Degradation functions for railway station equipment quality based on maintenance-influenced data

Maren Maus[1], Felix Lampe[1], Nadine Friesen[1], Mark Heckmann[2], Lea Elfert[2] and Nils Nießen[1]

Abstract— **To find the optimal maintenance policies, the DB Station&Service AG – a railway infrastructure company which manages the majority of the railway stations in Germany – needs to describe the cause-effect-relationship between the funds for maintenance and the quality of the station equipment. To determine the influence of funding on the infrastructure quality, it is necessary to predict the cost and effect of maintenance measures as well as the behaviour of unmaintained items. However, the available degradation data do not contain any condition measurements of items which are not influenced by maintenance. Here, we provide a model which allows to describe the average maintenance intervals and the degradation during these intervals based on such maintenance-influenced data. This model can be used to predict the development of the equipment quality and is intended to be used to determine the necessary budget for a specific equipment quality or vice versa.**

I. INTRODUCTION

Up to 2009 each replacement in the German railway network had to be arranged separately between the German Government and the German railway infrastructure companies, DB Netz AG, DB Station&Service AG and DB Energie GmbH. To reduce this administrative effort, a contract between the railway infrastructure companies and the German state was established (Leistungs- und Finanzierungsvereinbarung (LuFV)) [11].

Since then, the DB Station&Service AG and the other infrastructure companies receive funds from the German government for the replacement of infrastructural elements in one amount and distribute it at their own discretion. For maintenance, DB Station&Service AG is obliged to invest their own funds. Replacement and maintenance ensure the equipment condition at a high quality level. Note that throughout this paper the term equipment refers to the equipment as a whole, i.e., the entire equipment of a station, a region, or country whereas the term item is used for a specific component. The equipment quality is measured by penalized and non-penalized quality indicators. If certain quality requirements are not met for penalized quality indicators, penalty payments to the German government are required. In the following only penalized quality indicators are considered.

The equipment of the DB Station&Service AG consists of items from different classes (see Table I). Each item is composed of subcomponents. The types of subcomponents an item has depend on its class. The quality indicator of each item, for example, platforms, staircases, etc., is a grade between 1 and 6. Hereby, 1 is the best and 6 the worst grade. To calculate an item's grade, a manual assessment in which the item's subcomponents receive a score based on their condition is conducted. The weighted sum of the scores across all its subcomponents is then converted into the item's grade. Based on the grades for each item, a grade for a station, a region, or the whole network can be calculated as a weighted average. The network-wide grade is then compared to the grade necessary for the achievement of the agreed quality targets.

Since no direct relationship between the amount of funds used and the quality to be expected exists, the government and the infrastructure companies agreed on the development of a tool to determine the cause-effect-relationship between the equipment quality and the funds in the LuFV III [12]. To anticipate the amount of funds needed for a given network-wide grade and reversely, we need to predict the frequency and time of maintenance as well as the item condition.

The given data provide information on grades of each item on a regular basis and additionally after maintenance measures. These data cover 13 years and we do not know how many and which maintenance measures each item has had before this time period. The data also do not contain any information about how an item would degrade without any maintenance measures.

The contribution of this paper is to provide a data-driven method to model the average degradation and maintenance time points of an item where the data do not contain any information about items without maintenance.

This paper is structured as follows: In the second chapter, we provide a short overview of the literature considering data-based degradation functions and maintenance planning in the railway sector as well as a broad comparison to other areas. In Chapter III we indicate the requirements the data has to fulfill for our approach to be applicable and describe the details of our use case. Afterwards, we explain our handling of the maintenance influence in the data. In Chapter IV, we show the obtained degradation function. Finally, we conclude this paper and provide suggestions for further research.

II. LITERATURE REVIEW

The infrastructure companies in the German railway sector have to provide a cause-effect-relationship between the infrastructure quality and the funds they partly receive from

[1] Institute of Transport Science (VIA), RWTH Aachen University, Mies-van-der-Rohe-Str. 1 52074 Aachen, Germany [last name]@via.rwth-aachen.de

[2] DB Station&Service AG, Washingtonplatz 2 10557 Berlin, Germany [first name].[last name]@deutschebahn.com

*This work is partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 2236/1 and NI 1597/4-1.

the government. This relationship has first been examined for the DB Netz AG which is responsible for the railway tracks, bridges, etc., in Germany in [5].

