

## A decision support approach for condition-based maintenance of rails based on big data analysis

Jamshidi, Ali; Hajizadeh, Siamak; Su, Zhou; Naeimi, Meysam; Núñez, Alfredo; Dollevoet, Rolf; De Schutter, Bart; Li, Zili

**DOI**

[10.1016/j.trc.2018.07.007](https://doi.org/10.1016/j.trc.2018.07.007)

**Publication date**

2018

**Document Version**

Accepted author manuscript

**Published in**

Transportation Research Part C: Emerging Technologies

**Citation (APA)**

Jamshidi, A., Hajizadeh, S., Su, Z., Naeimi, M., Núñez, A., Dollevoet, R., De Schutter, B., & Li, Z. (2018). A decision support approach for condition-based maintenance of rails based on big data analysis. *Transportation Research Part C: Emerging Technologies*, 95, 185-206. <https://doi.org/10.1016/j.trc.2018.07.007>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

1 **Abstract**

2 In this paper, a decision support approach is proposed for condition-based maintenance of rails relying on  
3 expert-based systems. The methodology takes into account both the actual conditions of the rails (using axle  
4 box acceleration measurements and rail video images) and the prior knowledge of the railway track. The  
5 approach provides an integrated estimation of the rail health conditions to support the maintenance decisions  
6 for a given time period. An expert-based system is defined to analyze interdependency between the prior  
7 knowledge of the track (defined by influential factors) and the surface defect measurements over the rail. When  
8 the rail health conditions is computed, the different track segments are prioritized, in order to facilitate grinding  
9 planning of those segments of rail that are prone to critical conditions. In this paper, real-life rail conditions  
10 measurements from the track Amersfoort-Weert in the Dutch railway network are used to show the benefits of  
11 the proposed methodology. The results support infrastructure managers to analyse the problems in their rail  
12 infrastructure and to efficiently perform a condition-based maintenance decision making.

13

14 **Keywords:** Decision support system, Condition-based maintenance, Rail surface defects, Fuzzy inference  
15 system, Axle Box Acceleration (ABA) system.  
16

## 17 **A decision support approach for condition-based maintenance of rails based on big data** 18 **analysis**

19 Ali Jamshidi<sup>1</sup>, Siamak Hajizadeh<sup>1</sup>, Zhou Su<sup>2</sup>, Meysam Naeimi<sup>1</sup>, Alfredo Núñez<sup>1</sup>, Rolf Dollevoet<sup>1</sup>, Bart De  
20 Schutter<sup>2</sup>, Zili Li<sup>1</sup>

21 <sup>1</sup>*Section of Railway Engineering, Delft University of Technology, Delft, The Netherlands.*

22 <sup>2</sup>*Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands.*

23 *Corresponding author: Alfredo Núñez, [a.a.nunezvicencio@tudelft.nl](mailto:a.a.nunezvicencio@tudelft.nl)*

24

### 25 **Abstract**

26 In this paper, a decision support approach is proposed for condition-based maintenance of rails relying on  
27 expert-based systems. The methodology takes into account both the actual conditions of the rails (using axle  
28 box acceleration measurements and rail video images) and the prior knowledge of the railway track. The  
29 approach provides an integrated estimation of the rail health conditions to support the maintenance decisions  
30 for a given time period. An expert-based system is defined to analyze interdependency between the prior  
31 knowledge of the track (defined by influential factors) and the surface defect measurements over the rail. When  
32 the rail health conditions is computed, the different track segments are prioritized, in order to facilitate grinding  
33 planning of those segments of rail that are prone to critical conditions. In this paper, real-life rail conditions  
34 measurements from the track Amersfoort-Weert in the Dutch railway network are used to show the benefits of  
35 the proposed methodology. The results support infrastructure managers to analyse the problems in their rail  
36 infrastructure and to efficiently perform a condition-based maintenance decision making.

37 **Keywords:** Decision support system, Condition-based maintenance, Rail surface defects, Fuzzy inference  
38 system, Axle Box Acceleration (ABA) system.

