Original Article

Institution of MECHANICAL ENGINEERS

Journal of RAIL AND RAPID TRANSIT

Proc IMechE Part F:
J Rail and Rapid Transit
0(0) 1–9

DOI: 10.1177/0954409717695650
journals.sagepub.com/home/pif
SAGE

Survival analysis at multiple scales for the modeling of track geometry deterioration

Negin Alemazkoor[1], Conrad J Ruppert[2] and Hadi Meidani[1]

Abstract

Defects in track geometry have a notable impact on the safety of rail transportation. In order to make the optimal maintenance decisions to ensure the safety and efficiency of railroads, it is necessary to analyze the track geometry defects and develop reliable defect deterioration models. In general, standard deterioration models are typically developed for a segment of track. As a result, these coarse-scale deterioration models may fail to predict whether the isolated defects in a segment will exceed the safety limits after a given time period or not. In this paper, survival analysis is used to model the probability of exceeding the safety limits of the isolated defects. These fine-scale models are then used to calculate the probability of whether each segment of the track will require maintenance after a given time period. The model validation results show that the prediction quality of the coarse-scale segment-based models can be improved by exploiting information from the fine-scale defect-based deterioration models.

Keywords

Track geometry defects, red tag defects, deterioration modeling, survival analysis

Date received: 21 June 2016; accepted: 29 January 2017

Introduction

Railway is an important mode of transportation in the United States. It carries the largest share of freight moves on a ton–mile basis[1] and moves millions of people on a daily basis in most of the major cities in the US. Therefore, railroad safety is a crucial issue, as train accidents and derailments can cause casualties, damages to infrastructure and trains, delays, and environmental harm. Track geometry defects have been identified as a major cause of derailments.[2] To maintain the safety of railroads, the Federal Rail Administration (FRA) has set track safety standards, which includes a section on track geometry defects. Individual track geometry defects with amplitudes greater than the safety standards must be corrected or protected within a prescribed time limit.[3]

In current practice, track geometry cars classify the defects into two severity groups, commonly named as "red tags" and "yellow tags". Red tag defects are those that exceed the FRA standards and yellow tags are those that exceed a railroad company's maintenance threshold. A deterioration model for track geometry defects, which that makes reliable predictions on when a yellow tag defect will turn into red tag can therefore help railroad companies to better plan preventive maintenance. This would lead to reducing the number of unplanned corrective maintenance actions in response to red tag defects, thereby reducing delays and the maintenance costs for rectifying defects.

Deterioration is a complex stochastic process, as many factors that impact the deterioration cannot be captured by the available data. Therefore, probabilistic models that account for the stochastic nature of deterioration should be preferred over deterministic regression models. Deterministic and probabilistic models have been developed to estimate the degradation of track segments. Although these models may provide accurate estimations of future condition for a track segment, for example in terms of track quality index (TQI), the future condition of isolated defects within that segment cannot be extracted from them. This is while even a single isolated defect that is in violation of FRA safety standards requires immediate protection or corrective maintenance actions. In other words, segment-based

[1]Department of Civil and Environmental Engineering, University of Illinois at Urbana–Champaign, Champaign, IL, USA
[2]Rail Transportation and Engineering Center (RailTEC), University of Illinois at Urbana–Champaign, Champaign, IL, USA

Corresponding author:
Hadi Meidani, Department of Civil and Environmental Engineering, University of Illinois at Urbana–Champaign, 1211 Newmark Civil Engineering Bldg, 205 N. Mathews, Urbana, Champaign, IL 61801, USA.
Email: meidani@illinois.edu

predictions may fail to capture the actual condition of defects within a segment, and result in overconservative or underconservative actions. Therefore, additional work is needed regarding isolated defects deterioration modeling.

In this paper, we use the survival analysis to model the lifetime of yellow tag geometry defects and calculate their probability of turning into red tag defects, referred to as failure probability, within a given time period. Given the failure probability for each defect within a track segment, the probability of having at least one failure within the segment can be consequently calculated. We compare the results of coarse-scale survival analysis that directly calculates the failure probability for a track segment with the results of fine-scale defect-based models.

The paper is structured as follows. The next section reviews the current practice on deterioration modeling for track geometry. Next, a section is specified to describe the survival analysis approach in detail. Then the description of the dataset along with parameter estimation and discussion of fine-scale and coarse-scale track geometry survival models are provided. Finally, the last section compares the accuracy of coarse-scale and fine-scale defect-based survival models in estimating the failure probability for a track segment.

Background on modeling track geometry degradation

There are two different approaches in the literature for track geometry deterioration modeling: (a) mechanistic approach, where the mechanism of degradation is investigated and (b) statistical approach, where the relationship between degradation and track features that best describe the degradation data is identified. Some of the main mechanistic models are those developed in Shenton,[4] Sato,[5] and Chrismer and Selig.[6] In what follows, we provide a survey of the statistical models developed for track geometry deterioration.

