed speed. This is used as another feature as trains often try to approach this maximum speed. A third feature to estimate the train speed is based on the average speed of trains logged per 100 meters. This logged data is based on recordings of only one day but is likely the best estimate for the actual train speed. Based on this data, the difference in speed is also calculated. This feature indicates the acceleration and deceleration which indirectly is an estimate for the electrical current.

Passed trains

Between the most important railroad switches, the number of passed trains is measured. For this, a distinction is made between trains transporting passengers and goods which are both used as a feature. Also the total number of passed trains is a feature. The number of trains is an indicator for the number of passed pantographs that have been in contact with the contact wire. This feature can be adjusted if more or fewer trains will pass in the future.

Transported tons

The amount of tons is measured in the same way as the number of passed trains. The amount of tons is known for trains transporting passengers, goods and both combined. For all three measurements, a feature is created. The amount of tons transported can be used as an estimate for the total electrical load exerted on the contact wire.

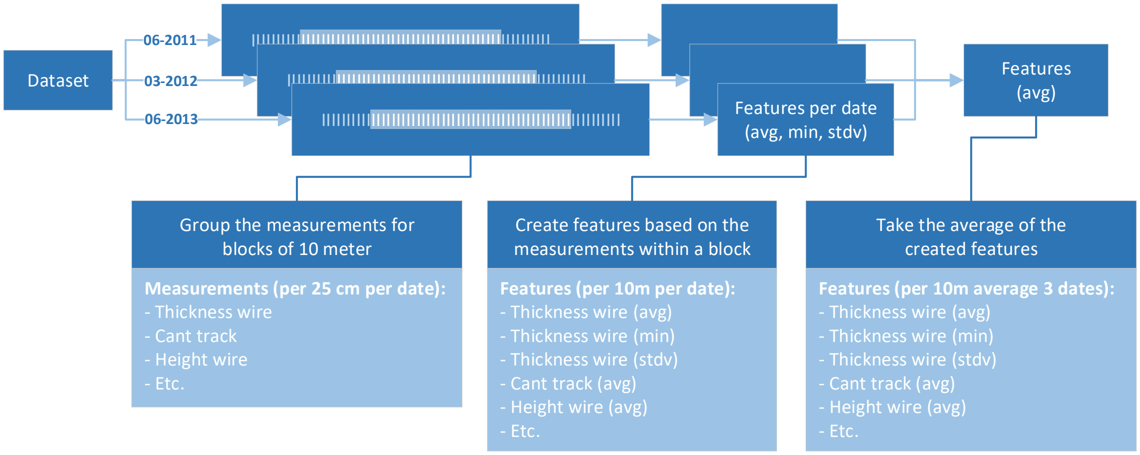

Figure 8. Transformation from raw data to features

Historical trend

Based on the datapoints used for the input of the model, a wear rate trend can be created for this specific period. With linear regression, a trend line can be calculated based on the worn area for the selected datapoints. The historical trend is used as a feature which is an indication of the expected wear rate for an extended time. Increasing the number of used datapoints will improve the stability and accuracy of the slope. The R^2 score of the regression line is also used as a feature to indicate the reliability of the slope. This feature must not be confused with the label of the model. Both apply the same principle, however, this feature only uses the datapoints of the input horizon instead of all datapoints.

3.3.2. *Importance of features*

To identify which features are most important for predicting the wear rate, an analysis is made. For the correlation, each feature is compared with the wear rate label of the model. A feature with a strong correlation is likely a good predictor [13]. However, this is not always the case because of confounding variables or lacking causation. When performing linear regression, the model assigns a weight to each variable. When the features are normalized, an importance score for the linear regression model can be determined for each feature. As these weights can be different per run due to the optimization process, the average of 3 runs is used to determine the importance. These weights are then scaled so the maximum score is 1.

3.3.3. *Train and test model*

The machine learning models that are considered are multi-linear regression, random forest, gradient-boosted tree and neural network. For each model type, the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R-Squared (R^2) are measured. These metrics represent the accuracy of the predicted wear rate compared to the actual wear rate. The MAE gives the average absolute difference between the prediction and the actual value. The MSE is the average squared difference between the predicted and actual values. The RMSE is the square root of the MSE metric which converts the squared difference back into the original units. The R^2 score indicates the percentage of variance between the predicted and actual values. The R^2 scores ranges from 0 to 1, whereby an R^2 of 1 means that the predictions are identical to the actual values. In this study, the focus is on the RMSE score when comparing different models.

3.4. *Test long-term performance*

3.4.1. *Divide datapoints*

To determine the long-term performance of the model, the model is tested on a subset of datapoints. In the example of Figure 9, it can be seen that the first four datapoints are used as input for the model. The four datapoints are creating the feature values which are used by the model to generate a predicted wear rate. This wear rate will be projected into the future. The model has no knowledge about any information within the 'prediction horizon' and is only meant to compare the prediction with the actual values.

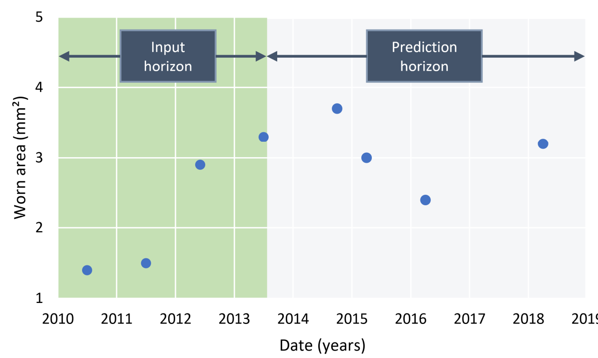

Figure 9. Division of datapoints for model input and datapoints to test the long term performance

3.4.2. *Starting point of prediction*

The model gives a predicted wear rate as output. This predicted wear rate is a single value and will appear as a slope when plotting over time. In Figure 10, the predicted wear rate is drawn in orange and deviates in this example slightly from the actual wear rate. In order to produce a good prediction in the future, the starting point of the predicted slope is important. Theoretically, the best starting point matches the regression line of the actual wear rate at the end of the limited horizon. In the figure, this theoretical best starting point is shown with a blue diamond. If the prediction is drawn from this optimal starting point, the error at the target point would be as small as possible. However, this point cannot be known by only having information about the input horizon. Therefore, an experiment will be conducted in a later stage to find the optimum starting point.

