Insights in Mining Science & Techonology
ISSN: 2689-4629

Research Article
Volume 2 Issue 4 - February 2021
DOI: 10.19080/IMST.2021.02.555594

Insights Min Sci technol

Variable Selection for Conveyor-Belt Mean Wear Rate Prediction

Joanna Z Sikorska*, Callum Webb, Nazim Khan and Melinda Hodkiewicz

University of Western Australia, Crawley Australia

Submission: January 13, 2021; **Published:** February 26, 2021

***Corresponding author**: Joanna Z Sikorska, University of Western Australia, Crawley Australia

Abstract

Rubber belt conveyors are an integral part of many mining and bulk haulage applications. The belts are designed to wear in-service and thus need to be replaced periodically. Currently, belt life management is done by measuring the thickness reduction and estimating remaining life once the belt has been operating for some time. This approach is not applicable for new conveyor belts. In this paper we present a process for building a model, and results thereof, to predict the life of new conveyor-belts based on a variety of design and operating parameters. The work analyses wear readings from operating conveyor belts and constructs linear regression models for predicting the wear-rate of out-of-set conveyors. Ultrasonic readings from 165 iron-ore rubber, steel-cored conveyor belts, installed in 95 conveyors were modelled against various belt and conveyor features. Nine linear regression models and/or modelling approaches were evaluated with nested cross-validation to measure prediction error. Results show that these models achieved an improvement in performance error of up to 46% compared to relying on the mean wear rate alone and indicate that the mean wear rate of heavy-duty conveyor belts is, at least in part, related to the compound variable of belt-speed squared divided by belt-capacity (length x width). Although the models described herein offer a significant improvement to the current best practice for estimating life of new conveyors, models were only able to account for a maximum of 75% of the observed wear behaviour and prediction error still remains in the same order as the data variance, suggesting additional variables will be required to improve end-of-life prognosis. This work also demonstrates, that for this dataset, in which two explanatory variables dominate, performance error is largely unaffected by variable selection approach. Finally, the work shows how widely used data science methods can be applied to commercially impactful equipment life prediction. The work can be easily replicated by conveyor owners to improve their own belt maintenance planning.

Keywords: Belt conveyor; Model selection; Linear regression; Predictive maintenance; Equipment reliability

Introduction

Rubber belts are often the largest component of a conveyor's lifecycle costs. In 2018, the global market for conveyor belts was worth US$5.68Billion and was forecast to reach US$6.9B by 2024 [1]. Therefore, to ensure cost effective production, it imperative that conveyor belt life be managed judiciously, and replacements be planned to minimize the operational impact. The traditional approach to estimating belt life is based on collecting thickness measurements at various locations along the width and length of the conveyor and then extrapolating these measurements to predict a replacement date. Thickness testing is usually performed using manual ultrasonic, eddy current or laser probes, which requires the conveyor to be shut down. The thickness data is then supplemented with staff experience and historical belt replacement frequency records to determine when belts should be changed. On-line condition monitoring techniques are improving, with real-time approaches being developed to reduce the need for conveyors to be removed from service for monitoring [1-3]; however due to their cost and reliability these are not yet widely employed.

The physical and operating differences between conveyor belts mean that remaining life prediction is specific to an individual conveyor belt. Unfortunately, due to the cost of downtime and/or measurement, not all conveyors can be measured regularly. Utilization of the conveyor as a function of actual throughput and running time is also not considered in typical thickness-versus-calendar time models, reducing the accuracy of predictions in operations where throughput is non-uniform over calendar time. The biggest limitation, however, is that wear rate can only be accurately estimated and extrapolated when enough thickness measurements for a belt have been undertaken, so this method cannot be used on newly installed units.

Conveyor Belt Wear

Conveyor belts are made of several layers of different material, selected for the application into which the belt will be installed. Long belts are often comprised of shorter belt sections spliced together. These splices can be replaced individually or as a set.

Figure 1 shows a simplified representation of the components of a conveyor system. Material is typically loaded onto the belt through a transfer chute that feeds the material onto the belt at a defined impingement angle. The belt is supported by freely rotating idlers, which are more closely spaced in the loading zone to absorb the energy of material impacting the belt. (Impact beds or other mechanisms are sometimes used in place of idlers in the loading zone). The belt is driven through one or more pulleys and discharges material at the head pulley. Belt tension is controlled through a take-up pulley. Belt cleaners or scrapers are used to prevent material from clinging to the belt.

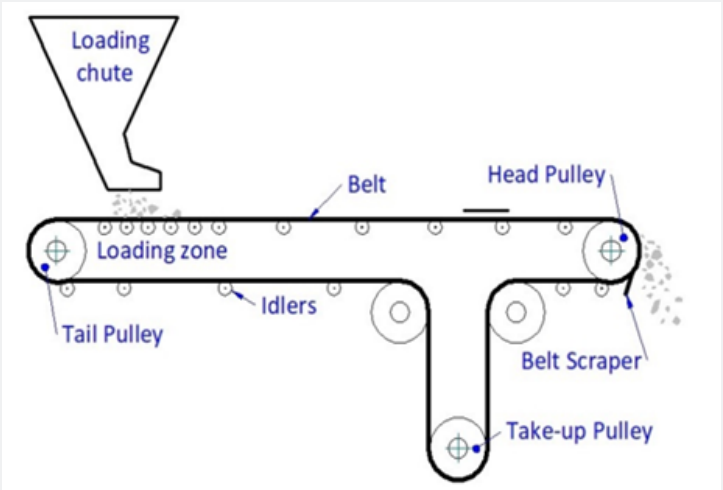

Figure 1: Belt conveyor parts.

Typically, the belt comprises an inner load bearing carcass of fabric or steel cords, surrounded by top and bottom rubber covers. The carcass is the principal structural component of the belt, providing tensile strength. The rubber covers protect the carcass from damage, with the top cover contacting the conveyed material and the bottom cover contacting pulleys, idlers and drive systems. Consequently, the top cover is usually thicker than the bottom cover, as it experiences more aggressive wear.

Wear of the rubber cover is predominantly due to abrasion by the material being transported, resulting in a reduction in thickness over time [4]. Unlike more rigid materials, when a smooth rubber surface is abraded, parallel ridge patterns emerge. Cracks at the roots of these ridges deepen, eventually leading to material tearing off from the bulk [5]. Both two and three body erosion are known to occur in conveyors. Observed wear rate is affected by which of the two mechanisms is dominant in a particular conveyor [6]. Hutchings derived an equation for conveyor wear that describes the abrasive wear as a function of the conveyor's operating parameters [7]:

$$t = \frac{kmv^2}{4\mu wl} \qquad (1)$$

where t is the thickness of rubber removed, m is the total mass of ore moved, v is the speed of the belt, w is the width of the belt, l is the length of the belt, μ is the coefficient of friction between the belt and the bed of ore and k is the specific wear rate for abrasion of rubber by the ore, which will differ depending on the type of wear that is occurring. A similar equation was published in [8] based on empirical studies by a conveyor belt manufacturer for the wear life in mega-tonnes (WMT):

$$W_{MT} = \frac{T_f \cdot \rho_{ore} \cdot A \cdot l \cdot t_w \cdot C_\theta}{30v^2} \qquad (2)$$

where T_f is an empirically derived tonnage factor based on how the conveyed material is distributed across the width of the conveyor (developed for material bulk density of 1000kg/m^3), ρore is the relative density of the material conveyed, A is the relative abrasion resistance of the belt cover material (=1 for ASTM grade M rubber), l is the conveyor belt length, t_w is the thickness of the wearable cover, C_θ is a correction factor for the angle of impingement, and v is the belt velocity. The authors note that a shortcoming of this formula is that it does not account for any differences in sharpness or abrasiveness of the material being conveyed. The work also did not present the data from which the various factors were developed or the methodology.

In practice wear rates are not always constant throughout the life of the belt due to changing operating conditions, bulk material properties or maintenance practices. For example, if material accumulates underneath the conveyor, or if belt cleaners are improperly installed, higher wear rates are usually observed because wear also occurs at the location of the belt cleaners or on the return journey (when the belt is on bottom of the conveyor and not carrying material) [9]. Wear of natural rubber is also affected by aging at elevated temperatures as the temperature

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

affects the rubber's elastic properties [6,10]. Storage also affects performance once in service due to increased risk of cracking [11].

Belt working life is therefore regarded as a function of:

a) The belt's material properties (tensile strength, Shore A hardness, shear strength) which collectively provide a certain resistance to abrasion, shearing, impacts, heat and ageing

b) The conveyor's design (e.g. length, width, speed, height and design on loading chute) that affect how the load impinges the best and is distributed across the width of the conveyor and how often loading occurs

c) The physical properties of the material being transported (size, density, shape, hardness) [4,6,12-15]. Unfortunately, in practice, very few of these parameters are measured or readily available for analysts.

During service, belts are also subjected to damage from impacts, scratching and gouging, usually occurring at time of belt loading [16]. Although not specifically the subject of this study, thickness measurements collected from operating conveyors may also reflect these events. For example, material can become trapped between the conveyor sealing system (skirts) and the belt, causing wear in grooves along the length of the belt underneath the skirts. Pitting from improper loading is also a common damage mechanism, for which predictive models based on laboratory studies have recently been developed [11, 17-19]. These studies identified that the likelihood and severity of pitting damage increased with drop height and shape of the dropped material and was also affected by whether the belt was new, refurbished or had been stored for some time prior to use.

Although conveyors are rated based on their performance in standardized tests for abrasion resistance and tensile strength, there is little correlation between standard and custom abrasion tests and field performance "due, in part, to the varying nature of surface conditions" [20]. To ensure adequate life for an application, manufacturers offer belts and covers in numerous grades, each of which has a unique combination of wear and impact resistant properties, the specific details of which are not disclosed. Belts are considered 'worn-out' when the top rubber covering the steel or mesh core has reduced to a nominal thickness (e.g. 3mm) at any point across the width of the belt. Wear is usually the greatest near the center of the belt as this is corresponds to maximum load. The belt is either replaced or refurbished by re-coating the worn top of the belt with new rubber.

The current approach to planning belt maintenance is based on extrapolating thickness wear (specifically, the rate of top cover wear) per unit of time. This process is repeated at equally spaced intervals across the belt width, producing a picture of the top cover cross section as shown in Figure 2. Measurements may also be taken at multiple places along the belt's length, often coinciding with different splices. Unfortunately, it is not always possible for the belt thickness to be measured in the same place every time. This complicates analysis when the belt is made of multiple splices that may have been installed at different times.