39

### 40 **1. Introduction**

41  
42 The increase in train traffic and axle loads affect the health conditions of railway infrastructure. Hence, efficient  
43 infrastructure monitoring and maintenance is among the major concerns of infrastructure managers in order to  
44 improve the performance of railway operations (Ahren and Parida, 2009). As such, infrastructure health  
45 conditions should be monitored and considered in the maintenance decision making process. Effective  
46 management of infrastructure health conditions is crucial to guarantee the desired asset quality level (Parida  
47 and Chattopadhyay, 2007; Gandomi and Haider, 2015). It also plays an important role in meeting the demands  
48 for the whole system performance when the infrastructure is upgraded e.g. when increasing traffic capacity, the  
49 maintenance regime should be adapted to avoid compromising safety and infrastructure health requirements.  
50 To keep the infrastructure system working at an effective level, a conditions-based maintenance system is  
51 required not only to consider the actual health conditions but also evolution during the maintenance decision  
52 horizon (Jamshidi et al., 2016).

53 Condition-based monitoring is used in railway infrastructures to estimate the actual health conditions of the  
54 assets, so that degradation processes can be effectively controlled. It helps to keep the infrastructure manager

55 continually informed of the estimated health of the railway infrastructure. Condition-based monitoring is  
56 supposed to collect information that will allow an effective operation by reducing maintenance cost,  
57 eliminating unnecessary operations, and focusing on places where the problems are located and where they will  
58 be in the coming period. Furthermore, the enhancement in usage of the railway infrastructure needs a  
59 systematic monitoring plan to keep the trains running safely by considering all related data influencing the  
60 health conditions. The data for a typical railway infrastructure includes a large amount of frequent  
61 measurements from the monitoring systems of the assets involved in the railway operations. To ensure the  
62 required performance level, a huge amount of data should be collected, transmitted, processed, and properly  
63 stored so that it can be used as historical information. This whole process reflects a transition from raw  
64 infrastructure data into actionable maintenance knowledge. Therefore, the database constituted from continuous  
65 monitoring will become larger and larger over time and applying big data analysis approaches is inevitable  
66 (Fumeo et al., 2015). In order to design proper maintenance plans in railways, it is necessary to explore and  
67 analyse the growing amount of data and to extract useful information. To do so, different sensors can be used to  
68 collect the data obtained in railway track monitoring at different times, environmental conditions, and  
69 frequencies. These data can exhibit different characteristics: (1) discrete or continuous, (2) spatial or temporal,  
70 (3) signal and images among others (Lasisi and Nii, 2018; Nii, 2017; He et al., 2013; Liu and Dick, 2016;  
71 Ghofrani et al. 2018).

72 In condition-based maintenance for railways, the monitoring data are mostly collected periodically with regular  
73 sampling intervals. For some critical assets, the monitoring can be adapted to other possible needs including  
74 continuous measurement. The essential concept for the monitoring data is to take the degradation of the  
75 infrastructure into account, in particular for critical infrastructure like rails. This paper focuses on rail  
76 conditions monitoring, which has a critical role in the network performance (He et al., 2015; He et al., 2013). In  
77 an intensively used network, a considerable amount of the maintenance budget has to be allocated for the rail,  
78 e.g., in the Dutch railway network, this amounts to almost half of the annual maintenance budget  
79 (approximately 60.000 euro/km) (Zoeteman et al., 2014). As a high percentage of failures are directly related to  
80 the rail, it is important to assess the rail conditions in order to obtain a proper condition-based maintenance  
81 approach. More specifically, the health conditions analysis involves detecting the rail surface defects that can  
82 potentially result in rail breaks and derailment in extreme cases (Liu et al., 2011; Liu et al., 2012; Islam et al.,  
83 2016).