Regression analysis has been widely used in track degradation modeling.[7–9] However, it fails to account for inherent uncertainty in the degradation process, something only identifiable in probabilistic models. In Andrade and Teixeira,[10] this issue was addressed by considering the constant value and the coefficients in the regression model as uncertain parameters with prior distributions. A Bayesian framework using inspection data was employed to compute the posterior distributions for initial longitudinal deviation and rate of deterioration in a single variable tonnage-based linear regression model that predicts the standard deviation of longitudinal defects over a segment.

In Meier-Hirmer et al.,[11] the uncertainty in track deterioration was considered and gamma process was used to model the degradation of track segment in terms of the synthesized Mauzin indicator of the longitudinal leveling. Gamma process accounts for uncertainty by giving a probability distribution, instead of a single value, for cumulative deterioration (i.e., the increase in the amplitude of Mauzin indicator within a given time). However, it does not account for the features of the track such as tonnage, class of track, etc. Therefore, in practice, it would be necessary to first classify the tracks based on their features and then estimate the parameters of gamma process for each class, which increases the modeling cost.

In Bai et al.,[12] some features of the track were accounted for in a probabilistic track deterioration model. Deterioration was modeled as a Markov process with four condition states. Condition states were defined based on the TQI, which was taken to be the sum of standard deviations of seven geometrical parameters over a segment. Then, exponential hazard models, with gross tonnage and percentage of curve track in the segment as explanatory variables, were used to estimate the transition probabilities between condition states.

The above-discussed models predict the overall degradation of a track segment in terms of TQI. Although TQI for a segment can give insights on the condition of the segments, it does not provide specific information on whether isolated defects in the segments have exceeded or will exceed the safety limits or not. Therefore, in order to make defect-based predictions, e.g. when yellow tag defects turn into red tag defects, and plan preventive maintenance actions, it is necessary to also develop models for isolated defect deterioration.

It is expected that the uncertainty and variability in the degradation of individual defects are larger than that in the overall degradation of track segments. Hence, probabilistic approaches should be preferred in defect-based deterioration modeling. Different probabilistic methods can be used to analyze geometric defects, one of which is survival analysis. Survival analysis concerns about analyzing the waiting time until a specific well-defined event happens. Survival analysis has applications in different fields, such as biomedical, engineering, and economics, where the organ, equipment, or system of interest has two states: a functioning state and a failed state.[13,14] The survival analysis aims to estimate, for a component, the probability distribution of its lifetime, i.e. the time spent in the functioning state before transition into the failed state.

Analyzing the track geometry defects can be readily done using survival analysis, as geometric defects can be considered to belong to two states, the yellow tag and red tag states. The objective is then to determine when a yellow tag becomes a red tag. The time a defect spends in the yellow state can be considered as the lifetime of the defect and the transition to the red tag is considered the failure event.

Survival analysis was used in Larsson-Kråik et al.[15] to estimate the probability of having a failure, i.e. at least one exceedance of safety limit, in a track segment. However, the initial condition of the track segment or the track features was not considered in the model. In this work, we adopt a similar approach but in a finer scale. We calculate the probability of failure for each yellow tag defect, which can further be used to estimate the probability of having a failure in a segment. Moreover, we parameterize the failure probability distribution to account for initial amplitude of yellow defects and track features. We then compare the results of coarse-scale survival analysis that directly calculates failure probability for track segment with the results of these fine-scale defect-based models.

Survival analysis

As mentioned earlier, the survival analysis focuses on modeling the lifetime of systems with two states. We denote the functioning or good state by G, and the failed state by F. A sequence of the observed state conditions for such a system has the form $\{G, G, \ldots, F, F, \ldots, F\}$. When a failure happens, the system will remain in the failed condition unless a maintenance action is done. The time that system stays in the working state may depend on a number of factors, or explanatory variables. The survival analysis aims to describe this dependence mathematically. Consider lifetime, T, as a continuous random variable in $[0, \infty]$ with probability density function $f(t)$ and cumulative distribution function $F(t) = P(T < t)$. $F(t)$ gives the probability that a failure has happened by time t. The probability that a failure has not happened by time t, referred to as the survival probability, is given by $1 - F(t)$. Therefore, the survival function is defined as the complement of the cumulative distribution function

$$S(t) = \int_t^\infty f(t')\mathrm{d}t' \tag{1}$$

Using the concept of conditional probability, let $h(t, \Delta t)$ denote the probability that a failure happens between time t and $t + \Delta t$, given that no failure has happened by the time t. $h(t, \Delta t)$ is given by

$$h(t, \Delta t) := P(t < T < t + \Delta t | T > t) = \frac{F(t + \Delta t) - F(t)}{S(t)} \tag{2}$$

The average rate of failure per unit of time can be obtained by dividing $h(t, \Delta t)$ with Δt, and the instantaneous rate of failure, denoted by $z(t)$, called hazard rate function, can then be obtained by choosing Δt to be very small

$$z(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t | T > t)}{\Delta t} = \frac{-\mathrm{d} \ln S(t)}{\mathrm{d}t} \tag{3}$$

The hazard rate function gives fundamental information about the nature of the event that is modeled. A constant hazard rate implies system's lack of memory. Monotonically decreasing hazard rates indicate that the failure is more probable to happen in the early life of the system. For example, errors and bugs are usually detected early after releasing a new computer program. Monotonically increasing hazard rate indicates that the probability of failure increases by increasing the time spent without failure. Since the infrastructures are prone to material aging, fatigue, and stress, they follow monotonically increasing hazard rates.