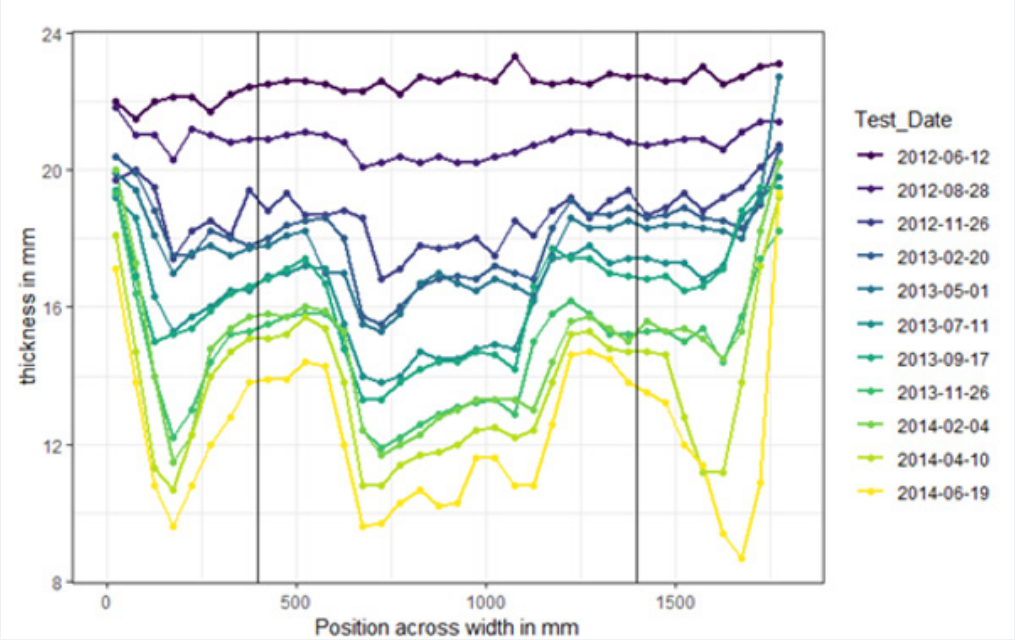

Figure 2: Typical wear pattern (conveyor C_038 from dataset used herein). Vertical lines are shown at 400mm from each edge.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

Very few studies on predicting conveyor belt wear have been published, with most studies focusing on impact damage. In [21] the authors used multiple linear regression (MLR) to model economic life of 18 conveyor belts against a number of different variables and found that length and tonnage were statistically significant in estimating the economic life of conveyors. The authors were able to account for 85.3% of the life variability with only length and tonnage. However, this study only used one observation per belt and provided little detail on how the processing was achieved. Belts were also very short and narrow; the longest belt in the dataset was only 196m in length, and the widest only measured 1400mm across. Crucially, the model was not tested on unseen data, so its generalization ability is unknown.

The lack of relevant studies is actually a much broader issue facing the field of engineering prognostics and health management [22,23] (PHM). To illustrate the scale of the problem, we analysed how researchers were validating PHM models. Using the broad search terms of '+data-driven +prognostics +model' we categorized the methods used by researchers to validate prognostic models published in any Science Direct accessible journals during 2018. Search results were then further filtered using '+engineering' to remove health related models and then further assessed manually to remove irrelevant articles, such as reviews or concept-only papers; in total 146 papers were removed, leaving 273 papers that described a model relevant to engineering prognostics and health management. The analysis of these 273 papers showing how each paper's model was validated are provided in Table 1. This data shows that most models are only tested using either.

a) simulated or numerical datasets

b) a single illustrative and highly simplified case study

c) using measurements collected under laboratory conditions.

Table 1: Model Validation Methods of Papers Published in Science Direct Journals During 2018.

Method of model validation	No of papers	Percentage of total
Simulated data only	61	22%
Simulated data & experimental data	15	5%
Simulated data & illustrative case	4	1%
Simulated data & industry case study	4	1%
Simulated data & complex industry dataset	1	0%
Experimental data only	122	45%
Experimental data & illustrative case	1	0%
Experimental data & industry case study	2	1%
Experimental data & complex industry dataset	1	0%
Illustrative case only	28	10%
Industry case only	17	6%
Industry dataset only	15	5%
Unclear where data originated	2	1%
	273	100%

In only 11% of cases were models tested using data collected from equipment operating in industrial plants (shown as industry case study or industry dataset).

Data Set Description

Unless otherwise stated, all data extraction, processing and visualization for this paper was undertaken using the software language R [24]. Existing R packages and functions were utilized wherever possible [25-31].

Data for this study was collected from 165 heavy-duty steel-core belts on 95 troughed conveyors at two different bulk material handling sites in the north-west of Western Australia and spans a period of 8 years. These conveyors transferred between 0.02 and 0.8 mega-tonnes of iron ore per week (with an average of 0.4 mega tonnes/week). The final data set contains one record for each conveyor belt, each of which contains 2 alternative dependent variables modelled in this work:

a) MeanWearRate_mm/week

b) MeanWearRate_mm/MT

These variables were derived from 41,324 ultrasonic thickness readings and 2,354,127 conveyor operations. Each record also contains nine explanatory variables: four numerical variables (belt speed, belt strength, drop height, % fines), one categorical variable (conveyor duty), three transformed variables (1/belt width, 1/belt length, belt-speed^2) and two compound variables (loading frequency and V1 as defined later in this paper).

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

Thickness Data

Ultrasonic thickness measurements were collected periodically (but at irregular intervals) by specialist contractors from each conveyor whilst it was shut down. Readings were taken at one or more locations along the length of the conveyor and at 50mm intervals across the width of the conveyor. These measurements were stored by the contractor in Excel© spreadsheets; in principal each conveyor belt had its own Excel© spreadsheet which was maintained until that belt was replaced. 78557 thickness records were extracted from 239 spreadsheets into a single table and anonymized using R. Thicknesses at width positions within 400mm of each edge were removed from the dataset, as excessive wear in this region indicates issues with the skirts and not steady state wear. This is illustrated in Figure 2, which shows the wear pattern over time for one conveyor in the dataset. Thickness readings that preceded tonnage records for that conveyor could not be used to determine a wear rate per MT and were also removed from the dataset. Similarly, thicknesses associated with 8 conveyor belts that had less than 3 thickness readings with valid tonnage values were also removed. The resulting usable set of thickness readings contained 41324 records.

Identifying Spliced belts (Pools)

Visualizing the thickness data for each conveyor as a function of the date on which the thickness reading was collected (referred to herein as reading date) revealed a problem with the original data collection process, as illustrated in Figure 3. In this figure, the degradation of only three conveyor belts can be identified, despite the data originating from 9 separate reports.

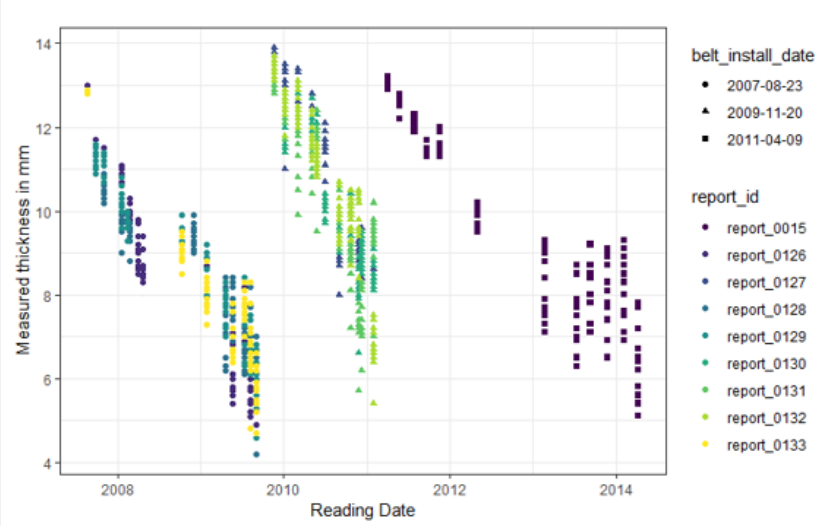

Figure 3: Data from conveyor c_053 showing multiple overlapping reports.

When all conveyors were similarly analyzed, 79 spreadsheets were found to contain overlapping readings that could be clearly attributed to only 18 separate conveyor belts. These were identified as separate sets of measurements and not merely duplicates because the thickness measurements were different. We assumed that readings had been collected from the same belt at multiple longitudinal locations due to the presence of splices, or because the belt was particularly long or operationally critical. (There was no way of confirming this at the time of modelling.)

To ensure that our final modelling set contained only one record for each conveyor belt, the belt-install-date was used to pool the data from these concurrent sets of readings, to then determine a single set of metrics for each belt. In this paper, we refer to all measurements that have been collected from the same conveyor concurrently as a pool. It is not referred to as a belt simply because the underlying data referred to each report as a separate belt, rather than a splice or part of a belt. By adding the term 'pool' we clearly designate this as aggregation that we have added to the data.

Raw Tonnage Data

Tonnages were measured by weightometers positioned throughout the supply chain. It is not known how often these instruments were re-calibrated. Tonnage data was extracted using R Code from a spreadsheet containing 475191 rows, which had been generated from the Company's material tracking system. This spreadsheet contained a list of all equipment movements over an 8-year period. Records were not standardized, so multiple conveyors could be associated with a single movement. When this was standardized and relevant conveyor data extracted, 2354127 movements were identified that related to the conveyors in our study.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

Only movements with the following characteristics were included in the dataset: movement duration ≥ 0, start-date-time ≠ NULL, end-date-time ≠ NULL, 300≤ tonnes per hour ≤ 15000, tonnes >0, type of product ≠ NULL. These filters were necessary to ensure that subsequent analysis could be performed and to ensure that only records relevant to the study were included in the dataset.

Equipment Data

All readily available conveyor and belt design parameters were initially considered for use as explanatory variables. Belt width in mm, belt length in m, belt tensile strength in kN and manufacturer's belt grade were extracted from Company records. The drop height of the loading chute (measured as the vertical distance from the center of the head pulley in the feeding conveyor to the center of the tail pulley on the conveyor of interest) was manually extracted from design drawings. 5 belts were removed from the dataset at this step due to one or more missing explanatory variables.

Conveyors were also classified into 8 duty categories. When belts are designed, it is usually assumed that equipment belonging to the same duty-class will have similar operating requirements. Some conveyor duties had only a handful of records, and were similar in design to other larger groups, so these were merged into the larger groups. Specifically, Car-dumper conveyors were combined with Transfer conveyors and Wharf and Tunnel duties were combined with Yard conveyors. The initial and final spread of data across duties is provided in Table 2.

Table 2: Conveyor distribution amongst duty category.

Duty	Number of Conveyors	Number of Belts	Belts with multiple splices
Car Dumper	4	4	0
Transfer	52	56	4
CarDumper_Transfer	56	60	4
Wharf	6	13	6
Tunnel	1	2	1
Yard	14	18	6
Wharf_Tunnel_Yard	21	33	13
Reclaimer	5	23	1
Shiploader	6	27	0
Stacker	7	22	0
	95	165	18

Calculated Variables

Wear Rate

To ensure that modelling both dependent variables (wear-rate/MT and wear-rate/wk) could be undertaken using the same dataset, thickness readings that preceded the earliest conveyor movement were removed from the dataset. For remaining readings, all movements of both lump and fines between the earliest reading-date for that conveyor belt and the reading-date, were extracted and summed independently. Fines-tonnes and lump-tonnes were then added together to get a total value of tonnage for each reading. Total time-in-service (since the earliest reading) was also calculated for each reading and is a function of calendar time between the two dates (rather than operating time). Wear rates at each width-position (location across the width of the belt) were determined by extracting all readings for a pool (i.e. belt) and width position, along with their associated time-in-service and tonnes (Table 3). These values were then modelled separately using linear regression to get a wear-rate/MT (mega-tonnes) and wear-rate/week for each pool and width combination. Table 4 and Figure 4 summarize the wear rates and regression coefficients for all 3229 pool-width combinations.