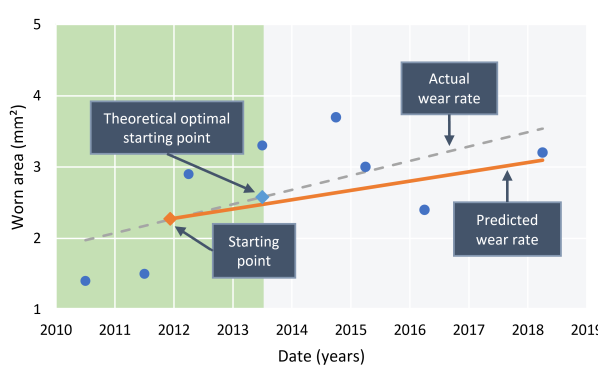

Figure 10. Example of a starting point for the predicted the wear rate

For the experiment to determine the optimal starting point, 6 different starting points are tested. In Figure 11, these starting points are visualized. The simplest starting point is the last known worn area within the input horizon and is indicated as 'last'. Because noise can make this last datapoint unreliable, also a starting point is based on a regression line for all datapoints within the input horizon. This starting point is called ‘last trend’. The same is done for the first datapoint which is labelled as ‘first trend’. The other 3 starting points are based on an average of multiple datapoints. In the example 4 datapoints are included in the input horizon. The starting point ‘avg 4 of 4’ takes the average value for the X- and Y-values of the datapoints. The average for X will then be the average date and for Y the average worn area. The starting point ‘avg 3 of 4’ does the same, but instead of using all 4 datapoints, only the 3 most recent datapoints are included. The same principle applies to ‘avg 2 of 4’, whereby only 2 datapoints are used to determine the average of the X- and Y-axis.

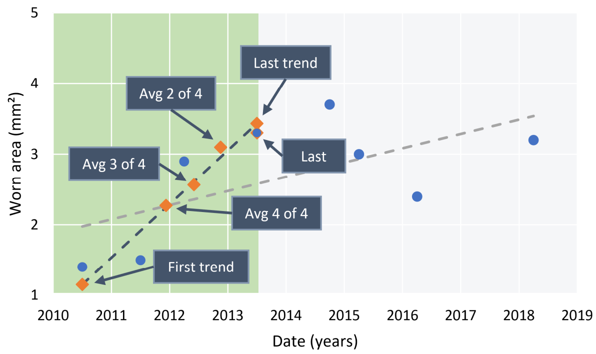

Figure 11. Different starting points for the predicted wear rate

As can be seen in the figure, most starting points are not at the end of the training horizon. Because of this, bad predictions have a larger error as it now has a longer prediction period. However, this starting position might be closer to the regression line of the label which eventually produces better long-term predictions. The best starting point is determined by its long term prediction performance which will be explained in the next section.

3.4.3. Error for long term predictions

The model is trained to replicate the real wear rate based on the regression line of historical data. To measure the difference between the predicted and actual worn area, the two values can be compared. This comparison takes place at the latest point in time and is called the target. The target is considered to be the actual worn area and is based on the regression line which is also used for the label of the model. In Figure 12, the target is shown with a blue diamond. The error of the prediction must be minimized and is quantified by the MAE, MSE, RMSE and R^2 metrics.

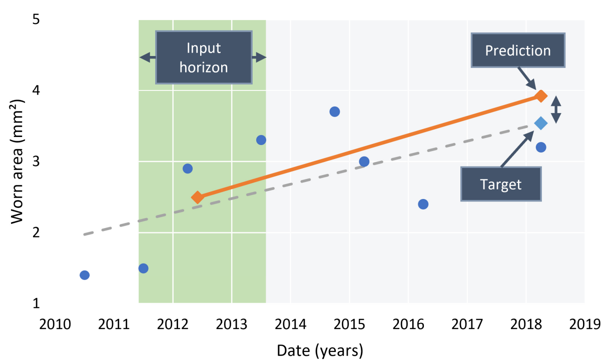

Figure 12. Calculation of prediction error by comparing predicted value with the target (3/4 datapoints)

The further the predictions are in the future, the greater the errors will be. The calculated error depends on the duration of the input horizon and prediction horizon. For this reason, an experiment is conducted that tests the performance for multiple combinations. For the experiment, the start of the prediction horizon will be held the same while varying the input horizon. In this way, the distance to the target will stay fixed which gives a fair comparison in terms of performance. In Figure 13 the setup of the experiment is shown. The green squares represent the input horizon and the grey squares the prediction horizon. The performance will be measured at the target which is the latest datapoint. For all other analysis, 4 datapoints are used for the input horizon in this study.

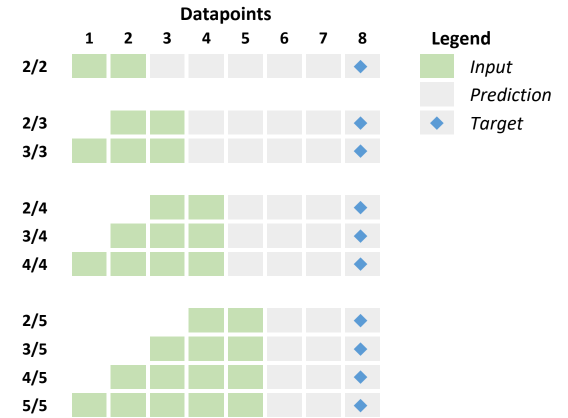

Figure 13. Visualisation of experiment with various prediction horizons while varying input datapoints

3.4.4. Compare different prediction methods

Many methods are possible to make future predictions with the available data. The earlier explained method is most suitable for long-term predictions with noisy data. To prove this, this prediction method is compared to 5 other methods which are using a fundamental different principle. These 5 alternative methods are compared in long term performance with the main model which is called the ‘extended wear rate’ method. The alternative methods are explained below.