84 Rolling contact fatigue (RCF) affects the health conditions of the rail due to the contact in the interface between  
85 wheel and rail (Makino et al., 2016). RCF is a generic term describing a range of rail surface defects and has  
86 been an interesting challenging research topic, in particular the influence of RCF on maintenance decision  
87 making (Sciammarella et al., 2016). Moreover, its influence is related to other factors including traffic type,

88 train speed, traffic load, rail/wheel profile, train characteristics, and maintenance policy (Popović et al., 2013).  
89 Once RCF appears, it induces considerable dynamical forces on the rail surface, and subsequently cracks are  
90 initiated and propagated from the surface (Zhuang et al., 2018; Makino, 2012). The most important cause of the  
91 appearance of defects is the large number of trains passing over rail critical components, most significantly at  
92 welds, joints, and switches (Molodova et al. 2014). Early detection of surface defects is important to mitigate  
93 induced maintenance costs as well as unforeseeable consequences of rail breaks. There are different methods to  
94 diagnose the conditions of rail defects, including ultrasonic measurements (Fan et al., 2007), eddy current  
95 testing (Song et al., 2011), and guided-wave based monitoring (Mariani et al., 2013). In this paper, the focus is  
96 on a type of rail surface defect called squat. The costs for treating these defects in the Dutch railway network  
97 are considerably high (more than 5000 euro/km per year) (Molodova et al., 2014). The maintenance of squats  
98 should be different according to their severity. For late-stage squats, a rail replacement plan is a proper decision  
99 while for the light squats, grinding a thin layer from the rail surface is the most effective solution. Hence, when  
100 all residual damages are removed, grinding is effective and the rail will be turned to a healthy condition. To  
101 optimally plan grinding operations, condition-based maintenance relying on early detection of the squats is  
102 required. Although a defect detection method could give an indication of the health of the rail, the infrastructure  
103 manager requires prior knowledge to (1) be aware of all influential factors, (2) analyse interdependency  
104 between the rail observations and the influential factors, and (3) obtain a future view of the track conditions. In  
105 this paper, we relate influential factors to the rail health conditions to show the effect of the track characteristics  
106 in the rail observations. Hence, by having knowledge about the track characteristics, potential risks about the  
107 rail can be anticipated due to the effect of the influential factors on the appearance of defects and consequently  
108 on the rail health conditions. Therefore, an analysis of influential factors should be taken into account to give at  
109 the most a proper prospect of the infrastructure health conditions.

110 Mixed Integer Linear Programming (MILP) is a common approach for track maintenance scheduling. An  
111 MILP model is developed in (Wen et al., 2016) for optimal condition-based preventive maintenance for a  
112 single track divided into multiple segments, considering various economic and technical factors such as train  
113 speed limits and track quality. The optimal planning of routine maintenance activities and projects like grinding  
114 to minimize maintenance costs and track possession time for a single track is formulated as an MILP in (Budai  
115 et al., 2006). The optimal scheduling of rail, sleeper, and ballast renewal at a network level is formulated as an  
116 MILP problem in (Caetano and Teixeira, 2016) to minimize the expected life-cycle cost and track  
117 unavailability. In (Peng and Ouyang, 2014), the optimal clustering of track maintenance jobs into projects to  
118 minimize total maintenance costs for the network of track is recast as a Vehicle Routing Problem (VRP). The  
119 track maintenance problem considering different priorities for each section in the railway network is formulated  
120 as an VRP with customer costs in (Heinicke et al., 2015). A time-space network model is developed in (Peng  
121 and Ouyang, 2012) for the optimal scheduling of capital maintenance projects like rail replacement. A

122 metaheuristic based on simulated annealing is developed in (Santos and Teixeira, 2011) to determine the  
123 optimal tamping length of a tamping machine, minimizing the associated logistic costs and fixed machine  
124 costs. In this paper, we use a simplified MILP model to optimize the rail grinding decision plan into clusters  
125 that can be related to the actual conditions of the rail. The proposed MILP model not only uses different  
126 clusters for determining the most critical pieces of tracks, but also simultaneously takes time and budget  
127 constraints into account. Moreover, the model benefits a new method for estimating rail health conditions as an  
128 input data. This eases implementing the condition-based maintenance strategy, reaching an effective  
129 maintenance plan in terms of rail health conditions and also reduces the high cost of track maintenance  
130 activities. In this paper, we propose a condition-based maintenance methodology taking both the observations  
131 and the prior knowledge of the track into account. The idea is to find the interdependency between the defect  
132 status and all major influential factors of the track prior knowledge. The defect status is defined in terms of  
133 number and severity of the defects. We investigate the interdependencies between the influential factors and the  
134 defect appearance by studying the track characteristics. Once the interdependency is studied, a set of rules is  
135 generated to connect rail conditions to the influential factors. The results then indicate which pieces of the rail  
136 are prone to be defective. The infrastructure manager is then able to propose maintenance planning according  
137 to the critical pieces of railway track. The methodology uses big data analytics, with real-life data measured  
138 from a Dutch railway track, using Axle Box Acceleration (ABA) measurements and rail video images  
139 (Molodova et al., 2014, Li et al., 2015; Hajizadeh et al., 2014).