For this cause-effect-relationship, the development of the equipment condition over time and the influence as well as the need of maintenance measures have to be known. Therefore, degradation functions are required. These have been widely studied in different contexts. For example, ordinal regression functions have been determined in [1] for bridges whose condition is represented by grades. A review of different options to describe the degradation of railway tracks is given in [3].

Neuhold describes in [7] how to prepare track condition data and obtains linear regression functions for the standard deviation of the track geometry.

Saadat analyses the mean-time between failures for track-circuits in [9].

Quiroga states a mathematical optimization approach to maintenance scheduling for railway tracks and develops a heuristic to solve the problem in [10]. In [6], Letot uses a random coefficient Wiener degradation-based process to model track degradation. The effect of maintenance is described by a probabilistic model and the maintenance costs are assessed with a Monte-Carlo approach. In [8], Prescott models the deterioration of the ballast by a Markov process and also considers the deteriorating effect of the maintenance measures to analyse the effects of different maintenance strategies.

In railway stations, items of many different classes are used. Some of these can be compared to structures used in other contexts. For example, platforms can be compared to sidewalks whose deterioration was analysed in [2] with finite element analysis. Heitel compares life cycle cost analysis for pedestrian overpasses in [4].

To the best of the authors' knowledge, no degradation functions for the equipment in railway stations have been given yet.

III. METHODS

To predict the required amount of financial resources for a specific quality or vice versa, it is necessary to know the average degradation of the station equipment as well as the effect of maintenance measures. Here, we use a data-driven approach to model the average life cycle of different types of station equipment. In this section, we first describe the requirements the data have to fulfill in order for our approach to be viable in a general context and then present the details of the implemented approach.

A. Data requirements

The model we derive in this section quantifies the effect of maintenance on the age induced degradation of constructional objects by a data-based approach. The data at the basis of this model needs to consist of several time series per class of item to be modeled. Each time series needs to contain the item's age and quality at the time of measurement. We furthermore need to be able to distinguish regular quality measurements from those taken right after maintenance and replacement. A last important factor is an item's estimated service life which indicates the age after which it is usually replaced.

TABLE I

LIST OF THE 21 EQUIPMENT CLASSES AND THEIR SERVICE LIVES.

no.	class name	service life (years)
1	platform halls	80
2	pedestrian overpasses	95
3	pedestrian underpasses	100
4	floors and staircases in public areas	50
5	entrance doors	23
6	flat roofs	104
7	steep roofs	117
8	facades	63
9	windows	32
10	underground passenger stations	100
11	walls	67
12	platform roofs	80
13	ramps on platforms	80
14	staircases on platforms	90
15	platforms	70
16	passenger information systems	12
17	passenger elevators	15
18	escalators	13
19	dodgers	40
20	windbreaks	40
21	lighting poles	40

In particular we applied our approach to the data from DB Station&Service which contains data from 21 different classes (see Table I for the full list) and covers a span of 13 years.

The item's quality is indicated by a grade between 1 and 6, where 1 specifies a very good condition and 6 corresponds to the poorest possible quality. According to [12] the grades are assessed in the following steps: First, each subcomponent (e. g., the platform's floor or the platform's guide stripes for the blind in the class of platforms) of an item is evaluated separately, yielding a so-called base value for each subcomponent. These base values are then weighted according to the corresponding subcomponent's importance for the item's functionality. The weighted sum of the base values of all subcomponents yields the item's score, which is given by a number of points between 0 and 240. A low score corresponds to a good condition whereas a high score corresponds to a poor one. In the last step, the calculated scores are converted into the mentioned grade between 1 and 6. This nonlinear conversion is illustrated in Fig. 1.

Further, it is possible to calculate a grade for a whole class by forming the arithmetic mean of all items of this class in a station. The grade for a whole station is then calculated as a weighted mean of the class grades for the different object classes. In the same manner, a grade for a whole region or the entire country can be calculated as the sum of the grades for the single stations, weighted by the station's importance in terms of the number of travelers, in each case.