Limited wear rate

The main model uses an extended wear rate based on a regression created from all datapoints, while the trend of the wear rate within the training horizon is used as a feature. This feature of the wear rate trend can be an important indicator as it reveals how fast the wear rate has been over a shorter period of time. A downside of this approach is that the wear rate over all years might generalize too much as it includes many datapoints. For this reason, a model is created which uses the wear rate generated within the random training horizon as the label. Therefore, the feature of the wear rate trend will be dropped as this has now become the label of the model. In Figure 14, the alternative method is visualized.

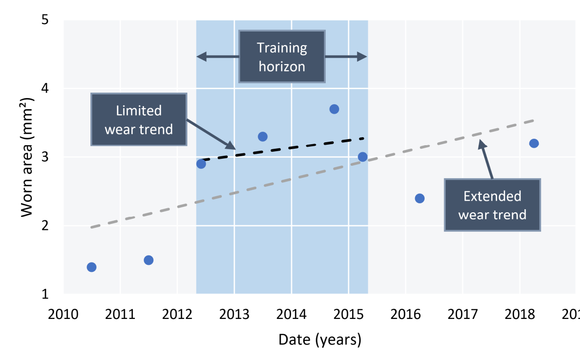

Figure 14. Alternative method with the limited wear rate trend as label for the model

Direct prediction worn area

The main model uses two steps to predict the worn area for the future. First, the wear rate is trained which then starts from a calculated starting point. By having two steps where errors occur, the predictions can be less accurate. Therefore, a model is created that tests the accuracy if the worn area is predicted directly. The input is the datapoints within the training horizon and the label is the target. The target is the actual worn area for the last datapoint based on the regression line of all years. The training horizon is no longer random but is now always made in such a way that the prediction horizon is as long as possible. This principle can be seen in Figure 15, whereby the input horizon always starts at the first datapoint. An extra feature for this model type is created which indicates the length of the prediction horizon.

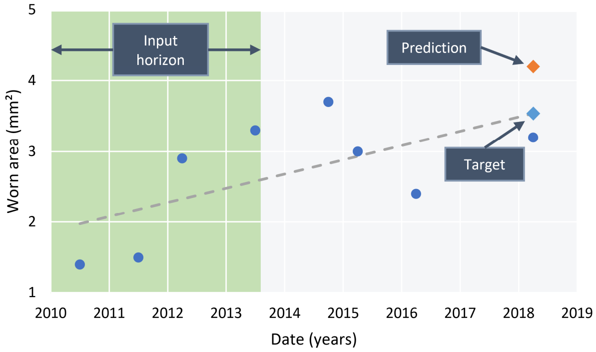

Figure 15. Alternative model whereby the target is used as label and the worn area is predicted directly

Optimal wear rate

The previously explained model needs to learn two principles by itself, namely the wear rate and the optimum starting point. As this can be complex to learn, a model is created which helps with the starting point. The principle of the model is similar to the previously explained model that predicts the worn area directly. However, instead of predicting the worn area, the delta between the starting point and the target is calculated. Thus, the amount of worn area which occurred within the prediction horizon. By doing this, a wear rate can be calculated that would be optimal if started from the assigned starting point. In Figure 16, it can be seen that the optimal wear rate is determined by the position of the starting point and the target.

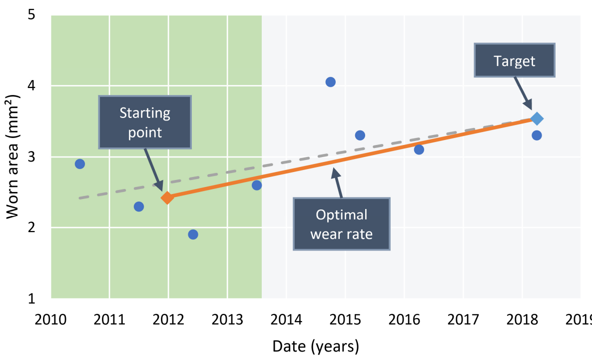

Figure 16. Alternative method with the wear rate label based on the slope between the optimal starting point and target

Extrapolation historical trend

Another method for predicting the worn volume in the future is by simply extrapolating the historical trend. This method does not require any features as it only relies on historical data of the worn area. In Figure 17, it can be seen that the historical trend is projected into the future. This method is currently one of the tools used by ProRail.

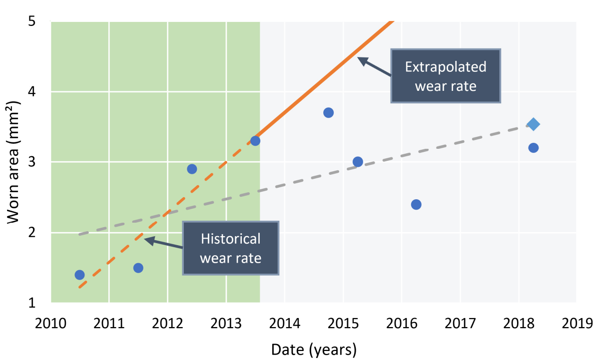

Figure 17. Alternative method whereby the historical wear rate trend is extrapolated

3.4.5. *Cluster predictions per wire section*

For every 10 meters, the worn area is predicted for the future. A single contact wire can have a length of more than a kilometre and will be replaced as a whole. The predictions per 10 meters must thus be combined into a single prediction for the whole wire. For this, the average is calculated based on all predictions within the section. Also percentiles can be useful for the replacement criteria as the places containing the worst conditions are the limiting factor for the life span. Wire sections with less than 5 predictions are removed as these results are unreliable.