Table 3: Wear Rate.

	(a) thickness ~ time	(b) thickness ~ MT
Mean R^2	0.867	0.872
Std deviation of R^2	0.165	0.164
% of fits with R^2<= 0.5	4.06%	3.90%
% of fits with R^2>=0.8	80.70%	82.00%

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

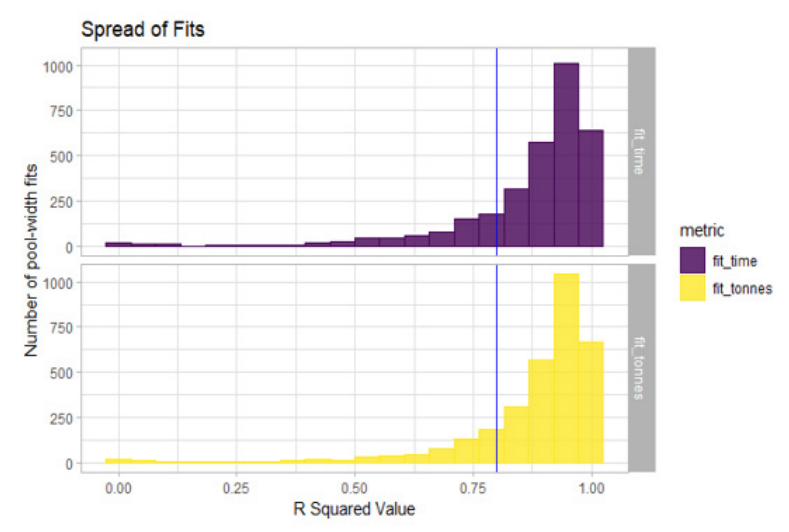

Figure 4: Spread of 3229 regression coefficients of determination (R^2).

As can be seen in Figure 4 the coefficients of determination (R^2) for most of the models exceeded 0.8, indicating that linear models were adequate for estimating lifetime wear rate. The linear fits between wear and tonnes were slightly better than between wear and time-in-service. 30

The mean wear rates for each metric and pool were then determined by averaging the wear rates for all width positions:

$$\mu_{metric,pool} = \frac{\sum_1^{n_{w,pool}} m_{w,pool}}{n_{w,pool}} \quad (3)$$

where $n_{w,pool}$ is the number of width positions in a particular pool and $m_{w,pool}$ are the gradients of the linear regression curves for each width position in the pool. These became the response variables for each record in the modelling dataset. The distribution of all wear rates for all width positions is provided in Figure 5. The corresponding mean and variance of each metric are:

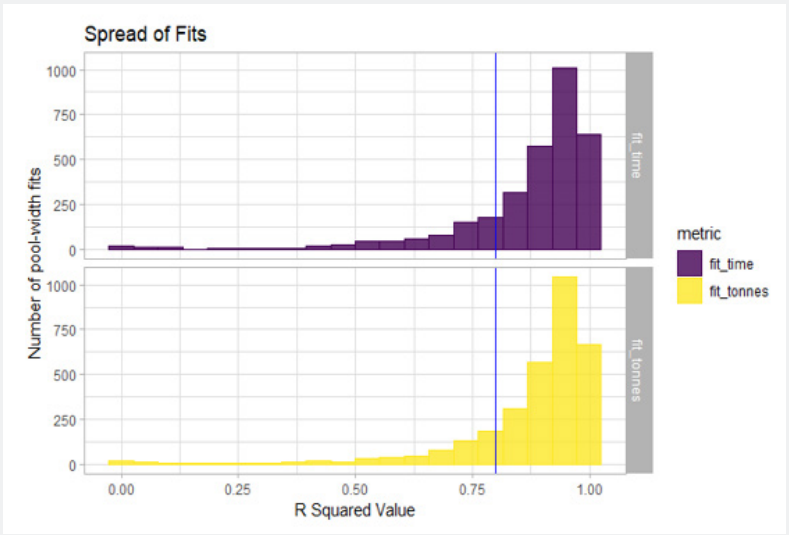

Figure 5: Spread of mean wear rates (N=165).

Calculated Explanatory Variables

In addition to the variables extracted directly from the data (belt-width, belt-length, belt-strength, belt-speed, drop-height, percentage-fines, belt-grade), five derived variables were determined for each conveyor belt record. Loading frequency was calculated as:

$$load\_freq = \frac{belt\_speed}{belt\_length} \quad (4)$$

A second variable, simply termed V1 , was obtained from the equation (1) derived by Hutchings [7]:

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

$$V1 = \frac{v^2}{4wl} \quad (5)$$

where v is the belt-speed, l is the belt-length and w is the belt-width (in m). The more comprehensive equation presented in (2) from [8] could not be used because many of the required design variables were not available. As (5) contains transformations of other variables, these transformations were also included in the models to determine whether they were individually influential. Namely:

$$width\_inv = \frac{1}{belt\_width} \quad (6)$$

$$length\_inv = \frac{1}{belt\_length} \quad (7)$$

$$belt\_speed\_sqr = belt\_speed^2 \quad (8)$$

Categorical Variables

As the conveyor-duty variable is a categorical variable with five discrete levels, it was converted automatically by R into binary variables using an effects coding scheme called treatment coding (also called dummy coding). In this scheme, a categorical variable with k levels is described with k−1 coded 'dummy' variables. These variables are included in the model, and are equal to 0 or 1, depending on the level of the categorical variable considered [32].

Belt-grade was initially considered as another categorical variable, but the data was deemed to be insufficiently trustworthy for modelling. The belt-grade had been assigned by plant staff retrospectively, based on the memory of key engineers, rather than having been accurately recorded from the markings on the belt at the time it was installed. In a significant number of records (>50%), this assignment did not match with other information that independently implied a grade.

Exploratory Data Analysis

Figure 6 shows the correlation between all numerical explanatory variables. As expected, loading-frequency and V1 have a high inverse correlation with belt-length. Loading-frequency is also very highly correlated with the new variable V1, which again is unsurprising as V1 can be rewritten as a function of loading-frequency.

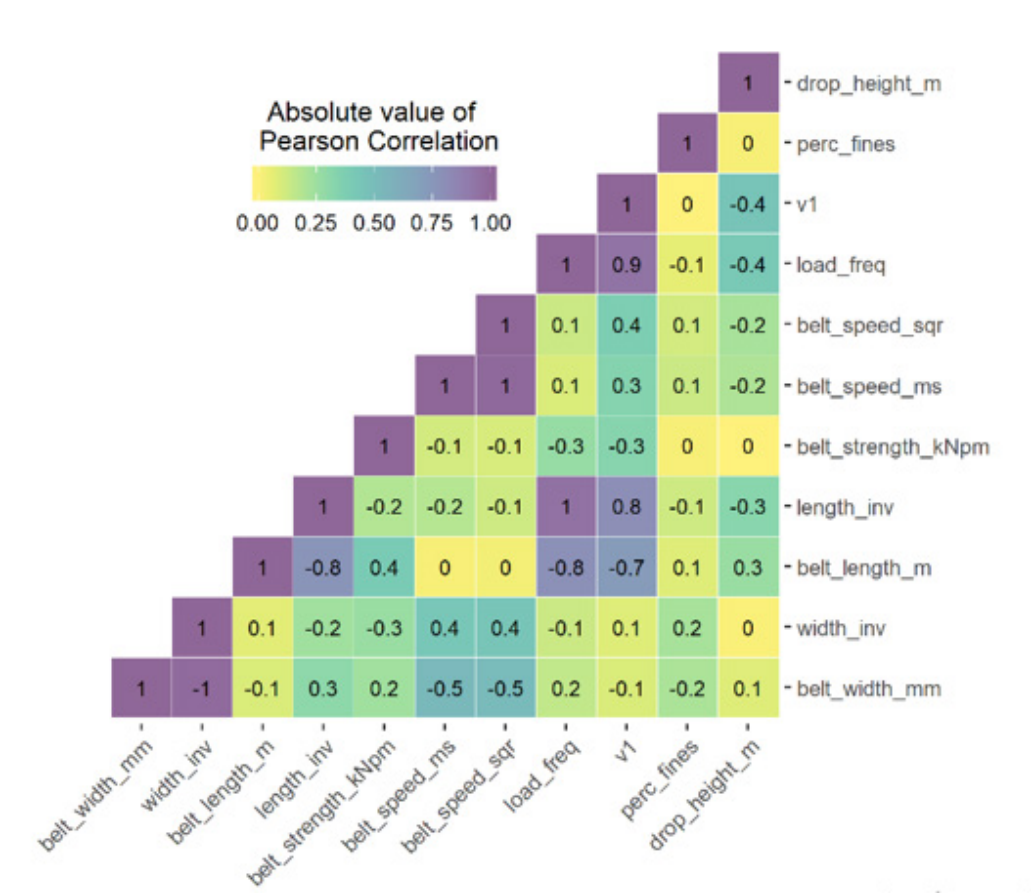

Figure 6: Correlations between numerical explanatory variables.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

As conveyor-duty is a categorical variable, its correlation with other explanatory variables cannot be evaluated directly. Instead, the relationships between conveyor-duty and other 6 explanatory variables are shown graphically in Figure 7.