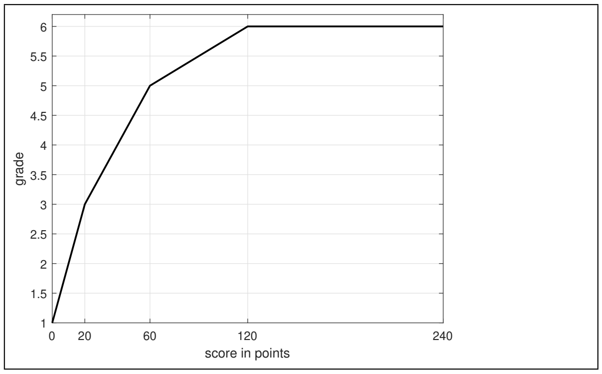

Fig. 1. **Relationship between an item's grade and score.** The plot illustrates the last step in the process of aggregation of an item's grade: the conversion from the score (number of points from 0 to 240) into the final grade (1 to 6). In both cases, the lowest value corresponds to the optimal condition and the highest value to the poorest possible condition.

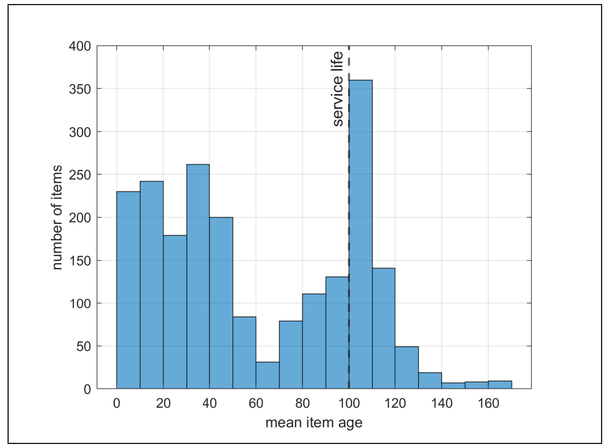

Fig. 2. **Distribution of the mean item age.** The histogram illustrates that the majority of items is younger than the service life, here for the example of pedestrian underpasses.

B. Grade development including maintenance measures

For the development of the equipment life cycle, we work with the measurements of the single items on the level of scores.

In a first step, we model the score development of an average item over time for each class.

A main challenge in the model development is that the number of items decreases significantly with increasing age. This is mainly a result of the intended replacement of an item as soon as it reaches its service life. However, replacement resources are limited and usually do not suffice to replace all items that are older than their service life. On the one hand, data referring to an age larger than the service life of the corresponding class therefore comprise a significantly smaller amount of items than the data corresponding to ages smaller than the service life: Taking into consideration all classes in the case of DB Station&Service, 87.1% of the 72 425 items in total have a mean age smaller or equal to their service life whereas 12.9% are older than their service life (see Table I for the service lives of the single classes). Fig. 2 shows the complete distribution of the mean item age for the class of pedestrian underpasses. It can be seen that the number of items significantly decreases for higher ages with a delay of a couple of years after the service life. In line with the results for all classes, 72.7% of the 2 142 items in this class have a mean age smaller or equal to the service life (100 years) whereas 27.3% are older than the service life. On the other hand, the limited replacement resources provoke a survivor ship bias: Items that are in the worst condition are replaced first while items of the same age in a good condition are not replaced.

As a result of these two effects, the data after the service life comprise only few items which are additionally in a disproportionately good condition. The data after the service life is thus not representative for the whole class. Therefore we only include data up to the service life to model the score development of an average item over its age. The model incorporates all types of measurements, i.e., also measurements after maintenance measures at this stage.

We perform a logarithmic regression on the score dependent on the equipment age, using the model

$$a \cdot \log(bx + 1) \tag{1}$$

and the method of least squares, with $a \in \mathbb{R}$ and $b \geq 0$.

Fig. 3 illustrates the resulting logarithmic fit for the example of pedestrian underpasses (full line). The data points represent the average score over the item's age.

C. Handling of maintenance influence

The current model so far only describes the average development of an item including its maintenance measures. Since there are only limited data available and there are items which are older than the time span for which data was collected - meaning that the history of the station equipment is not completely available -, the influence of past maintenance measures remains an unknown factor. This is revealed by the distribution of the item score averaged over age: the gradient of the curve declines over time, with a clear break at the service life (see data points in Fig.