4. Results

4.1. *Importance and correlation of features*

In total, 24 features are generated which can be divided into 7 categories. Within these categories, it is possible that features have a mutual correlation. Each model deals differently with correlated features so initially all features are used in the model. As can be seen in Table 4, the linear regression and random forest model both consider different features important. Both models put the emphasis on the properties of the wire. The linear regression model also uses the train passages, while this is insignificant for the random forest model. In turn, the random forest model values the trend of the wear rate more. In some cases the importance for one of the two model types is high, but the correlation between the feature and the label is low. This can be explained by interaction among multiple features. This can also be the other way around, whereby the correlation is low but the importance is high. The table shows the importance scores when applying 4 datapoints as input. A general description per feature can be found in Appendix II.

When the number of datapoints is increased or decreased, the importance values will slightly change. It must be stated that especially the importance scores for the linear regression model are volatile. The importance can vary significantly depending on the split between training and validation. However, features with a really low importance score tend to stay low regardless of the split in data.

Table 4. Importance and correlation for average thickness sorted by importance linear regression model.

Feature	Importance linear regression	Importance random forest	Correlation
Thick_avg	1.00	0.27	-0.56
Trains_total	0.93	0.01	-0.05
Trains_travel	0.88	0.01	-0.03
Thick_min	0.58	0.24	-0.58
Thick_avg_min	0.41	0.24	-0.58
Wear_rate_trend	0.20	0.56	0.51
Thick_min_dev	0.20	0.49	0.67
Thick_min_delta	0.19	1.00	0.70
Thick_min_min	0.18	0.48	-0.61
Tons_travel	0.14	0.02	-0.01
Tons_total	0.13	0.01	-0.08
Tons_goods	0.13	0.01	-0.13
Thick_avg_delta	0.06	0.03	0.48
Trains_goods	0.06	0.01	-0.12
Horizontal_avg	0.03	0.01	0.03
Thick_avg_dev	0.02	0.82	0.68
Cant	0.02	0.04	-0.05
R2_wear_rate_trend	0.02	0.02	0.01
Delta_height	0.02	0.00	0.08
Speed_field	0.01	0.01	0.00
Speed_local	0.01	0.00	0.01
Height_avg	0.01	0.00	0.11
Speed_measure	0.00	0.00	-0.01

4.2. *Performance of model types*

Using all 24 features, 4 machine learning models are tested for their performance. As can be seen in Table 5, Linear regression has the best performance with an RMSE of 0.146 and is for this reason the main model in this study. The random forest model and gradient-boosted tree are scoring worse but still give reasonable results. The neural network model fails to find patterns and cannot give proper predictions. This model type normally performs well when dealing with non-linear data. However, most relationships are linear and the data contains a lot of noise. In those cases more simple models can outperform this more complex method [14].

When only including the most important features or removing correlating features, the performance did not increase. Also tuning the model did not result in better performance. For this reason all other analyses in this paper are based on using all 24 features.

Table 5. Performance metrics of the model using all 24 features

Model type	MAE	MSE	RMSE	R2
Linear regression	0.094	0.021	**0.146**	0.682
Random forest	0.097	0.024	0.154	0.644
Gradient boosted tree	0.097	0.029	0.170	0.568
Neural network	0.135	0.034	0.185	0.020

In Figure 18, a scatterplot is shown of the predicted wear rate versus the actual wear rate for the linear regression model. The predictions have a relatively small error for the lower wear rates. When the wear rate increases, the predictions become slightly less accurate.

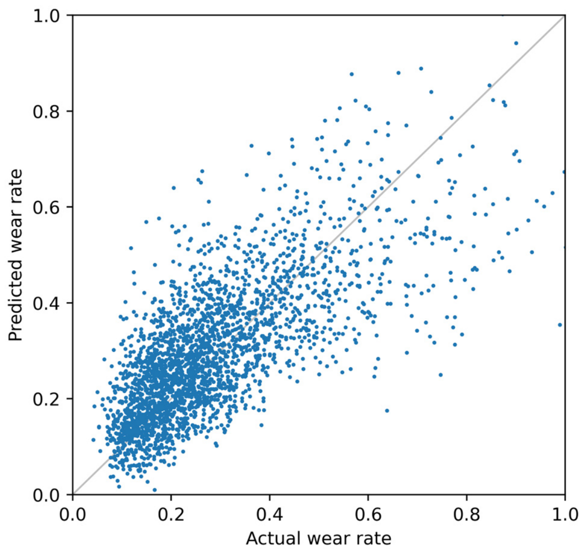

Figure 18. Scatterplot of the linear regression model showing predicted versus actual wear rates

In Figure 19, a histogram of the prediction errors can be found for the linear regression model. The distribution reassembles a bell-shaped curve, which indicates a normal distribution. The distribution can be considered centred and symmetric, which means that the overall trend is captured well.

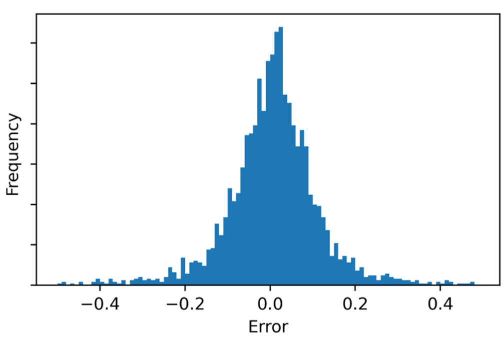

Figure 19. Distribution of the wear rate prediction error

The number of datapoints used as input affects the importance of the features. Especially the feature about the historical wear rate is affected as more datapoints will give a more reliable regression line. More datapoints also reduce the noise for all other features. In Table 6, the RMSE scores are shown per number of input datapoints used for the model. Only linear regression and random forest are considered as these models are best performing. It can be seen that the more datapoints are included, the better the prediction can mimic the wear rate label.