Weibull hazard function

A survival analysis usually begins with choosing a distribution for the hazard function. The simplest choice for hazard function is a constant function, which means that the probability of failure remains the same overtime. A constant hazard function implies that lifetime has an exponential density function

$$z(t) = \lambda \tag{4}$$

$$S(t) = \exp(-\lambda t) \tag{5}$$

$$f(t) = \lambda \exp(-\lambda t) \tag{6}$$

Exponential distribution has only one parameter, namely the scale parameter, λ. Weibull distribution is a generalized form of the exponential distribution and includes an additional parameter, namely the shape parameter. Therefore, Weibull distribution is more flexible and better applicable for modeling lifetime and hazard occurrence.[16,17] The failure probability distribution function associated with the Weibull distribution is given by equation (7). The corresponding survival distribution and hazard function can be given using equations (1) and (3) as

$$f(t) = \lambda^p p t^{p-1} \exp(-(\lambda t)^p) \tag{7}$$

$$S(t) = \exp(-(\lambda t)^p) \tag{8}$$

$$z(t) = \lambda^p p t^{p-1} \tag{9}$$

where p and λ are the shape and the scale parameters, respectively. When $p = 1$, the Weibull distribution takes the form of the simple exponential distribution and hazard rate takes the constant value of λ. If $p > 1$, then the hazard function, $z(t)$, is monotonically increasing, whereas for $0 < p < 1$, it is monotonically decreasing. Figure 1 shows the Weibull probability distribution for $\lambda = 1$ and different shape parameters, and demonstrates the flexibility of Weibull distribution in modeling various behaviors.

The survival analysis is aimed to represent the dependence of lifetime on the explanatory variables.

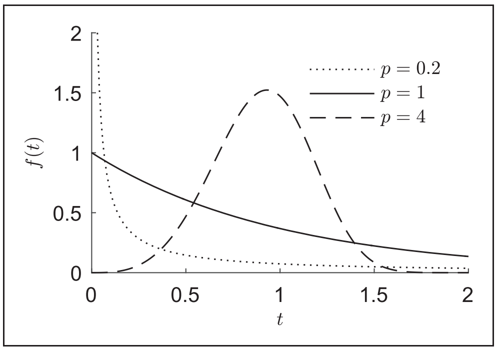

Figure 1. Weibull probability densities for $\lambda = 1$ and different shape parameters.

However, the Weibull hazard function does not account for the impact of the explanatory variables. To overcome this limitation, an extension of the Weibull distribution is used, in which the scale parameter λ is considered to be a function of exogenous variables according to Kleinbaum and Klein[18] and Mishalani and Madanat[19]

$$\lambda = \exp(-\beta \mathbf{X}) \tag{10}$$

$$z(t) = \exp(-p\beta \mathbf{X})\, pt^{p-1} \tag{11}$$

$$S(t) = \exp(-(\exp(-\beta \mathbf{X})t)^p) \tag{12}$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is the column vector of explanatory variables and $\beta = (\beta_1, \beta_2, \ldots, \beta_3)$ is the row vector of the coefficients. X_1 has a constant value of one and $X_2, \ldots, X_n$ are the value of the explanatory variables considered in the model. Vector of coefficients, β, and the shape parameter, p, should be estimated by the modeler.

Parameter estimation

Parameters p and β in equations (11) and (12) can be estimated using maximum likelihood method based on the observed lifetimes of the modeled system. The observations may be uncensored or censored. Uncensored observations are those for which the lifetime, T, is known, i.e. the start time of life and failure time is known. For censored observation, T is unknown and the only available information is that T is greater or less than a certain limit. Censored infrastructure deterioration observations are most likely right-censored, meaning that the construction time or latest repair time is known but the failure has not happened by the time of last inspection.