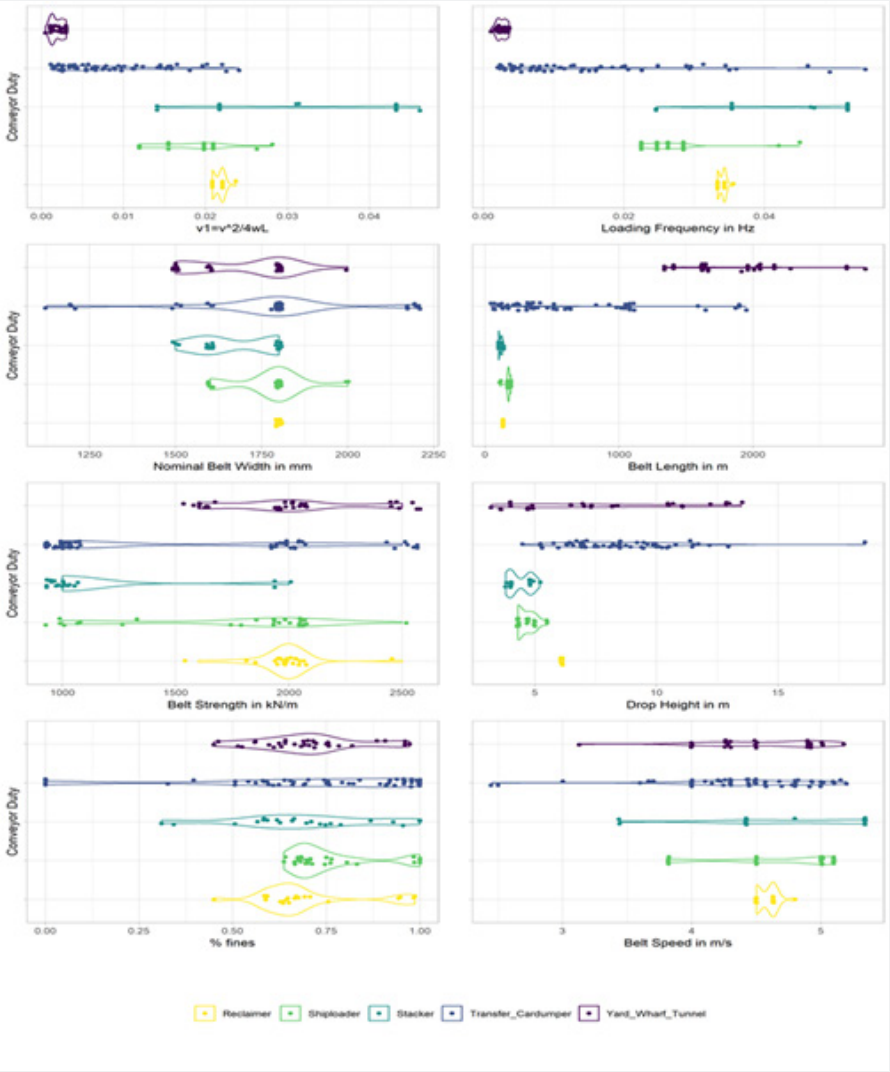

Figure 7: Distribution of variables across conveyor-duties.

Several observations about the relationships between explanatory variables can be made from these plots that have implications on subsequent predictive modelling:

a) V1 is very highly correlated (0.94) with loading-frequency as would be expected. The strength of the correlation is probably due to the limited range of speed and width values in the dataset, resulting in a much stronger relationship with belt-length (the other component of loading-frequency). This is supported by the strong inverse correlation between V1 and belt-length (-0.72).

b) Transfer-CarDumper and Yard-Wharf-Tunnel duties have the widest range of lengths, widths and speeds, which may be due to these having the greatest numbers of conveyors, but it may also be the case that these conveyor duties are in fact very diverse in the range of operational tasks that they perform. In particular, the Yard_Wharf_Tunnel group has no short conveyors (<1000m) and is the only group with conveyors that are over 2000m long. Stacker, shiploader and reclaimer belts have very similar, short lengths (<500m).

c) V1 discriminates between different conveyor-duty better than other continuous variables, indicating that V1 may be better correlated with conveyor-duty than other parameters.

From these plots, it was concluded that both loading-frequency and V1 would probably not be required in the same model due to their very high correlation (0.94). Subsequently,

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

when computational problems emerged due to the simultaneously inclusion of both variables (specifically a failure to converge on a best model and high sensitivity to changes in the dataset division during cross validation), load-frequency was removed as a variable.

a) As per convention, when the inverse of a variable was included in the model, the original variable was omitted, whilst the squared variable and its original were always retained.

Model Development

Cross-validation for Prediction Models

It is widely accepted that predictive and inferential models trained on data, should be tested on data that has not been used to build the model [33,34]. When the dataset is sufficiently large, it is usually best to divide the available data into separate learning (a.k.a. training), validation and test sets. The test set is only utilized to estimate out-of-sample error after modelling is complete. Often however, splitting the data into three groups leaves insufficient data for training. For smaller datasets alternative techniques are required. Resampling techniques such as k-fold cross-validation (KFCV) [35], leave-one-out cross-validation (LOOCV is an extreme case of k-fold cross-validation with k equal to the number of samples) and bootstrapping [36] are now widely used for this application [34,37]. These approaches use the same dataset multiple times for all tasks and thus facilitate using smaller datasets for predictive modelling.

This work uses k-fold-cross-validation to determine the performance error. KFCV involves splitting the data into k folds. On each iteration, one of the folds is set aside as the test-set (a.k.a. hold out set), whilst the remaining k-1 folds are used as a learning set. Once the model has been trained, the loss function is calculated using the test-set.

The loss associated with any model can be considered a combination of three main types of error. Bias error is related to how much the model deviates from a true model of the data (assuming one exists). Variance error is how much the predictions vary around their mean. The final component of error is how much variation occurs in the actual measured data (which is a function of measurement errors and noise) and cannot be affected by the modelling process [38]. Selecting an optimal prediction model therefore requires finding a balance between the two types of errors, which is commonly known as the bias-variance trade-off [34,39]. Adding more variables to a model can increase its fit to the training data, but this eventually reduces its ability to predict or infer from new information. Simpler models may predict better but will not be as useful for exploratory purposes. The benefit of cross-validation for prediction, is that it measures overall error, and not just one of the two components (bias or variance).

Selecting the optimal number of folds is an important decision in cross-validation because it affects the relative amounts of bias and variance error being reported by the performance estimate. Variance in CV models generally decreases as the number of splits decrease, due to the larger test set. It is also indirectly affected by the variability of the predictor and thus it is sensitive to how the dataset is split into learning and test folds. Whether this results in the lowest overall error, depends on the size of the bias error. For predictive models, the optimal number of folds is often reported as being between 5 and 10 [34,40], but this is not always true and, in part, depends on the signal-to-noise ratio of the data [41]. In fact, the case for a recommended number of folds that applies to all applications is far from conclusive [42-44]. In practice, determining the optimal number of folds for each modelling application is often an iterative process. For a more detailed discussion, the reader is referred to [43,45].

For k-fold cross-validation, repeating the entire process multiple times, called repeated- or nested-CV, has been reported to reduce variance error and decrease error variance for regression and classification [43, 46,47] at the cost of additional computation time. In addition to selecting the number of folds, and if desired the number of repeats, it is also important to decide how the data will be split into folds. The most common approach within the cross-validation literature is to randomly split data into folds. However, in some cases this is not advisable as it results in highly correlated data points being contained in both the training and validation sets, thus resulting in an underestimate of the error. Stratified data splitting aims to ensure that each fold contains approximately the same proportion of different 'types' of data. This is advocated by some authors [40,46,48], whilst others report no improvement from applying stratified splitting in regression [49]. It is important to note that cross-validation estimates the average loss function; it is not an estimate of the error that can be expected for a specific prediction [35,50,51].

Model Selection

Model- (more specifically variable-) selection is known to be a potential source of error during model building (by selecting a sub-optimal model or measuring its error inaccurately [52-55]. The biggest factor affecting the choice of variables for this work was simply what data could be found to describe these conveyors and their belts. Therefore, it was anticipated that a more formal variable selection approach would be required to determine which of these was significant to wear modelling. To implement the measurement of cross-validation error correctly, model selection needed to be performed within each fold. This necessitated the use of an automated model selection algorithm.

There is significant dissension in the published literature as to the suitability of some commonly used model selection algorithms [56-58]. Rather than subscribe to any one side of the ongoing debate, we instead trialed several algorithms to determine if model selection methodology was important in this application.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

Several linear models were analyzed without any automated variable selection. These were implemented using the 'lm' method of the train function from the caret package [25]. In these cases, the train function also implemented the repeated cross-validation process, with the settings described below. For brevity, this paper only discusses the following list of basic linear models:

a) Null hypothesis - linear model without explanatory variables.

b) Full linear model with all explanatory variables (including compound variables and transformations)

c) Model without any compound variables (simple variables only, including single variable transformations)

d) Model with only conveyor-duty as a variable

e) Model with only V1 as a variable

f) Model with both conveyor-duty and V1 as variables

Three widely used model selection approaches were then tested:

g) Stepwise model selection: This was performed using the Step AIC function from the MASS package [24]. The selected performance metric was AIC (Akaike Information Criteria) [59-61]. This algorithm combines forward and backward selection into one algorithm. Stepwise selection has been criticized extensively by the data science community following the criticism in [55], but it is still widely utilized for linear regression due to its simplicity and computational efficiency and is regularly included in textbooks on linear regression [34]. It has been found to perform adequately when a few covariates have a strong relationship with the outcome variable [56].

h) Model averaging was undertaken using the glmulti package [53]. Although we call this a model selection method, this approach in fact avoids selecting a single model. Instead, the glmulti function analyzes all possible models and ranks them by their AIC value. (Alternate metrics are available but were not used in this study.) This AIC value is used to determine a ΔAIC, which is defined as the difference between a candidate model AIC and the minimum AIC in the set: $\Delta AIC = |AIC - AICmin|$. Candidates that differ from the best model (i.e. the model with the lowest AIC weight) by a ΔAIC>10 are considered so unlikely of being the top model, that they are often excluded from further analysis [53]. Conversely, models that have ΔAIC≤ 2 are considered as strong contenders for the best model. The process then determines an Akaike weighting for each candidate that is akin to a likelihood that a candidate is the 'best' model. When a prediction is required, the top n candidate models are multiplied by their respective Akaike weightings and summed. The number of candidate models (n) to include in a solution is defined by the user when calling the function. In this analysis we used n=200, to ensure that all models with ΔAIC<10 would always be included in the final aggregate model.

i) Lasso was implemented using the gglasso package as per [31]. This package uses the glmnet package [30] to perform the Lasso but adds the ability to scale and select factor variables as a set, rather than removing individual dummy variables. Lasso is often presented as the preferred approach to model selection instead of stepwise selection [56-58], particularly when the explanatory variables do not have a strong relationship with the predictor. Lasso is a method that forces the sum of the absolute value of the regression coefficients to be less than a fixed value *(λ)*, and it does so by setting coefficients of less important variables to zero. Lasso can be problematic when covariates are not independent, in which case a variation of Lasso called the Elastic Net may be required [62]. In our work, we also elected to have the algorithm use a further internal cross-validation process to select its optimal value for λ, chosen to minimize a selected metric (in our case the Mean Square Error). This was done by using the cv. gglasso function to implement Lasso within each fold. These Lasso functions can return both the best model, which uses the value of lambda that minimizes the MSE, and an optimal model. The optimal model uses a different value of lambda that returns the simplest model that is still within one standard deviation of MSE from the best model. We show both models in this work to help describe the difference in performance resulting from the addition of lesser predictors.

Prediction error associated with all models was calculated using a repeated k-fold cross-validation (KFCV) process. Unless otherwise reported, we used 10 folds and 50 repeats to stabilize results and provide higher precision, as in [43,46,47]. The overall predicted error was then calculated for each model option as detailed in the following section. All numerical independent variables were standardized to have a mean of 0 and a standard deviation of 1 within each fold (or for the entire dataset prior in the case of model refitting).