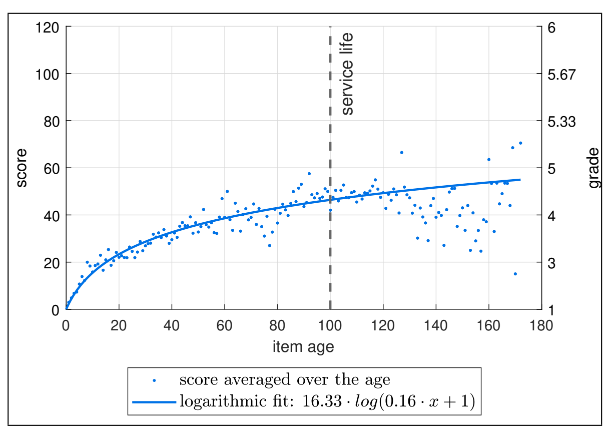

Fig. 3. **Development of an average item's quality over time.** Plot of the item's quality given by a score (left y-axis) or the corresponding grade (right y-axis) averaged over age for all available measurements in the class of pedestrian underpasses (dots) and the derived model for the average development (line).

3). Consequently, the poorest grade is never reached in the average of the data. To ensure that this effect is not a result of an increasing amount of maintenance measures after the service life has passed, we also examine the data excluding the measurements directly after maintenance as well as the distribution of the number of maintenance measures over age. However, the described effect persists. To account for the difficulty that the data cannot be divided into pure degradation and pure maintenance data, we describe the life cycle by explicitly including the maintenance measures in our model.

First, we determine the points in time at which an item on average receives a maintenance measure: For every age x, we calculate the fraction p of items that has obtained maintenance at this age via

$$p(x) = \frac{n_{\text{maintenance}}(x)}{N_{\text{items}(x)}}, \tag{2}$$

where $n_{\text{maintenance}}(x)$ denotes the number of maintenance measures on items of age x and $N_{\text{items}}(x)$ the number of items that have age x at any time point within the investigation period. Thereafter, we derive the cumulative frequencies at each age from these fractions:

$$F(x) = p(\text{age} \leq x) = \sum_{k=0}^{x} p(\text{age} = k). \tag{3}$$

By setting $F(x) = y$ and resolving for the age x, the ages of the y^{th} maintenance measure can be obtained. This procedure is illustrated in Fig. 4 which shows the cumulative frequencies for the example of pedestrian underpasses and indicates how times of maintenance are obtained from this function.

In addition to the times of maintenance, the average effect of a maintenance measure is needed. For this purpose, we evaluate the scores of the items at their last assessment before a maintenance measure is applied, and of their first assessment thereafter. This data can be best modeled by

$$\max(0, a \cdot \log(bx + 1) + c) \tag{4}$$

with $a, c \in \mathbb{R}, b \geq 0$ in the logarithmic function. The maximum ensures that we do not obtain negative scores. The evaluation of the condition at the last assessment before (red) and after (green) a maintenance measure is depicted in Fig. 5. The average development of the item score including maintenance is shown as in Fig. 3 (blue).

Together with the maintenance times, the evaluation allows assigning a specific effect to each maintenance measure in terms of the difference between the number of score points of an item's condition before a maintenance measure and the one thereafter.

IV. Results

In this section, we use the gained information to build a full life cycle model that is capable of reproducing the average item behavior within one class and can be applied to all the different classes distinguished in the data that describes the use case we apply our method to. For this purpose, as described in Subsection III-C, the scores before and after a maintenance measure are evaluated. These scores define the upper boundary function (red) and the lower boundary function (green) in Fig. 6. The upper boundary function (red) shows the intervention threshold. All items whose score is above this function for the respective age should receive either a maintenance or a replacement measure. The need for measures is divided into the need for maintenance and the need for replacement. The distinction is drawn based on the service life of an item: All items that have a score exceeding the intervention threshold and are older than their service life are considered to be in need of replacement. If the item score is above the intervention threshold but the item has not exceeded its service life, it is considered to be in need of