Table 6. RMSE per number of datapoints used for training to predict the wear rate

Datapoints input	Linear regression	Random forest
2	0.153	0.164
3	0.148	0.159
4	**0.146**	0.154
5	0.138	0.149

4.3. *Optimal starting point*

To determine the best starting point for the predicted wear rate, an experiment is conducted. In Table 7, the RMSE scores per starting point are shown for the long term performance with 4 datapoints as input. The starting point that gives the best performance is 'avg 4 of 4'. This starting point takes the average of the X- and Y-axis for all datapoints included in the input horizon. Other variations of the starting point produce significant worse predictions.

The starting point 'avg 4 of 4' is the center of all datapoints that are used for input. The features that are used to predict the wear rate are also based on the average of all datapoints. The feature values are thus centered to the same average date as the starting point. The RMSE score suggests that the best starting point is bound to the transformation of features where the model is trained on. The best starting point which uses the average of all datapoints is used for every analyses in this study.

Table 7. RMSE per starting point

Starting point	RMSE
Avg 4 of 4	**0.863**
Avg 3 of 4	0.957
Avg 2 of 4	1.039
Last	1.043
Last trend	1.041
First trend	1.193

4.4. *General long term prediction performance*

To test how well the model performs if predictions are made for the future, the predicted worn area is compared to the actual worn area. In Table 8, the RMSE scores can be found for 4 different prediction horizons using the 'extended wear rate' model. Because of the limited available datapoints, not all prediction horizons can be evaluated with 5 datapoints as input for the model. In the table, the maximum number of datapoints for the input of the model are shown on the y-axis. When using fewer datapoints as input, more datapoints are available to test the long term performance. By testing for multiple prediction horizons, the performance can be compared for different years of prediction. As the prediction horizon is different when changing the number of maximum datapoints, the comparison between RMSE scores should only be row-wise. The column of the prediction horizon is based on the average years of prediction which is dependent on the dates of the datapoints and are therefore not whole numbers.

Table 8. RMSE of individual long-term predictions

	Datapoints input				Prediction
Max	2	3	4	5	horizon (years)
2	0.94	-	-	-	6.8
3	1.14	1.06	-	-	5.8
4	0.93	0.90	**0.86**	-	4.6
5	0.74	0.76	0.75	0.73	3.4

As can be seen in the overview, the performance for each number of maximum datapoints does not show a very clear pattern. When looking at a maximum of 5 datapoints, using 2 datapoints as input gives a similar performance as using 5 datapoints for input, while 3 and 4 datapoints show lower performance. In general it seems slightly better to include more datapoints.

Table 6 showed that more datapoints can better replicate the actual wear rate. However, more datapoints do not necessarily result in better long-term performance. This is because more factors are important when predicting the worn area over time. The model with less datapoints as input has a higher responsiveness to change as it considers a shorter time frame. Meanwhile, the model with more datapoints is more passive as it considers a longer time frame but is more stable due to noise filtering. Also the starting point for both models is different. The starting point with more datapoints is further back in time, which increases the error over time due to a longer prediction length. At the same time, the starting point has a higher accuracy as it is based on more datapoints. The results imply that the mentioned factors are almost equally important as the models produce similar performance scores.

In Figure 20, a scatterplot is shown for the model which uses 4 datapoints as input and with a prediction horizon of 4.6 years. The prediction errors are symmetric and form a normal distribution. It can be seen that the accuracy of the predictions slightly decreases when the wire is worn down further. This can be explained by the measurement accuracy. Due to the round shape of the wire, a new wire decreases faster in thickness compared to an older wire. This principle is explained in Figure 4. The accuracy of the measurements is based on the contact area of the wire which is thus more precise to measure for newer wires.

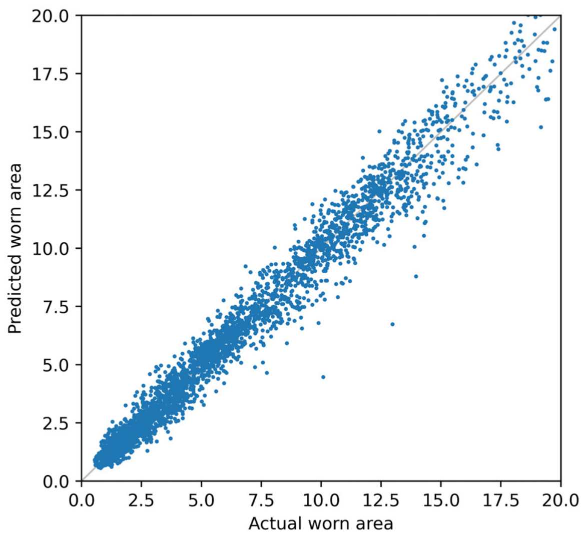

Figure 20. Scatterplot of long-term performance using worn area

When the worn area is converted back to thickness, the results are easier to interpret. In Figure 21, a scatterplot can be found for the thickness. It can be seen that most contact wires have a thickness of more than 10 mm. As the predicted thickness is especially important for thinner wires, the predictions for these thin wires need to be accurate. The model has a very slight bias for the thinnest wires in the dataset. The accuracy for this category could be improved if the majority of the training data consists of thin wires.