Consider a total of m observations, the first n of observations are uncensored, and the next $m-n$ observations are right-censored. The likelihood can be given by

$$L = \prod_{i=1}^{n} f(\beta, p, \mathbf{X_i}, t_i) \prod_{i=n+1}^{m} S(\beta, p, \mathbf{X_i}, t_i) \tag{13}$$

and the optimal parameters to be used in survival analysis can be given by

$$\{p^*, \beta^*\} = \underset{p,\beta}{\operatorname{argmax}} \sum_{i=1}^{m} \ln f(\beta, p, \mathbf{X_i}, t_i) \times \sum_{i=n+1}^{n} \ln S(\beta, p, \mathbf{X_i}, t_i) \tag{14}$$

Since equation (14) includes non-linear function of p and β, iterative methods are usually used to find the values of p^* and β^* that maximize the log-likelihood.[20,21]

Coarse- and fine-scale track geometry survival models

The survival analysis can be used to predict failure probability, i.e. exceedance probability of geometry defects, at different scales. That is, for a system of interest, depending on the engineering need, the lifetime of the whole system or a specific component or sub-system can be modeled using the survival analysis. Almost all engineering and natural systems involve multiple spatial or temporal scales.[22] Coarse-scale models are developed at scales that are larger than the scale of governing microscopic processes, and aim to describe the behavior at a macroscopic level. These coarse-scale representations can be inaccurate due to loss of information from a lower scale, depending on the complexity of the modeled system. Developing a model at a microscopic level, on the other hand, can improve the accuracy. In what follows, we seek to investigate the impact of the scale on the quality of the survival models.

Fine-scale deterioration models, if computationally possible, can improve the accuracy and reliability of predictions. Reliable predictions on when and where a red tag defect will occur in a given track can be used to better plan preventive maintenance actions and reduce the maintenance cost. A detailed version of this predictive model, also referred to here as a fine-scale model, produces the failure probability, i.e. the occurrence probability of a red tag defect, at locations of yellow tag defects. However, in practice, this fine level of prediction is not pursued, as corrective actions are typically performed on a segment of a track, typically stretching half a mile, and estimates of failure probabilities are only needed for segments. That is why commonly used track geometry deterioration models are of coarse-scale (segment-based). These models predict the overall TQI for a segment or estimate its probability of containing at least one red tag defect.[10,11,12,15] At this coarse-scale level, detailed information about individual yellow tag geometry defects within segments are disregarded. That is, while this detailed lower-scale information is crucial in coarse-scale prediction as the yellow tag defects within a segment are most vulnerable locations to failure, the red tag defects most probably will occur at the location of yellow tag defects. Therefore, the accuracy of the segment-based deterioration models not incorporating the information on the individual yellow tag defects will be in question.

To improve the accuracy of track geometry deterioration models, the deterioration process of yellow tag defects within a segment should be analyzed. In this work, a fine-scale model, that predicts the probability of an individual yellow defect turning into a red tag, is used to predict a coarse-scale failure probability measure. This coarse-scale measure is taken to be the probability of having at least one yellow tag within the segment. We demonstrate the advantage of this approach using an actual track geometry dataset. This section continues with the description of the dataset, and the estimating parameters of fine-scale and coarse-scale track geometry survival models.

Measurement data

The dataset used in this work was provided as training and test datasets from the INFORMS 2015 Railway Applications Section Problem Solving Competition. The dataset includes records of the yellow and red tags for three defect types between 2007 and 2013 on four different tracks. The dataset was aimed to be blind; therefore, the information on the locations of the tracks were not given. The three defect types included in the dataset are dip measured with a 31′ chord (dip31),[23] surface measured with a 62′ chord (surface62), and crosslevel. Table 1 shows the values of the red and yellow tag limits used in this dataset. For each defect, the milepost, type, amplitude, and length of the defect are included in the dataset. The dataset also includes track code, class of track, operating speeds of freight, and passenger trains at the location of each defect, and total monthly gross tons traveling across each section. Although the dataset does not contain information on maintenance actions, it includes time and spatial information of all inspection runs. Whenever an inspection run was observed with no recorded defect at a location with previously recorded defects, it was concluded that a maintenance action has taken place at that time and location.

Table 1. Red tag and yellow tag limits according to 2015 INFORMS Rail Applications Section.[23]

	FRA track class				
	Class 1	Class 2	Class 3	Class 4	Class 5
Passenger speed (mph)	15	30	60	80	90
Freight speed (mph)	10	25	40	60	80
Surface62					
Red tag limit (in.)	3	2 3/4	2 1/4	2	1 1/4
Yellow tag limit (in.)	2 3/4	2 1/2	2	1 3/4	1
Dip31					
Red tag limit (in.)	3	2 3/4	2 1/4	1 3/4	1 1/2
Yellow tag limit (in.)	2 1/2	2 1/4	1 3/4	1 1/2	1 1/4
Crosslevel					
Red tag limit (in.)	3	2	1 3/4	1 1/4	1
Yellow tag limit (in.)	2 3/4	1 3/4	1 5/8	1 1/8	7/8