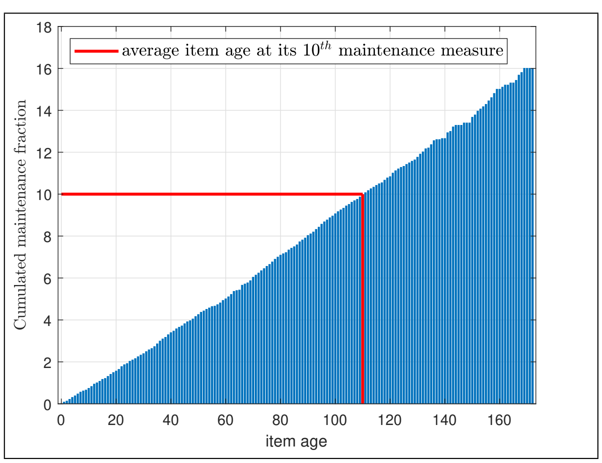

Fig. 4. **Cumulative maintenance frequencies.** Fraction of items having obtained maintenance at each age for the class of pedestrian underpasses plotted cumulatively. By this procedure, the average time points for maintenance measures within a class can be determined. The red lines, for example, indicate that the 10^{th} maintenance measure for a pedestrian underpass was performed at an age of 129 years on average.

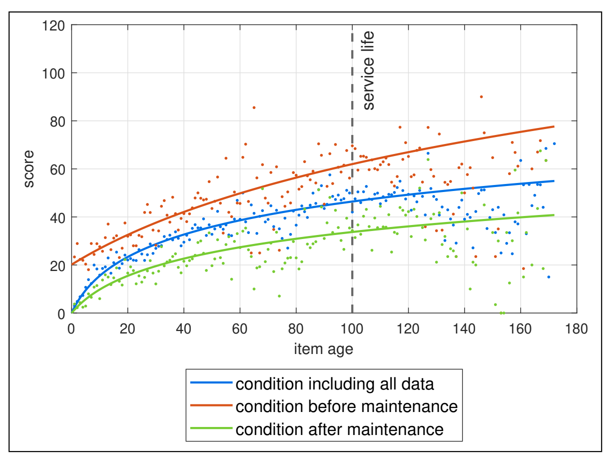

Fig. 5. **Evaluation of the item condition at the last assessment before and just after a maintenance measurement.** Plot of the average condition before (red) and after (green) a maintenance measure. The dots indicate the averaged data points whereas the full lines correspond to the performed logarithmic regression. The development of the average score over age, including regular measurements and measurements before/after maintenance, is depicted as in Fig. 3 (blue).

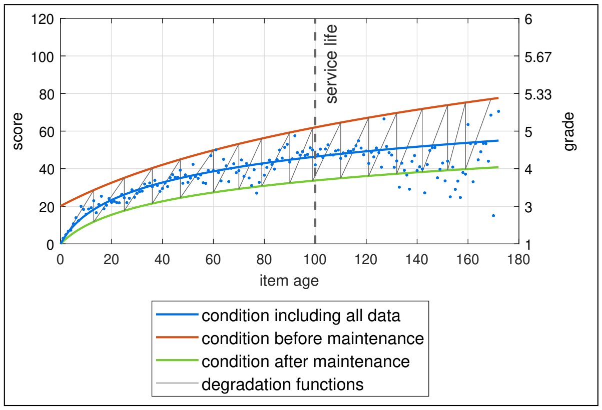

Fig. 6. **Life cycle function.** Plot of the intervention threshold (red) and of the condition after a maintenance measure (green). The blue line corresponds to the performed logarithmic regression including all available data. The degradation function of an average equipment is indicated by the gray lines.

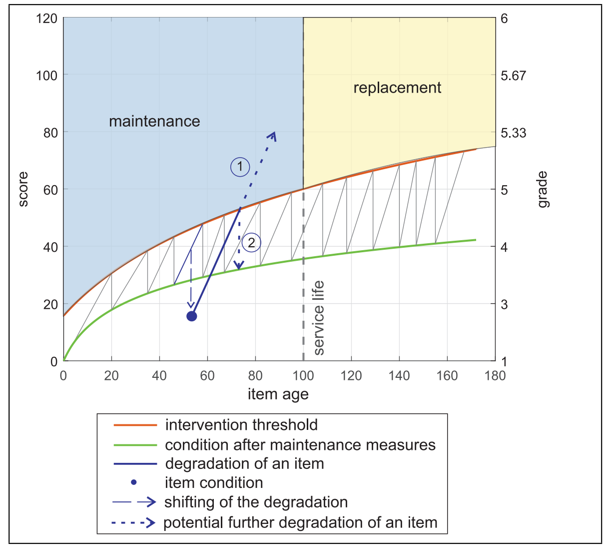

Fig. 7. **Assignment of degradation functions and classification of measures.** Plot of the intervention threshold (red) and of the condition after a maintenance measure (green). The purple dot indicates a sample item. The degradation of the equipment according to its age is shifted to the item (purple arrow). At the intersection with the intervention threshold, the item can either deteriorate further based on the shifted degradation function (dashed purple arrow 1) or receive a maintenance measure that improves the item's score to the lower boundary function (dashed purple arrow 2).