To minimize expected bias caused by assigning different belts from the same conveyor to both training and test sets in the same fold, we used stratified data splitting of folds, rather than random splitting. This process verified that each conveyor only appeared in a single fold (so that the same conveyor could never appear in a training set and the associated test set) and that each categorical variable appeared in at least two folds to ensure that every level of the variable was always included in training. The process by which this was implemented is shown in Figure 8. A side effect of the implementation was that fold sizes were not consistent, due to the varying number of data records for different conveyors. Test-set sizes varied typically between 9 and 31 records, with a median size of 16 records and mean of 16.5 records.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

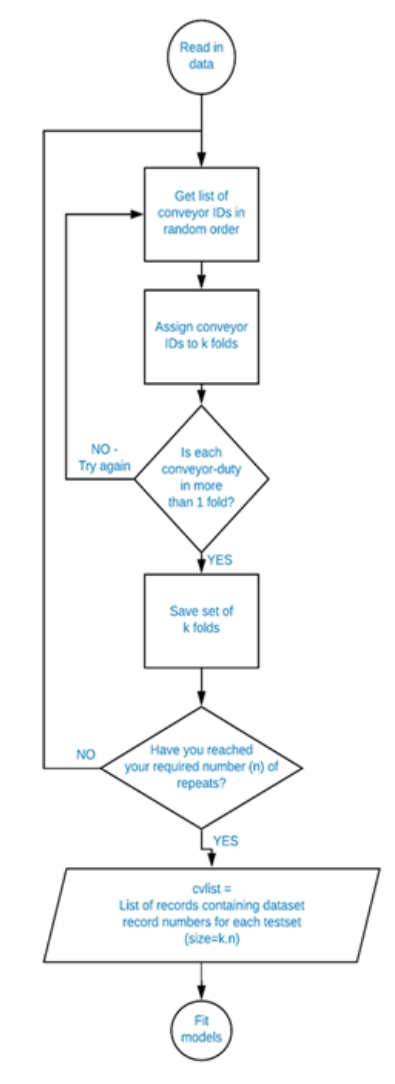

Figure 8: Stratifying folds.

Repeated cross-validation was implemented using standard caret functions [25]; additional methods were added to manage the glmulti and gglasso models within caret train functions. To determine the sensitivity of performance results to the number of folds, number of repeats and stratified versus random sampling of folds, select models were tested at range of values for each programming parameter to determine the effect on the loss function. As no unexpected findings eventuated from this analysis, results are omitted for the sake of brevity.

Model performance evaluation metrics

The process used in this paper to implement cross validation and evaluate model performance is summarized in Figure 9 and Figure 8. The following paragraphs describe how the error is calculated as well as describing other metrics used in this work.

KFCV Prediction error

Let $\varkappa$: {1, . . , N} → {1,...,K} be an indexing function that indicates the partition to which an observation, i, is allocated and let f^_(-k) (x) be the fitted regression model, computed with the kth partition removed. The cross-validation estimate of prediction error for a single cross-validation iteration is:

$$CV(\hat{f}) = \frac{1}{N_{-k}} \Sigma_{i=1}^{N_{-k}} L\left(y(i), \hat{f}_{-k}(x_i)\right) \quad (9)$$

where N-k is the total number of data points being analyzed [30]. In regression, the most common loss function is the mean square error (or its square root):

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

$$L_n = 1/n\Sigma_{i=1}^{n}\left(\hat{y}(i)-y(i)\right)^2 \quad (10)$$

where $\hat{y}(i)$ is calculated model output, y(i) is the actual measured value and n is the total number of points in the set. In KFCV, the process is repeated k times, each time using a different fold as the test set; this is then averaged to get an overall loss [34]:

$$L_{k,n} = \frac{1}{k}\cdot\Sigma_{j=1}^{K} L_n \quad (11)$$

When the cross-validation process is being repeated, the average is calculated over all folds and repeats.

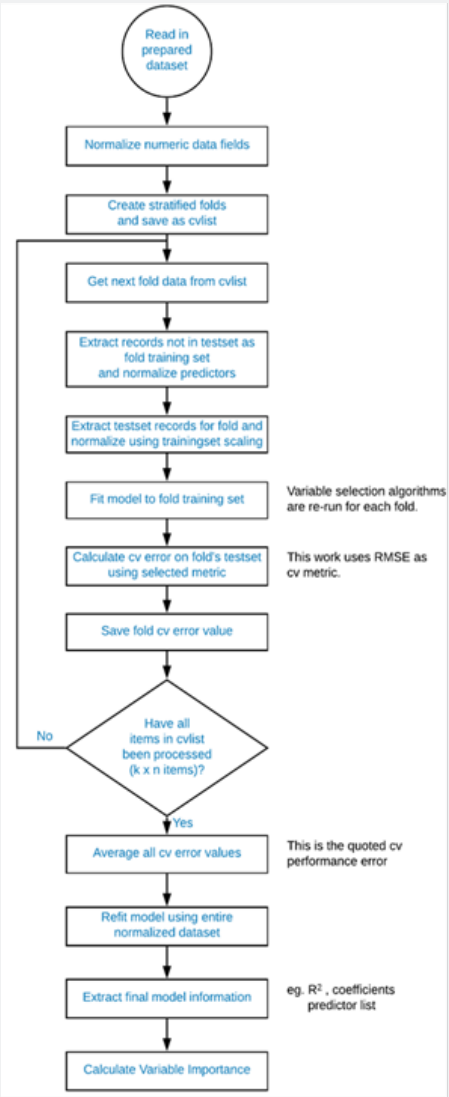

Figure 9: Modelling process used in this paper.

Model Fit

Other features used in the presentation of results are the R^2 and adjusted-R^2 value. These are used to quantify how much variation in the wear metric are described by the model's explanatory variables and are only calculated once modelling has been completed. Unlike the R2 value, which will always increase as the number of explanatory variables increase, the adjusted-R2 value penalizes additional variables that do not improve the performance of the model sufficiently and can therefore be used to compare models with differing numbers of parameters [63,64].

The R^2 and adjusted-R^2 are given by:

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}\left(\hat{y}(i)-y(i)\right)^2}{\Sigma_{i=1}^{n}\left(y(i)-\bar{y}\right)^2} \quad (12)$$

$$Adj.R^2 = 1 - \left[(1-R^2)\cdot\frac{n-1}{n-p-1}\right] \quad (13)$$

where n is the sample size, p is the number of explanatory variables in the model.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

Variable Importance

Variable importance methods measure the relative influence of variables on the predicted outcome. Unfortunately, the method by which this is calculated varies widely between model algorithms. For example, linear models most often use the absolute value of the t-statistic for each model parameter as the measure of variable importance, whilst modern algorithms such as random forests and boosted trees have methods intrinsic to the algorithms themselves. Collectively, these are known as model-based approaches. Averaging methods do not have a specific model-based approach and so instead the glmulti package simply reports the number of models in a set of best models, in which each variable appears; the more frequent the variable in the set of best models, the higher its 'variable importance'. Lasso methods typically report the absolute value of the coefficients in the final tuned model, and therefore require that all variables be normalized to have the same mean and standard deviation prior to calculating variable importance; this can be difficult to implement for categorical variables as the dummy variables need to be normalized as a group.

Model-free approaches are also useful, either when models do not have intrinsic methods of calculating the assessing the relative importance of predictors, or when an alternative method is desired because the standard approach fails to consider the relationships between predictors (as in the case for the linear method using the t-statistic mentioned above). The most common model-free approach is the filter method [37], but this also does not take into account either the final fitted model, nor the interactions between predictors [65].

An alternative model-based approach, proposed in [65] uses partial dependence plots (PDPs) to estimate the functional relationship between each predictor and the outcome of interest, whilst accounting for the average effect of the other predictors in the model. They can also rank and score each predictor in terms of its relative influence on the outcome. Although PDPs are invaluable for describing relationships in complex models, they can be misleading when predictors are correlated. A refinement to PDPs that overcomes this weakness was presented in [66] and uses the concept of individual condition expectation curves (ICE). The ICE method estimates the relationship between the response and a predictor of interest for each observation, which are then averaged to determine a PDP. They are particularly well suited to datasets with interaction effects between predictors [65]. In this work, we use the method proposed in [65] and available using the vip R package to calculate the variable importance of all methods.

Results

Each mean wear metric was modelled using the various model selection approaches, within a repeated-CV framework that used 10 stratified folds and 50 repeats. The results are shown in Table 4, which also shows the improvement of each model from the null hypothesis. This shows that all models achieve a measurable improvement in performance error over the null hypothesis. Once the performance error of each modelling approach was determined, each model was refit to the entire dataset. Results for throughput and time-based models are presented in Table 5 and Table 6.

Table 4: Summary of raw data linear fits.

		mean	var
Throughput metric	Mean_mm/MT:	0.277	0.0582
Time-based metric	Mean_mm/wk:	0.1	0.0065

Table 5: Throughput Model Results.

	Model selection approach	Final Model	R^2	No of terms	Adjusted R^2
1	Null Hypothesis	rate ~ 1	0	0	NA
2	Full model (All variables)	rate ~ 1 + conveyor-duty + V1 + belt-strength + drop-height + perc-fines + belt-speed + belt-speed2 + 1/belt-length + 1/belt-width	0.7762	13	0.7585
3	Simple variables only	rate ~ 1 + conveyor-duty + belt-strength + drop-height + perc-fines + belt-speed + belt-length + belt-width	0.7045	11	0.6854
4	Simple transformed variables	rate ~ 1 + conveyor-duty + belt-strength + drop-height + perc-fines + belt-speed + 1/belt-length + 1/belt-width	0.7638	11	0.7484
5	Conveyor-duty only	rate ~ 1 + conveyor-duty	0.6437	5	0.6348
6	V1 only	rate ~ 1 + V1	0.6261	2	0.6238

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

7	V1 + conveyor-duty	rate ~ 1 + V1 + conveyor-duty	0.7017	6	0.6923
8	Stepwise Model Selection	rate ~ 1 + conveyor-duty + V1 + belt-speed + belt-speed2 + 1/belt-width	0.7716	9	0.7599
9	Averaged Models (ΔAIC≤10)	NA - Statistics are for total weighted model using the best 100 models.	0.7744	13	0.7566
10	Lasso selection – best model	rate ~ 1 + conveyor-duty + V1 + belt-strength + drop-height + perc-fines + belt-speed + 1/belt-length + 1/belt-width	0.7714	13	0.7534
11	Lasso selection – optimal model	rate ~ 1 + V1 + 1/belt-length + 1/belt-width	0.6772	4	0.6712

Table 6: Time-based wear results.