Fine-scale deterioration modeling

At the fine-scale level, we use survival analysis to model the lifetime of yellow tag defects. The competition dataset is used to calibrate the survival analysis model. As mentioned earlier, the dataset contains a sequence of recorded defects for a given location, e.g. $\{Y_1, Y_2, R\}$, where Y_1 and Y_2 refer to two yellow tag defects and R to a red tag defect. A data point used for survival analysis should contain a beginning state and an end state. Instead of considering the sequence $\{Y_1, Y_2, R\}$ to be a single data point (Y_1, R), we turn it into two data points (Y_1, Y_2) and (Y_2, R), thus enhancing the calibration procedure. To analyze the lifetime of a data point with repeated yellow tags, e.g. (Y_1, Y_2), we assume that the lifetime of the yellow tag defect starts from the first report. This assumption is reasonable if the initial amplitude of the defects is considered as an explanatory variable. The model aims to predict the lifetime of a defect after its amplitude reached a certain value (defect's amplitude in the first report). If the defect did not turn into a red tag by the time of the second report, it is considered as a right-censored record, implying that the lifetime of the defect is unknown. In other words, the only known information is that the defect survived until the second report. Even for uncensored records, when the second inspection indicates red tag, the exact failure time is unknown. In this work, we assume that the failure has happened midway between the two yellow and red tag reports.

As mentioned, the dataset includes the records of three types of defects. In general, these defects have different deterioration processes. Therefore, a survival model is developed for each type of defect. To identify repeated defects, we considered a defect from the same type within 10 feet on either side of the previous defect to be a repeated defect.

We include the initial absolute values of amplitude and length of the defects, track code, class of track, operating speeds, and tonnage as explanatory variables in the model. The track code can be either tangent or curve state (i.e., tangent or curve the dataset does not include the spiral case). The track code parameter is set to 0 for tangent code and 1 for curve code. Since merely about 1% of the dataset belonged to classes other than 4 or 5, we only include the data from Class 4 or Class 5 tracks, with the value of the class parameter set to 0 for Class 4, and 1 otherwise. It should also be noted that the processed dataset for surface defects includes mostly defects within Class 5 segments and contains very few defects within Class 4 segments. Therefore, in surface defects survival

Table 2. Parameter estimation for dip, surface, and crosslevel survival model.

Defect	Dip31		Surface62		Crosslevel	
Variable	Coefficient	z-Score	Coefficient	z-Score	Coefficient	z-Score
Constant	17.17	7.82	13.11	10.97	15.46	11.39
Amplitude	−7.63	−5.73	−7.39	−7.00	−9.33	−8.17
Class5	−2.02	−5.19	NA	NA	−2.43	−8.02
1/*p*	0.78	−2.98	0.79	−3.45	0.68	−8.14

model, we only developed a model for Class 5 tracks, and considered that to be also applicable to Class 4 tracks.

Table 2 includes the estimated parameters for the survival models for dip, surface, and crosslevel defects, following the procedure in the Parameter estimation section. Only variables with a *z*-score less than −1.96 or greater than 1.96, i.e. statistically significant variables at 5% level, are included in the table. The *z*-score is a measure of statistical significance which indicates whether or not to reject the null hypothesis of the corresponding coefficient being zero. It can be seen that the signs of coefficients are intuitively correct, i.e. an increase in variables with negative coefficient would lead to increasing the likelihood of a shorter lifetime. For example, Class 5 tracks have smaller red tag thresholds than Class 4 tracks do. Therefore, a defect with a fixed amplitude is expected to have a shorter lifetime in a Class 5 track compared to that in a Class 4 track. This is reflected in the negative coefficient value for Class 5 variables. Also, one expects defects with larger amplitudes within a segment to have a shorter lifetime, which is consistent with the values shown in Table 2. The *z*-scores for the estimated values of *p*, i.e. the shape parameter of Weibull distribution, show that the *p* values are significantly greater than one at 5% level. This means that the hazard rates are monotonically increasing and the probability of failure increases with the time spent without failure. This agrees with our expectation that physical infrastructures are more likely to fail with time.

The fact that track code, tonnage, and operating speeds are found to be statistically insignificant variables does not necessarily imply that they do not impact the deterioration process. We should consider that only about 5% of the defects are located in curve segments. This amount of data may not be sufficient to inform the impact of being on a curve segment on the deterioration process. Also, about 95% of defects are located in segments with an average monthly tonnage between 4 and 7 MGT. The small variety in the average monthly tonnage in the data does not allow for a precise identification of its impact on the deterioration rate. To validate the calibrated survival models, an experiment is developed in which half of the data is used as training subset and the other half as test subset. We use the training subsets to estimate the parameters of survival functions. Then, we predict the failure probability for the test subsets based on estimated parameters. If the failure probability was greater than or equal to 0.5, we predict a failure; otherwise, we predict survival. All three survival models make more than 70% correct predictions for the test subsets, which shows the validity of the models.

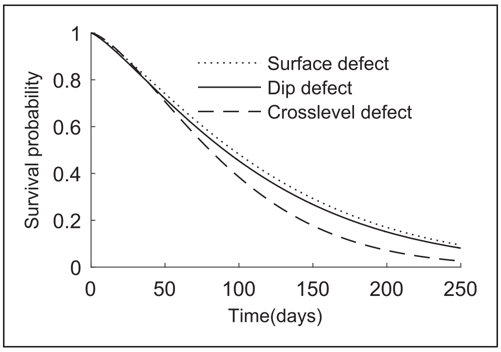

Figure 2. Survival probability for dip, surface, and crosslevel defects located in a Class 5 segment with initial amplitudes equal to nine-tenth of the safety limits.