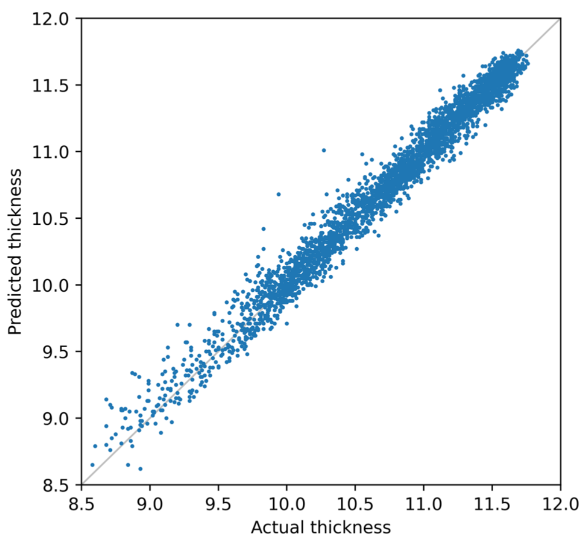

Figure 21. Scatterplot of long-term performance using thickness

The accuracy of the model is assessed by using different confidence levels. In Table 9, the accuracy is shown for the worn area and thickness of the wire. It must be noted that the accuracy of the thickness is dependent on the diversity of the sample. This is because the non linear conversion from worn area to thickness. As can be seen in the scatterplot of Figure 21, the majority of the wires are relatively new which results in a lower prediction accuracy compared to a majority of older wires. The average error is found to be 0.563 mm^2 for the worn area and 0.078 mm for the thickness when predicting 4.6 years into the future.

Table 9. Accuracy of long term predictions for 4.6 years per individual prediction

Confidence	Accuracy worn area	Accuracy thickness
90%	±1.23 mm^2	±0.16 mm
95%	±1.61 mm^2	±0.20 mm
99%	±2.67 mm^2	±0.30 mm

4.5. *Comparison of different prediction methods*

The long-term performance metrics allow for a fair comparison between multiple prediction methods as the goal for all methods is to accurately predict the worn area in the future. In Table 10, all considered methods can be found sorted from best to worst performance.

As can be seen in the overview, the best performance score is achieved by the 'optimal wear rate' method. For this method, two variants are tested; one that is trained with a fixed prediction horizon and one with a flexible prediction horizon. As can be seen in the table, the fixed method is performing significantly better. Despite this being the best-performing method, it is questionable if this is also true on new data. The model is trained on only the first few years and tested on the last year. Therefore, it cannot compensate for unwanted trends in the data. Another problem with this model is that it is trained on a fixed prediction horizon. It can therefore not reliably predict data for a different prediction length. This might be especially problematic for predictions further in the future, outside of the training data. For predictions with a shorter prediction horizon, a different model can be made using the preferred prediction period. The 'optimal wear rate' method with a flexible prediction horizon does solve these mentioned problems but performs worse. When excluding the 'optimal wear rate'-models with the mentioned drawbacks, the method 'extended wear rate' performs best. This method is considered most suitable as this model is more resistant to these issues. Therefore, every analysis in this paper is based on the 'extended wear rate' method.

Table 10. Performance of different methods for long-term predictions

Prediction method	MAE	MSE	RMSE
Optimal wear rate (fixed)	0.555	0.616	0.785
Optimal wear rate (flexible)	0.563	0.744	0.862
Extended wear rate [MAIN]	0.562	0.745	**0.863**
Limited wear rate	0.593	0.885	0.966
Direct prediction worn area	0.736	1.114	1.055
Extrapolation	1.596	7.663	2.768

4.6. *Clustering to wire section*

In Table 11, the RMSE scores can be found when the individual predictions are clustered per wire section. The pattern for the wire section is similar to that of the predictions per 10 meters.

Table 11. RMSE-scores for clustered wire sections

	Datapoints input				Prediction
Max	2	3	4	5	horizon (years)
2	0.62	-	-	-	6.8
3	0.69	0.65	-	-	5.8
4	0.58	0.55	**0.54**	-	4.6
5	0.49	0.47	0.45	0.44	3.4

When converting the worn area back to thickness, an average error (MAE) of 0.058 mm is obtained. By clustering values for the whole wire, the MAE decreased from 0.078 for individual predictions to 0.058. This means that to a certain extent the errors cancel each other out when clustered. Predictions that are too high are compensated by predictions that are too low. This is also the case for the accuracy of the predictions shown in Table 12. The accuracy is improved from ±0.20 mm for the individual predictions to ±0.15 mm for the clustered predictions. The average error is found to be 0.38 mm^2 for the worn area and 0.05 mm for the thickness.

Table 12. Accuracy of predictions for 4.6 years per wire section

Confidence	Accuracy worn area	Accuracy thickness
90%	±0.86 mm^2	±0.11 mm
95%	±1.02 mm^2	±0.12 mm
99%	±1.87 mm^2	±0.21 mm

5. Discussion

For all tested models, the current thickness appeared to be one of the most important feature categories. Takahashi, et al. [11] also state that the residual diameter is correlated with the local wear rate. Nonetheless, this feature is not mentioned often in other studies. In theory, converting the thickness to worn area should create a linear wear rate [12]. However, other external factors seem to make this statement incorrect forcing the model to use this feature to compensate for non-linearity. According to the data, new wires wear down faster even in terms of worn area. This non-linearity does not have a significant effect on the model. When using an acceptable timeframe for creating features and the label, this effect is barely noticeable as the wear of contact wires is relatively slow.

Another important feature category is the number of passed trains. This seems very intuitive as almost all wear mechanisms are based on contact between the pantograph and contact wire. Many studies have found that the number of train passages is a key factor for wear and tear [15][10][8].

Another feature that many studies found important is the speed of the train [4][6]. Remarkable is that in the data no relation is found between the speed and the wear rate. All three features which approximate the actual train speed fail to find a connection. Also the difference in speed does not give any useful information to the model. According to Wei et al. [6], significantly higher wear rates are expected in the acceleration and deceleration zones due to an increasing electrical current. In this study the difference in speed should indicate these zones but do not show a relation. For estimating the total electrical current, the amount of passed tons is used as a feature. This feature is found to be important for the wear rate. However, it is not clear if this is caused due to capturing the total electrical current or if this feature is related to the number of passed trains.