Model selection approach	Final Model	R^2	No. of terms	Adjusted R^2
Null Hypothesis	rate ~ 1	0	0	NA
Full model (All variables)	rate ~ 1 + conveyor-duty + V1 + belt-strength + drop-height + perc-fines + belt-speed + belt-speed2 + 1/belt-length + 1/belt-width	0.7506	13	0.7309
Simple variables only	rate ~ 1 + conveyor-duty + belt-strength + drop-height + perc-fines + belt-speed + belt-length + belt-width	0.7255	11	0.7076
Simple transformed variables	rate ~ 1 + conveyor-duty + belt-strength + drop-height + perc-fines + belt-speed + 1/belt-length + 1/belt-width	0.7369	11	0.7199
Conveyor-duty only	rate ~ 1 + conveyor-duty	0.6789	5	0.6709
V1 only	rate ~ 1 + V1	0.5254	2	0.5225
V1 + conveyor-duty	rate ~ 1 + V1 + conveyor-duty	0.7302	6	0.7218
Stepwise Model Selection	rate ~ 1+ conveyor-duty + V1 + belt-strength + belt-speed^2	0.746	8	0.7346
Averaged Models (ΔAIC≤10)	NA - Statistics are for total weighted model using all models with ΔAIC≤10.	0.7487	13	0.7289
Lasso selection – best model	rate ~ 1 + conveyor-duty + V1 + belt-strength + drop-height + perc-fines + belt-speed + belt-speed^2 + 1/belt-width	0.7487	12	0.7307
Lasso selection – optimal model	rate ~ 1 + conveyor-duty-stacker + conveyor-duty-transfer + conveyor-duty-yard + V1 + belt-strength + drop-height + belt-speed +1/belt-width	0.7107	11	0.6919

Outlier Management

Outliers were identified by reviewing residual plots to highlight whether any individual records were skewing the results. Any outliers identified would only be removed from the dataset if the record was identified as likely to be corrupted or untrustworthy. No such reasons were found and therefore no outliers were removed from the modelling set.

Throughput based wear results

Predictive Performance Error

Table 7 shows that the lowest CV performance error was achieved by the full linear model, with a 45.9% improvement over the null hypothesis. The model containing only V1 as a variable achieved a 40.1% improvement, whilst the conveyor-duty only model achieved 36%; combining these variables into a single model achieved a 42.4% improvement. The simple linear model containing transformed variables (1/belt-width, 1/belt-length) but without V1 or belt-speed^2, performed substantially better than a similar model without these transformations (i.e. using belt-length, belt-width). All variable selection algorithms performed with an improvement above 45%, except for Lasso's optimal option, which only performed similarly well to the V1 -only model.

Final Model Fit

Once performance error was measured, the modelling algorithm was rerun on the entire dataset to return a final fitted model. Table 8 shows the best model identified by each modelling option (if applicable) for the wear throughput metric, along with the respective R^2 and Adjusted-R^2 values. The lowest Adjusted-R^2 best fit of 0.7599 was achieved by the Stepwise algorithm due

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

to having fewer predictors but was closely followed by the Full model and the Weighted-Average model, both of which used all predictors. Very similar fits (Adjusted-R^2 ~0.75) were also achieved by the model with only simple transformed predictors and the Lasso's best model. Unsurprisingly, the worst fits were obtained by the simplest models, although these still all had Adjusted-R^2 values above 0.62.

Table 7: Cross Validation Performance Error Results.

Model selection approach	P-RMSE Mean mm/MT	% imp over null hypothesis	Ave R^2 of folds	P-RMSE Mean mm/wk	% imp over null hypothesis	Ave R2 of folds
Null Hypothesis – No variables	0.2286			0.0777		
Full model (All variables)	0.1237	45.90%	0.661	0.0435	44.00%	0.628
Simple variables only	0.1419	37.90%	0.575	0.0447	42.50%	0.623
Simple transformed variables	0.1253	45.20%	0.664	0.0437	43.70%	0.627
Conveyor-duty only	0.1462	36.00%	0.534	0.0462	40.60%	0.604
V1 only	0.137	40.10%	0.59	0.0524	32.50%	0.542
V1 + conveyor-duty	0.1317	42.40%	0.607	0.043	44.70%	0.635
Stepwise Model Selection	0.1249	45.40%	0.653	0.043	44.60%	0.629
Averaged Models (ΔAIC≤10)	0.1245	45.50%	0.659	0.0428	44.90%	0.634
Lasso selection – best model	0.1239	45.80%	0.663	0.0424	45.50%	0.639
Lasso selection – optimal model	0.1371	40.00%	0.617	0.0445	42.70%	0.615

Table 8: Throughput model candidates with ΔAIC≤2.

	Model formula	Weights	ΔAIC
1	rate ~ 1 + conveyor-duty + v1 + belt-speed + belt-speed^2 + 1/belt-width	0.102	0
2	rate ~ 1 + conveyor-duty + v1 + perc-fines + belt-speed + belt-speed^2 + 1/belt-width	0.084	0.39
3	rate ~ 1 + conveyor-duty + v1 + belt-strength + belt-speed + belt-speed^2 + 1/belt-width	0.062	0.98
4	rate ~ 1 + conveyor-duty + v1 + drop-height + belt-speed + belt-speed^2 + 1/belt-width	0.053	1.29
5	rate ~ 1 + conveyor-duty + v1 + belt-strength + perc-fines + belt-speed + belt-speed^2 +1/belt-width	0.052	1.36
6	rate ~ 1 + conveyor-duty + v1 + drop-height + perc-fines + belt-speed + belt-speed^2 + 1/belt-width	0.04	1.85
7	rate ~ 1 + conveyor-duty + v1 + belt-speed + belt-speed2 + 1/belt-length + 1/belt-width	0.039	1.91

Variable Selection & Importance

Both the Stepwise and the Weighted Averaging methods selected the same candidate (rate ~ 1 + conveyor-duty + v1 + belt-speed + belt-speed2 + 1/belt-width) as the best performing model based on the AIC metric. In terms of CV performance error, this model had 0.5% less improvement over the null hypothesis than the full linear model.

The model averaging algorithm was further interrogated to list top performing models to ascertain what variables appeared repeatedly. These are shown in Table 8, in order of increasing ΔAIC. Seven models were identified as potential candidates for the best model due to having a ΔAIC ≤ 2. Five variables appeared in each of these 7 models: 1/width, V1, conveyor-duty, belt-speed and belt-speed^2. Another 80 models had ΔAIC values between 2 and 10.

The Lasso best model only dropped belt-speed2 from its list of variables. The smallest optimal Lasso model however, contained only V1, 1/belt-width and 1/belt-length. The CV performance of this optimal model was 5% worse than the best model identified by Lasso. Variable Importance graphs, as presented in Figure 10, show that although all methods tested agree on the importance of 1/width and V1 to the wear-throughput outcome, they differ with regards to other parameters. Of interest is the exclusion of both belt-speed^2 and the reduced importance of belt-speed in the Lasso models; these parameters were deemed highly important to the full, Stepwise model and Weighted Averaged models (in which they were included in all top 7 best performing models).

Time based wear results

Predictive Performance Error

Results from modelling the time-based wear rate are presented in Table 7 and Table 6. As already mentioned, time-based models had slightly less improvement (by up to 1%) over the null hypothesis than throughput variants, but still except for the single variable V1 model, all options managed to reduce

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

performance error by at least 40%. In terms of performance error, the biggest contributors to the reduction in error again appears to be combination of V1 and conveyor-duty; the model containing only these two variables achieved a reduction in error of 44.7% over the base case, which was only 0.8% less than the maximum achieved by more complex candidates. Modelling conveyor-duty alone achieved a 40.6% improvement, whilst modelling V1 only achieved a 32.5% improvement. The lowest CV performance error, with an improvement of 45.5% over the null hypothesis, was achieved by the Lasso best model.

Final Model Fit

Of the set of time-based models studied, all multi-predictor options achieved similarly good fits above 0.7 for their final models. The Stepwise algorithm reported the maximum Adjusted R^2 value of 0.7346, followed by the Lasso best model with a value of 0.7312, the full model with a value of 0.7309, and the Weighted-Average approach with a value of 0.7312. The minimum for the V1 model was 0.5225, whilst the conveyor-duty model achieved a better fit with an Adjusted R^2 of 0.6709. The Lasso's optimal model had a fit of 0.6794, accounting for its 8 predictors.

Variable Selection & Importance

Again, the same candidate (wear rate ~ 1 + conveyor-duty + V1 + belt-strength + belt-speed^2) was selected by both Stepwise and Weighted Averaging functions as having the lowest metric (AIC) and thus deemed the 'best' model. The Lasso method selected a more complex model as its best option, with all variables being retained in this model except for percentage-fines and 1/length. For its simplest viable model, the Lasso selected a candidate containing only four variables: conveyor-duty, V1, 1/length and 1/width. This model only performed 2% worse than the best Lasso model when compared to the null hypothesis.

Fourteen models were identified by the model averaging algorithm as potential candidates for the best model by having $\Delta AIC \leq 2$. These are shown in Table 9; all candidates contain conveyor-duty, V1, belt-strength variables. Most of the top 14 candidates also include either belt-speed or belt-speed^2. Another 136 models had AICC weights between 2 and 10.

Table 9: Time based model candidates with $\Delta AIC \leq 2$.

	Model formula	Weights
1	rate ~ 1 + conveyor-duty + v1 + belt-strength + belt-speed^2	0.0498
2	rate ~ 1 + conveyor-duty + v1 + belt-strength + belt-speed	0.0469
3	rate ~ 1 + conveyor-duty + v1 + belt-strength + belt-speed2 + 1/belt-width	0.0421
4	rate ~ 1 + conveyor-duty + v1 + belt-strength + belt-speed + 1/belt-width	0.0378
5	rate ~ 1 + conveyor-duty + v1 + belt-strength + drop-height + belt-speed^2	0.033
6	rate ~ 1 + conveyor-duty + v1 + belt-strength + drop-height + belt-speed	0.0318
7	rate ~ 1 + conveyor-duty + v1 + belt-strength + drop-height + belt-speed^2 + 1/belt-width	0.0276
8	rate ~ 1 + conveyor-duty + v1 + belt-strength + drop-height + belt-speed + 1/belt-width	0.0255
9	rate ~ 1 + conveyor-duty + v1 + belt-strength	0.024
10	rate ~ 1 + conveyor-duty + v1 + belt-strength + 1/belt-length	0.0201
11	rate ~ 1 + conveyor-duty + v1 + belt-strength + drop-height	0.0191
12	rate ~ 1 + conveyor-duty + v1 + belt-strength + belt-speed + belt-speed^2	0.0189
13	rate ~ 1 + conveyor-duty + v1 + belt-strength + perc-fines + belt-speed^2	0.0189
14	rate ~ 1 + conveyor-duty + v1 + belt-strength + belt-speed^2 + 1/belt-length	0.0187

Variable Importance graphs for time-based models are presented in Figure 11. These show that V1 and conveyor-duty were important to all models, whilst belt-strength, belt-speed, belt-speed^2 and 1/belt-width, differed in the degree to which they were considered important by the different modelling approaches. 1/belt-length was not identified as particularly important by any modelling approach.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

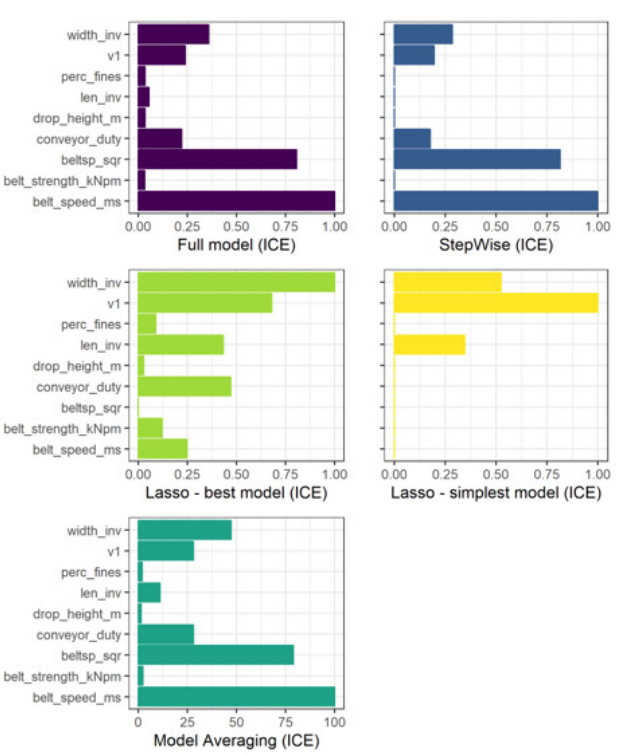

Figure 10: Variable importance graphs for different model selection approaches for throughput metric.