Figure 2 shows the survival probability for dip, surface, and crosslevel defects located in a Class 5 segment with initial amplitudes equal to nine-tenths of the safety limits. It can be seen that the deterioration rate of the three types of defects are close to each other. Crosslevel defect has the fastest rate of deterioration followed by dip defect and surface defect.

Fine-scale survival models provide the probability of individual yellow tag defects turning into red tag defects. However, railroad companies may prefer to perform segment-wise preventative maintenance. In other words, railroad companies need a deterioration model that predicts which segments will need maintenance, that is which segment will have at least one red tag defect. The question is how to use fine-scale defect-based measurement to make reliable predictions at the segment level. The naive approach would be to maintain the segments that include at least one yellow tag defect with failure probability higher 50% or a certain limit. This approach overlooks the probability of having a red tag defect in segments which include multiple yellow tag defects with low-failure probability. Using this approach, a segment with one yellow tag defect with 70% failure probability would be maintained, while a segment with four yellow tag defects with 30% failure probability would not be maintained. However, the probability of the occurrence of at least one red tag defect is larger in the latter case.

Reliable estimations of the failure probability of segments, probability of observing at least one red tag defect in a segment, can be used by railroad companies to make proper segment-wise maintenance decisions. We use fine-scale models to estimate the failure probability of segments. Failure probability of segments is approximated by calculating the union of failure probabilities of the existing yellow tag defects within the segments. Since red tag defects are more likely to occur at locations of yellow tag defects and fine-scale defect-based model accounts for failure probability of yellow tag defects, it is expected that the defect-based model outperforms coarse-scale (segment-based) models reviewed in the Background on modeling track geometry degradation section. Defect-based model is found to make more than 70% correct predictions. Later in the next section, the results from defect-based model are compared with the results from coarse-scale model in detail.

Coarse-scale deterioration modeling

In the previous section, we demonstrated how a fine-scale defect-based model can be used to forecast the lifetime of yellow tag defects. We also showed how a prediction at the segment level can be made using information from individual defects. In this section, we introduce a coarse-scale, or segment-based, survival model consistent with how the quality of a track segment is currently measured in practice. We assume a given track segment, depending on the geometry defects it contains, can be in one of these two condition states: a 'good' condition (G) or a 'failed' condition (F). A track segment containing only the yellow tag defects and no red tag defect is considered to be in the 'good' condition. If the segment contains one or more red tag defects, it is considered to be in the failed condition. A sequence of observed state conditions for a track segment, then, can be in this form: $\{G, G, \ldots, F, F, \ldots, F\}$. When a failure happens, i.e. at least one defect in the segment turns into red tag, and the track segment will transit to and remain in the 'failed' condition until a corrective maintenance action is taken.

Survival analysis is then used to estimate the time after which a segment in 'good' condition will transit into a 'failed' condition. These survival models are built using coarse-scale indicators for segments as explanatory variables, and details on individual defects within the segments are not incorporated as explanatory variables. Two of these coarse-scale indicators most widely used in practice are the K value[15] and TQI.[12] In K value K is the ratio of the total length of the segment above the yellow tag limit (l) to the total length of the track segment (L)

$$K = \frac{l}{L} \times 100 \tag{15}$$

TQI, denoted here by q, is defined to be the summation of standard deviations of seven geometry parameters in a segment

$$q = \sum_{i=1}^{7} \sigma_i \tag{16}$$

where σ_i is the standard deviation of ith geometry parameter, given by

$$\sigma_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n} a_{ij} - \overline{a_i}^2} \tag{17}$$

where n is number of sampling points in the segment, $\overline{a_i}^2$ is the average of amplitude of parameter i over sampling points, and a_{ij} is the ith parameter amplitude at the sampling point j. Since the data only contain the information on the defects of the three geometric parameters, we had to modify the TQI definition to be the summation of standard deviations of amplitudes of dip, surface, and crosslevel defects over a segment.

In fine-scale models, the amplitude of defects was found to be a significant variable in estimating the lifetime yellow tag defects. However, the coarse-scale indices, K value, and TQI, by definition, do not incorporate the amplitude of the yellow tag defects within the segment. This may lead to a poor performance by these indices in capturing the underlying deterioration process. To investigate this, we define a new index as the summation of relative amplitudes, SRAs, denoted by r which directly incorporates the amplitudes within a segment, according to

$$r = \frac{A_s}{T_s} + \frac{A_d}{T_d} + \frac{A_c}{T_c} \tag{18}$$

where A_d, A_s, and A_c are the summation of amplitudes of the dip, surface, and crosslevel defects within the segment, respectively, and T_d, T_s, and T_c are red tag thresholds for surface, dip, and crosslevel defects, respectively.