The contact force is a feature that is a promising feature according to literature. A strong correlation between the contact force and the label could not be found in the initial stage. Because a lot of data was missing, this feature is not used for the model in this study.

With an importance score of 0.20 for linear regression and 0.56 for the random forest, the wear rate trend feature pushes the predictions of the model in the right direction. It must be noted that the input horizon must be chosen wisely when including this feature. The more datapoints included, the closer the wear rate trend will be to the label of the model. On one hand this might be good as this historical wear trend is a good indicator for the wear in the future. On the other hand, the model might become unreliable if the two are too close. A good way to deal with this is to keep the prediction horizon close to the desired prediction length. If more data is available, more datapoints can be used as input without compromising the desired prediction horizon. In this way, the model can still be tested on the preferred prediction horizon, while increasing the correlation between the historical wear rate trend and the label.

A standard machine learning approach does not work well due to noise and missing data such as the age of the contact wire. Some tweaks were necessary to improve the performance. Despite this being a new methodology, the validity has been proven by comparing the predicted and actual worn area. It must be noted that the actual worn area is an approximation as the real value is not known. However, by being tested on a large dataset, the results can be considered reliable.

For training, all locations with a messy wear pattern have been removed. The wear rate label is created by applying linear regression on the worn area for all datapoints. The fit of this line is indicated with an R^2 score. All R^2 scores below 0.4 are filtered out, as this hinders the learning capability of the model. The performance of the model could probably be improved by increasing the minimum R^2 score. On the other hand, filtering on R^2 scores might create a bias for the model. It is not known if the average thickness of the section can be estimated better by including all cases or by predicting only a part of the cases with higher accuracy. By using low R^2 scores also the target becomes unreliable which makes it hard to validate the performance.

Another factor that must be taken into account is the presence of carbon deposits. This debris is caused by friction between the contact wire and the pantograph and makes the surface of the contact wire appear wider. The measurement train calculates the thickness of the wire based on the contact surface and interprets a thicker wire if carbon debris is present. This measurement error causes noise and induces a slight deviation in terms of thickness. Besides the carbon deposit, the location of the measurement train is not always accurate. Currently this problem is mainly solved by clustering the measurements per 10 meters. More precise data could be created by applying a synchronization algorithm which will most likely slightly improve the performance of the model.

The model assumes that the features will stay the same in the future. If it is known that certain values will change in the future and a forecast is available, the values for the features should be updated. If for example the number of trains is expected to increase by 10% in the coming years, this percentage can be added to the current value. For a prediction for 5 years with for example an expected increase of 10% after 2 years, the average value should be calculated. This would be an average increase of 6% over 5 years, which means that the current feature value should be multiplied by a factor of 1.06 to anticipate for this change.

The primary goal of this study is to accurately predict the end of lifetime of a contact wire. This can be done by predicting the average thickness of the whole wire and applying a threshold for the minimum allowed thickness. As the prediction of the wire section consists of many individual predictions, this allows for a more complex replacement criteria. Also percentiles can be used, so that for example at least x percent of the predicted values must be above a certain thickness threshold. More research can be done to determine which replacement algorithm is most suitable.

The measurement train logs the average and minimum thickness measured over a section of 25cm. Both thickness values can be important to determine the end of the lifetime of the wire. This study focussed on the average thickness. However, the same method can be applied to the minimum thickness. The model performance for the average and minimum thickness as label of the model are almost identical.

This study has shown that even with noisy data useful predictions can be generated. This method may be also useful for other assets whereby predictions must be made over a longer period of time. Especially if a lot of noise is present in the data, this method might perform better than most common solutions.

6. Conclusion

The goal of this study was to predict the thickness of the contact wire for 5 years in the future. In total 24 features are created which can be divided into 7 categories. The most important feature categories for the model are the thickness of the wire, the number of passed trains, the amount of transported tons and the historical wear rate trend. Multiple machine learning algorithms are tested whereby linear regression is best performing. Besides multiple machine learning algorithms, also different prediction methods are tested. Directly predicting the thickness based on the features does not produce optimal results. The model that performs best tries to predict the wear rate instead. For determining the wear rate, the thickness of the contact wire is first converted to the worn area to stimulate linearity. For each location multiple datapoints are available and are used to fit a regression line. This regression line represents the wear rate and is the label of the model. Once the machine learning model is trained and wear rates can be predicted, a starting point is needed. The best starting point lies in the centre of the datapoints used as input for the model. This centre can be seen as the centre of mass and is the average of the x-axis (date) and y-axis (worn area). To test the long-term performance, the predicted worn value is compared to the actual value up to 6.8 years in the future. The number of datapoints that are used as input for the model does not affect the performance significantly. Similar results can be achieved by using datapoints within the range of 2 to 5 datapoints as input. Adding more datapoints makes the model more robust but less responsive to change. The robustness and responsiveness are considered almost equally important as the performance is similar. The predictions are made per 10 meters and are eventually clustered per wire section. The estimated average thickness of a whole wire has an accuracy of ±0.12 mm at a 95% confidence level.

7. References

[1] CBS (Centraal Bureau voor de Statistiek), "Hoe druk is het nu daadwerkelijk op het spoor?," 2009. [Online]. Available: https://www.cbs.nl/nl-nl/nieuws/2009/09/nederland-heeft-drukste-spoor-van-de-eu.

[2] Centraal Planbureau, "Ontwikkeling Mobiliteit," 2020.

[3] S. Dridi, "Supervised Learning - A Systematic Literature Review," no. December, 2021, doi: 10.13140/RG.2.2.34445.67049.