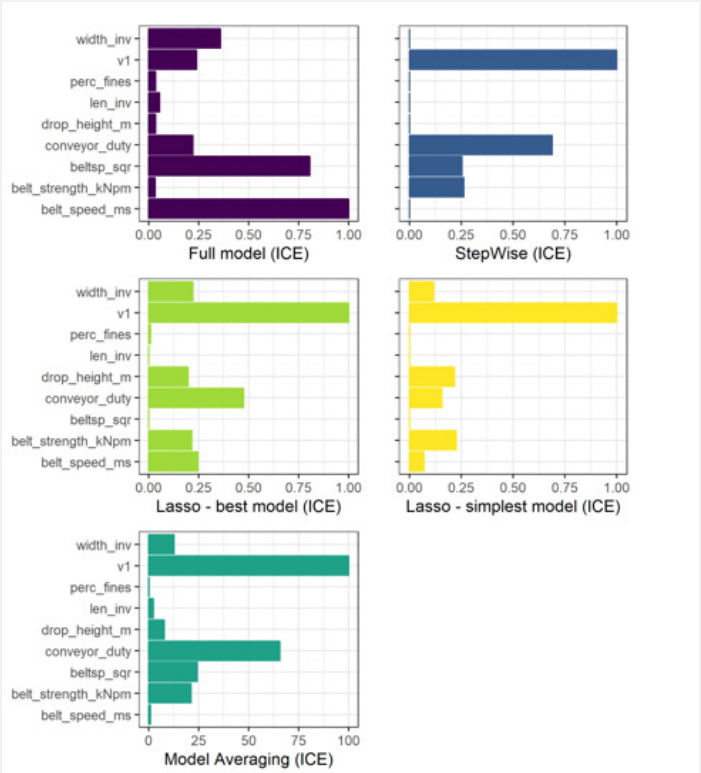

Figure 11: Variable Importance summaries for time-based models.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

Discussion

All modelling options (including the models with only a single variable) offered measurable and significant (up to 45.9%) improvements over the null hypothesis. Throughput models performed slightly better than time-based models (compared to the respective null hypothesis) but the difference was small (<2% for most models). For these conveyors, utilization was relatively constant over time, so the ability to account for fluctuating operating conditions intrinsic to the throughput models, was not influential.

For both wear rates, Stepwise selection, Weighted Averaging and Lasso modelling (using the best model option) resulted in almost identical CV performance errors and very similar quality of fits for their final model. In the case of time-based wear rates, model selection algorithms resulted in a marginally lower CV performance error than the full linear model. However, this was not the case for throughput models, for which the full linear model offered the lowest CV prediction error estimate. In both cases, the difference was small, indicating that for this dataset, variable selection methodology had little effect on prediction performance. Furthermore, the Stepwise selection did not miss the optimal model, which is a commonly reported risk of this specific variable selection approach [52,53,60]; in both cases, Stepwise selection chose the same candidate as the Weighted-averaging method, which independently ranked all 4096 options when using the same (AIC) metric. Another reported problem with Stepwise Selection is that variable selection in the presence of collinearity is made arbitrary [55]. This issue was observed when we initially attempted to include both V1 and load-frequency in the same models. Not only did the Stepwise algorithms take much longer to converge (hours rather than minutes), the final 'best' model and the overall performance error varied each time the data was grouped into different folds. Other methods also took longer to converge. These issues did not occur when load-frequency was removed from modelling. The results of this are difficult to report as processing was often aborted without a result, but the observation is worth noting.

Lasso selected different models for each metric, probably because it used a different metric for variable selection, MSE, rather than AIC as used by the Stepwise and Weighted-Averaging algorithms. However, as already mentioned the significance of this difference to prediction error was marginal for throughput models, and there was no difference for time-based models.

Other commonly quoted issues with model selection algorithms, first summarized in [55], include criticisms that the final R^2 is biased to be high, the standard errors of regression coefficients are biased low, confidence intervals for prediction estimates are too narrow and the regression coefficients are biased high in absolute terms and require shrinkage. Herein however, the magnitude of these effects seems to be small when results were compared against the full model.

The reason for the small impact of model selection on performance error is probably due to the fact that most of the improvement over the null hypothesis was due only to two variables: V1 and conveyor-duty, which together resulted in a 42.4 and 44.7% improvement for throughput and time-based wear rates respectively. The remaining variables only improved performance error by a further 3.5% for throughput models, and 0.2% for time-based models. Even without using V1 in the models, including the transformed variables that comprise V resulted in a smaller CV error than using the original untransformed variables (models 3 and 4). This effect was more significant for throughput models than for time-based models.

The smaller importance of V1 and belt-width, to time-based models can be easily explained. V1 had been derived from first principles to be proportional to wear rate per tonnage of ore transferred, which is the same as the throughput wear rate predictor. The time- based wear rate predictor was tonnage independent and so a relationship with V1 (or its constituent parts) was less likely. Similarly, belt-width is expected to be more important to throughput models than time-based options because the metric includes the amount of ore transferred in its calculation. Thus, stronger relationships between the throughput metric and both belt-width and belt-length would be expected. For these conveyors, belt-length also appeared to be strongly correlated with V1 and partly correlated with conveyor-duty, it is possible that the importance of this information to the wear rate, was in part accounted for by the other variables.

The slight differences in performance between V1 and conveyor-duty imply that although most of benefit of conveyor duty is probably due to the interaction between belt-length, belt-width and belt-speed (i.e. the number of times the conveyor is loaded and by how much), there is still some aspect of a conveyor's duty that has not be captured by these variables. Belt-grade is often selected based on expected duty, so it is quite plausible that inclusion of this variable, as well as more material parameters that also may differ between plants, may help quantify the missing effects and reduce the reliance on this categorical variable.

In this work, drop height was not an important variable for predicting conveyor wear. Although dropping an object from a greater vertical height may increase the likelihood and severity of pitting damage, conveyor loading chutes are designed to translate the potential energy of the ore into horizontal kinetic energy that matches the speed of the conveyor before it is dropped onto the belt [67]. A better variable in our models may be a measure of the relative velocity between the ore exiting the chute and the belt speed or of the impingement angle as suggested by [8].

Percentage fines did not seem to be a particularly helpful variable either. Unfortunately, it is difficult to provide a reason for this without knowing more about the physical property differences between the fines and lump; it is possible that this simple categorization did not adequately differentiate the physical

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

properties of ore that are known to be relevant to wear. Specific physical properties, such as average density, hardness or lump size may be more useful for modelling.

Conclusion

This work shows that the wear rate of heavy-duty conveyor belts is, at least in part, related to the compound variable of belt-speed^2 divided by belt-capacity (length x width). In this, our work supports the theoretical work derived in [7] and similar work briefly presented in [8]. We also demonstrate that this compound variable is more useful than using the same data as individual transformed variables, which themselves are more useful predictors than their original forms.

However, we also show that these variables alone do not entirely explain the wear patterns observed on this set of conveyor belts. Conveyor-duty was also critical to all models, which is somewhat problematic as it there is no universal definitions as to what defines a specific duty class. Further work is therefore required to determine how better to capture this information in a more transparent and quantifiable way.

With the variables available, the best models in our study only accounted for 70-75% of the variation observed in the data, and most of this was due to only one or two variables (conveyor duty and V1). The omission of belt-grade was particularly concerning, as this is a belt material property (albeit a categorical one) that directly relates to a belt's ability to withstand abrasive wear. However, due to inadequate collection practices, this data could not be trusted and was thus not used. Similarly concerning, the models tested contained no explanatory variables relating to the initial velocity of the ore landing on the conveyor, or its physical properties, such as density, particle size or hardness. Engineering and materials researchers have independently shown than these features affect rubber and conveyor belt wear. Therefore, it is likely that adding these variables into models would significantly reduce the prediction error and improve the usefulness of results.

This work also demonstrates how a predictive model can be robustly developed to predict wear rates for out of sample conveyors using nested cross-validation. Although the cross-validation approach detailed herein is widely used in many other fields, there have been few published works detailing its use to engineering reliability and maintenance management. Its suitability for both model selection and performance estimation on small datasets makes cross validation a particularly useful method in analyzing engineering failure data. Unlike other fields that are increasingly having to grapple with big data issues, improving equipment reliability means that failure datasets will likely become smaller as equipment reliability improves and the incidence of failures decline.

The resulting models offered in this paper provide a significant improvement over the current practice of using the mean of the population to predict conveyor life, reducing the prediction error by almost 50%. However, the prediction error of all models remains too high to be a reliable predictor of individual wear-rates, with prediction errors having the same order of magnitude as the prediction itself. Therefore, for individual conveyors, where the cost of an unplanned shutdown is greater than the cost of periodic or continuous thickness monitoring, using condition monitoring (be it thickness testing or another technique) to facilitate individual conveyor maintenance planning is still the best approach. The performance of prediction models would need to improve substantially before they could be a genuine alternative to individual belt monitoring. Alternatively, these models could be used as the initial, a-priori model in Bayesian methods that would then update models as each time new wear readings were received.

Although widely researched by academics, in practice, predictive modelling is still not widely utilized by industry maintenance professionals. This may be because the underlying data available to develop robust and reliable models is not readily available or of sufficient quality to be useful, which leads to disappointing results when modelling is attempted. In this project, like in most other modelling endeavors, most time was spent wrangling and cleaning the data, with some problems only becoming apparent when the raw data was assessed by experienced reliability engineers. To improve the uptake of statistical modelling techniques in maintenance, we strongly recommend that engineers and data scientists be more judicious about the variables selected for modelling. Potential explanatory variables should be primarily selected because they are known by engineers and scientists to relate to the failure mechanism or its precursory operating conditions. Analyzing data simply because it is readily available will continue to result in poor outcomes and reduced confidence in the field by industry users.