Next, we compare the validity of these three coarse-scale indices. We form three coarse-scale models, where each uses one of the three indices as an explanatory variable. Table 3 tabulates their calibrated parameters when the size of segments is considered to be 0.5 mile. It should be noted that the Class variable was

Table 3. Parameter estimation for coarse-scale survival models for segments with 0.5 mile length.

	K value		TQI		SRA	
Variable	Coefficient	z-Score	Coefficient	z-Score	Coefficient	z-Score
Constant	4.95	92.81	4.73	104.11	5.16	77.20
Quality index	−1.54	−14.11	−10.67	−8.82	−0.42	−12.61
1/*p*	0.72	−8.27	0.75	−7.32	0.73	−8.12

found to be an insignificant variable in these coarse-scale models. The reason is that the values of all three indices used as explanatory variables are indirectly dependent on the class of the track. In other words, for two tracks with identical characteristics except the track class, the quality indices such as *K* value, TQI, and SRA will take different values.

The same approach as for fine-scale models is used to validate the coarse-scale models. Coarse-scale models with *K* value, TQI, and SRA as the explanatory variables make 65%, 62%, and 65% correct predictions for test data, respectively. It can be concluded that TQI is the least informative quality index and SRA index is not any more informative than *K* value index. One way to explain this is that similar to *K* value, SRA does not inform the model about the amplitude of individual defects, and rather incorporate an aggregate value. For instance, a large SRA value could refer to one defect with large amplitude or several defects with small amplitudes.

In the next section, fine-scale and coarse models are compared. We include *K* value as the quality index in coarse-scale model as it is commonly used and is more informative than TQI.

Comparison between fine-scale and coarse-scale survival models

To investigate the comparison between fine- and coarse-scale survival models, we study the coarse-scale models at various spatial scales, i.e. given different segment lengths. Figure 3 shows the percentage of correct predictions for fine- and coarse-scale models versus segment lengths. It can be seen that fine-scale model outperforms the coarse-scale model, especially when the segment length is large. One possible reason is that as the length of the segments increases, the potential number of defects within the segments also increases, and subsequently, the *K* value becomes less informative. That is because for a short segment, the *K* value is likely to capture the length of the only defect within the segment. Longer segments, however, are more likely to contain more than one defect. Therefore, the *K* value index for a longer segment is more likely to refer to the total length from multiple of defects. In other words, the information that is lost when the length of the segment increases is the number of defects. This information is independent from the *K* value and significantly influences the overall failure probability of the segment. Thus, we recommend to use a fine-scale model for segment-based prediction, especially when the segment length is large.

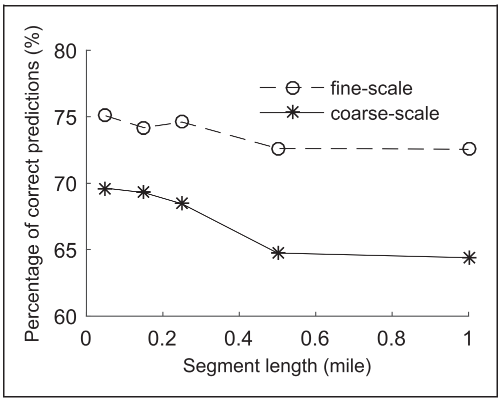

Figure 3. The percentage of correct predictions for fine- and coarse-scale models versus segment lengths.

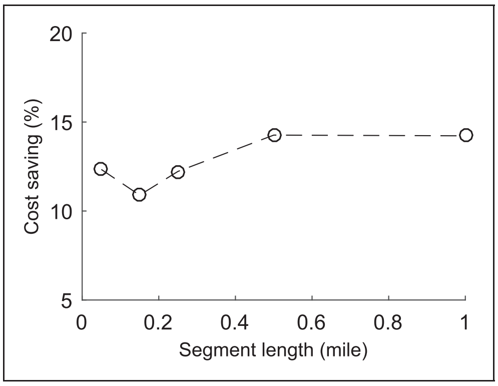

Figure 4. Percentage of maintenance cost reduction in case of using fine-scale model instead of coarse-scale model versus segment length.

We can also quantify the cost that will be saved by using fine-scale model in making maintenance decisions instead of coarse-scale model. For cost quantification we make the following assumptions: (a) the whole segment is maintained in both planned and unplanned maintenance, (b) an unplanned maintenance costs 1.5 times a planned maintenance for the same segment length, and (c) a planned maintenance is done if the model predicts a failure, and an unplanned maintenance is needed if the model predicts false survival for a segment. Figure 4 shows the percentage of maintenance cost reduction as a result of using fine-scale model in the prediction of segment failure. It can be seen that the cost saving is more than

10% and is significant. It should be noted that we did not consider the cost associated with derailment risk in the cost calculations. Cost saving would be even larger if we consider the potential derailment costs.