[4] G. Bucca and A. Collina, "Electromechanical interaction between carbon-based pantograph strip and copper contact wire: A heuristic wear model," *Tribol. Int.*, vol. 92, pp. 47–56, 2015, doi: 10.1016/j.triboint.2015.05.019.

[5] S. Derosa, P. Nåvik, A. Collina, G. Bucca, and R. Anders, "A heuristic wear model for the contact strip and contact wire in pantograph – Catenary interaction for railway operations under 15 kV 16 . 67 Hz AC systems," vol. 457, 2020, doi: 10.1016/j.wear.2020.203401.

[6] X. K. Wei, H. F. Meng, J. H. He, L. M. Jia, and Z. G. Li, "Wear analysis and prediction of rigid catenary contact wire and pantograph strip for railway system," *Wear*, vol. 442–443, no.

March 2019, 2020, doi: 10.1016/j.wear.2019.203118.

[7] T. Usuda, “Estimation of wear and strain of contact wire using contact force of pantograph,” *Q. Rep. RTRI (railw. Tech. Res. Institute)*, vol. 48, no. 3, pp. 170–175, 2007, doi: 10.2219/rtriqr.48.170.

[8] W. Shing, “A survey of contact wire wear parameters and the development of a model to predict wire wear by using the artifical neural network,” 2011.

[9] S. G. Jia, P. Liu, F. Z. Ren, B. H. Tian, M. S. Zheng, and G. S. Zhou, “Sliding wear behavior of copper alloy contact wire against copper-based strip for high-speed electrified railways,” *Wear*, vol. 262, no. 7–8, pp. 772–777, 2007, doi: 10.1016/j.wear.2006.08.020.

[10] H. Wang, A. Núñez, Z. Liu, Y. Song, F. Duan, and R. Dollevoet, “Analysis of the evolvement of contact wire wear irregularity in railway catenary based on historical data,” vol. 56, no. 8, pp. 1207–1232, 2018, doi: 10.1080/00423114.2017.1408919.

[11] A. Takahashi, T. Kishi, and H. Yamamoto, “Overhead Contact Line Monitoring and Prediction of Contact Wire Localized Wear Points,” no. 29, pp. 22–25, 2014.

[12] J. F. Archard, “Contact and rubbing of flat surfaces,” *J. Appl. Phys.*, vol. 24, no. 8, pp. 981–988, 1953, doi: 10.1063/1.1721448.

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*. 2017.

[14] M. Kuhn and K. Johnson, *Applied predictive modeling*. 2013.

[15] G. Bucca and A. Collina, “A procedure for the wear prediction of collector strip and contact wire in pantograph-catenary system,” *Wear*, vol. 266, no. 1–2, pp. 46–59, 2009, doi: 10.1016/j.wear.2008.05.006.

Appendix I – Example data measurement train

Only relevant columns are shown. The numbers in the headers refer to the wire that is measured. The measurement train can measure up to 4 wires at the same time.

ID section	Km	Wire section	Date	Thick avg 1	Thick min 1	Thick avg 2	Thick min 2	Thick avg 3	Thick min 3	Thick avg 4	Thick min 4
078_205BR_64.9	0.00	078_65.013_207R_229L	18/03/2015	9.9	9.9	10.1	10.1	11.8	11.8	11.8	11.8
078_205BR_64.9	0.25	078_65.013_207R_229L	18/03/2015	9.8	9.8	10.6	10.3	11.9	11.9	11.8	11.8
078_205BR_64.9	0.50	078_65.013_207R_229L	18/03/2015	9.9	9.9	10.1	9.9	11.8	11.8	11.8	11.8

Height 1	Height 2	Height 3	Height 4	Position 1	Position 2	Position 3	Position 4	Speed	Cant
-54.5	-56.5	-87.3	-87.3	-47.6	-87.3	288.7	249.0	78.8	-3.1
-55.5	-57.5	-87.3	-86.3	-49.6	-90.3	286.7	247.0	78.3	-3.2
-55.5	-58.5	-88.3	-86.3	-48.6	-90.3	284.7	244.1	78.5	-2.8

Appendix II – Features model

Category	Feature	Description	Calculation method
Properties wire	Thick_avg	Average thickness of the contact wire	Average of group
	Thick_avg_delta	Difference in average thickness between the left and right contact wire	Average of group
	Thick_avg_dev	Deviation of the average thickness along the wire	Deviation of group
	Thick_avg_min	Smallest value of the average thickness	Minimum of group
	Thick_min	Minimum thickness of the contact wire	Average of group
	Thick_min_delta	Difference in minimum thickness between left and right contact wire	Average of group
	Thick_min_dev	Deviation of the minimum thickness along the wire	Deviation of group
	Thick_min_min	Smallest value of the minimum thickness	Minimum of group
Position wire	Delta_height	Difference in height between the left and right contact wire	Average of group
	Horizontal_avg	Horizontal position of the contact wire	Average of group
	Height_avg	Height of the contact wire	Average of group
Cant	Cant	Cant of the rail	Average of group
Speed	Speed_difference	Difference in speed between current and previous 100-meter section	Average over period
	Speed_field	Average speed of trains of one day	Average over period
	Speed_local	Maximum allowed speed for trains	Average over period
	Speed_measure	Speed measured by the measurement train	Average over period
Passed trains	Trains_goods	Number of trains transporting goods	Average over period
	Trains_total	Total number of goods and passenger trains	Average over period
	Trains_travel	Number of trains transporting passengers	Average over period
Transported tons	Tons_goods	Total amount of tons transported by goods trains	Average over period
	Tons_total	Total amount of tons transported by goods and passenger trains	Average over period
	Tons_travel	Total amount of tons transported by passenger trains	Average over period
Historical trend	Wear_rate_trend	Slope of the regression line within the datapoints of the training set	Regression over period
	R2_wear_rate_trend	R2 score of the regression line within the datapoints of the training set	Regression over period