Finally, if models are important to business decision making (and if they are not, then you should question why the models are being developed), more effort is required at the time of data collection to ensure that information necessary to subsequent modelling is complete and reliable. Trying to wrangle data sometime after it has been collected will always be fraught with difficulty. More importantly, it will affect the likely success of modelling projects and continue to disillusion end users. Thus, it is important to include both the data scientist and reliability engineer in the data collection process.

Nomenclature

ρ_{ore} Relative density of the material conveyed in kg/m^3

μ Co-efficient of friction between the belt and the bed of ore

$\mu_{metric,pool}$ Mean wear rate for each metric and data pool

A Relative abrasion resistance of the belt cover material

C_θ Correction factor for the angle of impingement

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

k Specific wear rate for abrasion of rubber by the ore

k Number of folds used in k-fold cross validation

l Length of the belt in m

$m_{w,pool}$ Gradients of the linear regression curves for each width position in the pool

n Total number of points in the dataset being processed during each iteration

n Sample size in R^2 calculation

n Number of candidate models being averaged by the glmulti algorithm

N_{-k} Total number of points being analysed.

$n_{w,pool}$ Number of width positions in a particular pool

p Number of explanatory variables in the linear regression model

t Thickness of rubber removed in mm

T_f Empirically derived tonnage factor

t_w Thickness of wearable cover in mm

v Speed of the belt in m/s

V1 Derived variable defined in this paper

w Width of the belt in m

WMT Wear life in megatonnes

y(i) Measured value

$\hat{y}$(i) Calculated model output

Data Availability Statement

Code that initially extracted the data from key systems and prepares the data for modelling cannot be shared as this contains proprietary information. However, subsequent code that extracts the graphs for this paper and then performs the modelling is provided in the supplementary materials.

Acknowledgement

This work would also have not been possible without funding from the BHP Fellowship for Engineering for Remote Operations – supporting community projects in areas in which BHP operates.

References

1. (2017) New entrant to conveyor monitoring market in Australia. Bulk Handling Review.
2. Zimroz R, Błażej R, Stefaniak P, Wyłomańska A, Obuchowski J, et al. (2014) Intelligent diagnostic system for conveyor belt maintenance. Mining Science 21(2): 99-109.
3. Blazej R, Jurdziak L (2017) Condition-Based Conveyor Belt Replacement Strategy in Lignite Mines with Random Belt Deterioration. ed: Institute of Physics Publishing 95(4).
4. Schallamach A (1954) On the abrasion of rubber. Proc Phys Soc B 67: 883,
5. Fukahori Y, Yamazaki H (1994) Mechanism of rubber abrasion. Part I: Abrasion pattern formation in natural rubber vulcanizate. Wear 171(1-2): 195-202.
6. Molnar W, Varga M, Braun P, Adam K, Badisch E (2014) Correlation of rubber-based conveyor belt properties and abrasive wear rates under 2- and 3-body conditions. Wear 320(1): 1-6.
7. Hutchings, Shipway P (2017) Tribology: Friction and wear of engineering materials. 2nd Edition ed. Oxford, UK.: Elsevier Ltd.
8. Pitcher D (1999) Predicting the life of rubber covered conveyor belting. presented at the Beltcon 11 - International Materials Handling Conference, South Africa.
9. Swinderman RT, Lindstrom D (1995) Belt Cleaners and Belt Top Cover Wear. ed: Institution of Engineers. Australia pp. 95-98.
10. Fedorko G, Molnar V, Dovica M, Toth T, Kopas M (2014) Analysis of pipe conveyor belt damaged by thermal wear. Engineering Failure Analysis 45: 41-48.
11. Andrejiova M, Grincova A, Marasova D (2016) Measurement and simulation of impact wear damage to industrial conveyor belts. Wear 368-369: 400-407.
12. SCP Ltd (2019) Definitive Guide for Choosing the Right Conveyor Belt.
13. Friedrich K (1986) Erosive wear of polymer surfaces by steel ball blasting. Journal of Materials Science 21(9): 3317-3332.
14. (2016) Metso Conveyor Solutions handbook, Metso Sweden AB, Trelleborg Sweden.
15. Masaki MS, Zhang L, Xia X (2018) A design approach for multiple drive belt conveyors minimizing life cycle costs. Journal of Cleaner Production 201: 526-541.
16. Goldbeck L (1997) Conveyor belt damage: Causes & cures. Engineering & Papermakers: Forming Bonds for Better Papermaking. books Atlanta: Tappi Press 1-3(1): 189-193.
17. Andrejiova M, Grincova A, Marasova D, Fedorko G, Molnar V (2014) Using logistic regression in tracing the significance of rubber–textile conveyor belt damage. Wear 318(1-2): 145-152.
18. Andrejiova M, Grincova A (2018) Classification of impact damage on a rubber-textile conveyor belt using Naïve-Bayes methodology. Wear 414-415: 59-67.
19. Grincova A, Andrejiova M, Marasova D, (2016) Failure analysis of conveyor belt in terms of impact loading by means of the damping coefficient. Engineering Failure Analysis 68: 210-221.
20. (2015) ASTM D2228 Standard Test Method for Rubber Property - Relative Abrasion Resistance by Pico Abrader Method. ASTM, West Conshohocken PA.
21. Miriam A, Daniela M (2013) Using the classical linear regression model in analysis of the dependences of conveyor belt life. Acta Montanistica Slovaca 18(2): 77-84.
22. Astfalck L, Hodkiewicz M, Keating A, Cripps E, Pecht M (2016) A modelling ecosystem for prognostics. presented at the Annual Conference of the Prognostics and Health Management Society, Denver, Colorado USA.
23. Sikorska J, Hodkiewicz M, A de Cruz, Astfalck L, Keating A (2016) A collaborative data library for testing prognostic models. presented at the 3rd European Conference of the Prognostics Health Management Society, Bilbao, Spain.
24. (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

25. (2019) Caret: Classification and Regression Training.

26. Venables WN, Ripley BD (2002) Modern Applied Statistics with S. 4 ed. New York: Springer.

27. (2019 Cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'.

28. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

29. (2018) dplyr: A Grammar of Data Manipulation.

30. Friedman J, Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33(1): 1-22.

31. Yang Y, Zou H (2015) A fast unified algorithm for solving group-lasso penalize learning problems. Statistics and Computing, 25(6): 1129-1141.

32. UIfDR Education (2019) R Library Contrast Coding Systems for Categorical Variables. UCLA.

33. Hurvich CM, Tsai CL (1990) The Impact of Model Selection on Inference in Linear Regression. The American Statistician 44(3): 214-217.

34. Hastie T, Tibshirani R, Friedman J (2011) The Elements of Statistical Learning. (2nd ed.) Springer Series in Statistics. New York: Springer-Verlag 2009.

35. Stone M (1977) An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. Journal of the Royal Statistical Society: Series B (Methodological) 39(1): 44-47.

36. Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics 7(1): 1-26.

37. Kuhn M (2013) Applied predictive modeling. New York: Springer.

38. James G, Witten D, Hastie T, Tibshirani R (2018) An Introduction to Statistical Learning: with Applications. In: Gazzo R, ed: QUBES Educational Resources.

39. Shalev SS, Ben DS (2014) Understanding Machine Learning: From Theory to Algorithms. Cambridge: Cambridge University Press.

40. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Presented at the 14th International Joint Conference on Artificial Intelligence, Canada 2: 1137-1143.

41. Van Hasselt H (2013) Estimating the Maximum Expected Value: An Analysis of (Nested) Cross Validation and the Maximum Sample Average.

42. Burman P (1989) A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods. Biometrika 76(3): 503-514.

43. Zhang Y, Yang Y (2015) Cross-validation for selecting a model selection procedure. Journal of Econometrics 187(1): 95-112.

44. Rao R, Fung G, Rosales R (2008) On the Dangers of Cross-Validation. An Experimental Evaluation. (ed.) Philadelphia: Society for Industrial and Applied Mathematics pp. 588-596.

45. Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. Statistics Surveys 4: 40-79.

46. Krstajic D, Buturovic L, Leahy D, Thomas S (2014) Cross-validation pitfalls when selecting and assessing regression and classification models. Journal of Cheminformatics 6(1).

47. JH Kim (2009) Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Computational Statistics and Data Analysis 53(11): 3735-3745.

48. S Arlot (2007) Resampling and Model selection.

49. Breiman L, Spector P (1992) Submodel selection and evaluation in regression. The X-random case. International Statistical Review 60(3): 291-319.

50. Konishi S (2008) Information criteria and statistical modeling (Springer series in statistics). New York: Springer.

51. Rogers S, Girolami M (2016) A First Course in Machine Learning. Second Edition. Chapman \& Hall/CRC p. 427.

52. Lukacs P, Burnham K, Anderson D (2010) Model selection bias and Freedman's paradox. Annals of the Institute of Statistical Mathematics 62(1): 117-125.

53. Burnham KP (2002) Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. ed. New York: Springer.

54. Buckland ST, Burnham KP, Augustin NH (1997) Model Selection: An Integral Part of Inference. Biometrics 53(2): 603-618.

55. Harrell FE (2001) Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis (Springer series in statistics.). New York: Springer.

56. Tibshirani R (1996) Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58(1): 267-288.

57. Steyerberg EW, Eijkemans MJC, Harrell FE, Habbema JDF (2000) Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Statistics in Medicine 19(8): 1059-1079.

58. Mundry R, Nunn CL (2009) Stepwise model fitting and statistical inference: Turning noise into signal pollution. American Naturalist 173(1): 119-123.

59. Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6): 716-723.

60. Burnham KP, Anderson DR (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods & Research 33(2): 261-304.

61. Cavanaugh J, Neath A (2019) The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. Wiley Interdisciplinary Reviews 11(3).

62. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2): 301-320.

63. Kvålseth TO (1985) Cautionary Note about R^2. The American Statistician 39(4): 279-285.

64. Ramsey FL (2013) The statistical sleuth: a course in methods of data analysis. 3rd ed. Boston, Mass: Brooks Cole.

65. Greenwell B, Boehmke B, McCarthy A (2018) A Simple and Effective Model-Based Variable Importance Measure. arXiv.org.

66. Goldstein A, Kapelner J, Bleich E (2015) Pitkin Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. Journal of Computational and Graphical Statistics 24(1): 44-65.

67. Xie L, Zhong W, Zhang H, Yu A, Qian Y, Situ Y (2016) Wear process during granular flow transportation in conveyor transfer. Powder Technology 288: 65-75.

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.

DOI: 10.19080/IMST.2021.02.555594

How to cite this article: Joanna Z S, Callum W, Nazim K, Melinda H. Variable Selection for Conveyor-Belt Mean Wear Rate Prediction. Insights Min Sci technol.2021; 2(4): 555594. DOI: 10.19080/IMST.2021.02.555594.