Conclusion

We have applied the survival analysis methodology to develop a forecast model for the deterioration of track geometry, and demonstrated its validity in predicting the future defect conditions based on actual measurement data collected from track geometry cars provided by the committee organizing the 2015 INFORMS RAS Problem Solving Competition. A survival model proved efficient, as it fully depends on the data, and does not necessitate additional information from experts, which is subject to simplification or human errors. We also investigated how different choice of spatial scales for the survival model may lead to erroneous predictions. In particular, we considered the case where predictive models are developed for track segments instead of individual defects. This is particularly motivated by the current practice of track maintenance where maintenance is done for the whole segments and not individual defects. We compared the fine-scale defect-based model with coarse-scale segment-based survival model and we found that fine-scale model outperformed the coarse-scale model. Therefore, to predict the probability of having at least one red tag defect in a segment, we recommend using fine-scale defect-based survival models.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Martland CD, Paul L and Youssef K. Productivity improvements in the US Rail Freight Industry, 1980–2010. *J Transport Res Forum* 2012; 51: 83–107.
2. Liu X, Saat M and Barkan C. Analysis of causes of major train derailment and their effect on accident rates. *Transp Res Rec* 2012; 2289: 154–163.
3. He Q, Li H, Bhattacharjya D, et al. Track geometry defect rectification based on track deterioration modelling and derailment risk assessment. *J Operational Res Soc* 2014; 66: 392–404.
4. Shenton MJ. Ballast deformation and track deterioration. In: *Track technology*. London: Thomas Telford Publishing, 1985, pp.253–265.
5. Sato Y. Optimum track structure considering track deterioration in ballasted track. In: *Proceedings sixth international heavy haul railway conference*, Cape Town, South Africa, 6 April 1997, pp.6–10.
6. Chrismer S and Selig E. Computer model for ballast maintenance planning. In: *Proceedings of 5th international heavy haul railway conference*, Beijing, China, 6 June 1993, pp.223–227.
7. Faiz R and Singh S. Information analysis of rail track for predictive maintenance. *WSEAS Trans Comp* 2009; 8: 1123–1133.
8. Famurewa SM, Tao X, Matti R, et al.. Comparative study of track geometry quality prediction models. In: *International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, 18–20 June 2013, vol. 2, pp.1057–1068.
9. Westgeest F, Dekker R and Fischer R. Predicting rail geometry deterioration by regression models. *Adv Saf Reliab Risk Manag ESREL* 2011; 146: 926–933.
10. Andrade AR and Teixeira PF. A Bayesian model to assess rail track geometry degradation through its life-cycle. *Res Transport Econ* 2012; 36: 1–8.
11. Meier-Hirmer C, Riboulet G, Sourget F, et al. Maintenance optimization for a system with a gamma deterioration process and intervention delay: application to track maintenance. *Proc IMechE, Part O: J Risk and Reliability* 2009; 223: 189–198.
12. Bai L, Liu R, Sun Q, et al. Markov-based model for the prediction of railway track irregularities. *Proc IMechE, Part F: J Rail and Rapid Transit* 2015; 229: 150–159.
13. van Noortwijk JM and Frangopol DM. Deterioration and maintenance models for insuring safety of civil infrastructures at lowest life-cycle cost. *Life-cycle performance of deteriorating structures: assessment, design and management ASCE*. Reston: VA, 2004, pp.384–391.
14. Cleves M. *An introduction to survival analysis using Stata*. College station, Texas: Stata Press, 2008.
15. Larsson-Kråik PO, Nissen A, Juntti U, et al. Optimisation of track geometry inspection interval. *Proc IMechE, Part F: J Rail and Rapid Transit* 2014; 228: 546–556.
16. Pinder JE III, Wiener JG and Smith MH. The Weibull distribution: a new method of summarizing survivorship data. *Ecology* 1978; 59: 175–179.
17. Nelson WB. *Applied life data analysis*. vol. 577. New York: John Wiley & Sons, 2005.
18. Kleinbaum DG and Klein M. Parametric survival models. In: *Survival analysis*. New York: Springer, 2012, pp.289–361.
19. Mishalani RG and Madanat SM. Computation of infrastructure transition probabilities using stochastic duration models. *J Infrastruct Syst* 2002; 8: 139–148.
20. Mauch M and Madanat S. Semiparametric hazard rate models of reinforced concrete bridge deck deterioration. *J Infrastruct Syst* 2001; 7: 49–57.
21. Aitkin M and Clayton D. The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *J Appl Stat* 1980; 29: 156–163.
22. Weinan E and Engquist B. Multiscale modeling and computation. *Notice AMS* 2003; 50: 1062–1070.
23. 2015 INFORMS Rail Applications Section (RAS) Problem Solving Competition. Available at: https://www.informs.org/Community/RAS/Problem-Solving-Competition/2015-RAS-Problem-Solving-Competition (accessed 25 February 2017).