maintenance. The lower boundary function (green) describes the condition of the item after a maintenance measure has been carried out. The effect of a maintenance measure can hence be derived from the difference between the upper and lower boundary functions. After a replacement, the model assigns the best possible condition and age zero to the new item.

The number of maintenance measures and their timing are determined as described in Subsection III-C. Based on this, we know when to schedule a maintenance measure in the model on average. The effect of a maintenance measure is determined as described in the previous paragraph. The degradation between two maintenance measures, i. e., the development from the condition of an item after the first maintenance measure to the condition before the next measure has to be applied, is linearly interpolated. At the beginning of the life cycle or after the construction of an item, the best possible condition is assumed. The complete modeled average degradation functions and the improvements caused by maintenance measures between the two boundary functions are obtained as shown in Fig. 6 (gray). The figure indicates that the average item development including maintenance measures (blue) can be well described by this life cycle modeling.

To model the condition of all items and the influence of maintenance measures it is not only essential to describe the degradation of an average item but to be able to predict the future score for any item. Therefore, each item is assigned to the slope of the average degradation function according to its age as shown in Fig. 7. The slope of the corresponding function is applied to the initial condition of the considered item allowing the prediction of the item's degradation (purple arrow in Fig. 7).

At the intersections of the predicted degradation with the upper boundary function (red), the model schedules maintenance measures. A measure improves the score of the item to the value of the lower boundary function (green) regardless of the previous score. This case is indicated by the second dashed purple arrow in Fig. 7. If no maintenance measure can be carried out, the score of the item will deteriorate based on the assigned slope even beyond the intervention threshold until a measure is scheduled which is indicated by the first dashed purple arrow in Fig. 7. If the model indicates that an item is to be replaced, but the budget allocated for replacement is not sufficient, a maintenance measure can be scheduled to improve the item's condition given that there is still budget available for maintenance measures. Therefore, further maintenance measures beyond the service life are also included in the model.

V. Conclusion

To be able to predict the condition of DB Station&Service AG's equipment, the available data was analyzed and evaluated. This resulted in a logarithmic function being the most appropriate for describing the development of the equipment condition including maintenance measures. Since only data from 13 years and no further historical data are available, the influence of maintenance measures which led to a reduced average degradation at older ages could not be completely eliminated from the degradation functions.

Therefore, a life cycle function, which takes maintenance measures into account, was developed so that the quality development of an average item can be well described. For the description of the life cycle, an intervention threshold and a function for describing the item condition after a maintenance measure were derived from the data and implemented in the model. Within these two boundary functions, the life cycle of an average item was defined. To predict the future condition of an item with an arbitrary initial condition, the average item degradation is shifted parallel so that it starts at the condition of the considered item. In addition to degradation, the need for maintenance measures can be derived from the model.

The model can hence be used to predict the future condition of DB Station&Service AG's equipment and thus the network-wide grade, taking into account various financial resources for maintenance and replacement. This prediction is a valuable contribution to the optimization of maintenance policies.

The data-driven approach allows a straight forward generalization to fields where time-resolved information of item quality is available such as road networks and different kinds of constructional objects.

VI. Future research directions

Since data about the type of maintenance measures and the number of previous maintenance measures are not available, we do not consider the effects of previous maintenance measures and of combinations of maintenance measures in this paper. Including this information could lead to a more differentiated view on the effect of maintenance measures onto the immediate condition improvement and the deterioration rate between the maintenance measures.

A further simplification of our model could be achieved by replacing the dependence on explicit data points by an approximation of the underlying distribution for the maintenance probabilities. In the class of pedestrian underpasses considered in this study, for example, the highly linear nature of the cumulative maintenance frequencies (Fig. 4) suggests that a uniform distribution could be a viable assumption.

As a next step, the costs of the maintenance and replacement measures must be analyzed and evaluated in order to include them in the quality prediction. The costs to be assigned must reflect the measures depicted. Since different quality benefits result from the maintenance measures, it is useful to develop the prices for the maintenance measures as unit prices. The unit price is standardized to the price-enhancing feature of the item, for example, the surface area of a platform in square meters, but also to a quality increase of one quality score point. By using unit prices, costs can be assigned to each maintenance measure of an item. This allows the required budget to be determined for the entire network.

Furthermore, a prioritization logic must be developed to decide which measures are to be implemented first. Since there might not be enough budget for all the measures needed according to the model, both, the maintenance and replacement measures must be sorted with regard to this prioritization logic. The measures are then carried out according to the resulting sequence. Different prioritization criteria lead to different implemented measures and thus to different qualities in the network. Therefore, prioritization plays an important role in the cause-effect relationship.

Once the costs and the prioritization for measures are included, the model can be implemented in a software tool. This simulation has to be validated against real-life data to prove the model's plausibility. After this validation, the software tool should be capable of evaluating the cause-effect relationship between the funds and quality of DB Station&Service AG's equipment.

A potential issue in the practical application of our approach is the fact that the model does not distinguish between different maintenance measures of subcomponents (e.g., replacement of a platform's floor or replacement of guide strips for the blind). They are likely to incur systematically different costs and yield different improvements of the quality score. This fact is currently not reflected in our model as the maintenance measures are modeled as average measures.

Another practical limitation concerns the applicability of the model under the existence of clusters within the item classes. The model describes how an item of a class degrades on average. Yet, in the construction of items usually different materials, construction designs, etc., are used depending on their time of construction, their specific use case, etc.. This may yield different degradation functions if examined by clusters. However, on the basis of DB Station&Service AG's data, it remains unknown if this is the case, as the data history is too short and the data base is too small for an assessment of item clusters.

References

[1] A. Brint and M. Black, Improving estimates of asset condition using historical data, Journal of the Operational Research Society, vol. 65, 2014.

[2] B. Czarnecki and B. Poon, Concrete Sidewalk Design Analysis and Optimization for Improved Life Cycle and Sustainability,TAC 2017: Investing in Transportation: Building Canada's Economy–2017 Conference and Exhibition of the Transportation Association of Canada,2017.

[3] A. Falamarzi, S. Moridpour and M. Nazem, A review of rail track degradation prediction models, Australian Journal of Civil Engineering, vol. 17, 2019.

[4] S. Heitel, H.Koriath, C. Herzog and G.Specht, Vergleichende Lebenszykluskostenanalyse für Fußgängerbrücken aus unterschiedlichen Werkstoffen, Bautechnik,vol. 85,2008.

[5] T. Jacke, B. Dickenbrok, N. Friesen, A.A. Grub and N. Nießen, Ursache-Wirkung-Zusammenhang: Zusammenhang zwischen Mitteleinsatz und Infrastrukturqualität abbilden, Eisenbahntechnische Rundschau, vol. 68, 2019.

[6] C. Letot, I. Soleimanmeigouni, A. Ahmadi and P. Dehombreux, An adaptive opportunistic maintenance model based on railway track condition prediction, IFAC-PapersOnLine, vol.49, 2016.

[7] J. Neuhold, I.Vidovic and S. Marschnig, Preparing track geometry data for automated maintenance planning, Journal of Transportation Engineering, Part A: Systems, vol. 146, 2020.

[8] D. Prescott and J. Andrews, Modelling maintenance in railway infrastructure management, 2013 proceedings annual reliability and maintainability symposium (RAMS), IEEE, 2013.

[9] A. Saadat, Basic Prediction Model of Railway Track Circuit Failure, Quarterly Journal of Transportation Engineering, vol. 11, 2019.

[10] L.M. Quiroga and E. Schnieder, A heuristic approach to railway track maintenance scheduling, WIT Transactions on The Built Environment, vol. 114, 2010.

[11] Leistungs- und Finanzierungsvereinbarung I (LuFV I) zwischen der Bundesrepublik Deutschland und den Eisenbahninfrastrukturunternehmen der Deutschen Bahn AG sowie der Deutschen Bahn AG, 2009.

[12] Leistungs- und Finanzierungsvereinbarung III (LuFV III) zwischen der Bundesrepublik Deutschland und den Eisenbahninfrastrukturunternehmen der Deutschen Bahn AG sowie der Deutschen Bahn AG, 2020.