140 Figure 1 shows the flowchart of the proposed methodology in five steps. The major contribution of the paper is  
141 to propose a methodology for rail maintenance decision making that is a combination of new methods and also  
142 uses already existing models. Particularly, in Step 5 we make use of the model proposed in Su et al. (2017).  
143 Moreover, the proposed methodology is presented in an integrated framework to keep simplicity and coherence  
144 between steps. This helps not only to guarantee real-life implementation, but also to keep the infrastructure  
145 manager updated of the infrastructure health conditions. In Step 1, the rail defects are detected by using two  
146 sources containing the ABA signals and rail video images. A list of critical defects is then provided to represent  
147 rail observation. In Step 2, track influential factors,  $\gamma_j$ , are presented to give context on the prior knowledge of  
148 the track for each segment  $j$ . Step 3 explains the interdependency analysis between the influential factors and  
149 the rail observations obtained from Step 1. The aim is to investigate how the influential factors are related to the  
150 rail observations.

151 The analysis of the interdependency between track characteristics and the rail observations is to support expert  
152 judgments in order to develop health condition rules as proposed in Step 4. In Step 4, s expert decision system  
153 is proposed using an inference system. To do so, a fuzzy approach is used including two steps: (1) a

154 questionnaire filled out by experts and (2) a set of fuzzy health condition rules,  $R_1, R_2, \dots, R_R$ . The fuzzy rules  
155 help to link the influential factors to the health condition. The rules are generated according to expert judgment  
156 through a questionnaire. Thus, by generating the rule set, the inference system is built using Mamdani inference  
157 model. The aim for the Step 4 is to find the most critical segments that require maintenance among all rail  
158 segment. Therefore, the infrastructure manager will have the information of the critical segments. To include  
159 operational considerations for the maintenance decisions e.g. time slot limits, logistic concerns, etc., Step 5 is  
160 proposed. This results in optimal suggestions for the maintenance decisions. A real-life case from the Dutch  
161 railway network is provided to apply the framework and show the applicability of the framework.

162 "place Fig. 1 about here"

## 163 **2. Step 1: Intelligent rail conditions monitoring**

164 In this paper, we require a technology that can detect defects in an early stage. Hence, we consider to use ABA  
165 measurements (Li et al., 2008). To enhance the visualization, ABA measurements are combined with rail  
166 image videos (Jamshidi et al., 2017; Faghih-Roohi et al., 2016). In our case study, the ABA measurement and  
167 rail video images are used to study rail surface defects; specifically squats, as they are costly for railway  
168 networks. A global scheme of the measurement systems is given in Figure 2.

169 "place Fig. 2 about here"

170 Li et al. (2015) show the feasibility of early-stage squat detection using an ABA system. The ABA system can  
171 be employed to detect a range of surface defects, most importantly, squats, corrugations, and damaged welds,  
172 insulated joints, and switches. The ABA system can be embedded in in-service operational trains. Four  
173 channels are assigned for the ABA measurement including left rail and right rail, and horizontal and vertical  
174 accelerations to capture all the ABA signals.

175 The image data is collected by a set of high frame rate cameras that are mounted on a specialized measurement  
176 train. A top view camera is aimed at the rail surface defects, with each frame covering a length of 15 cm of the  
177 track along the longitudinal rolling direction. The recordings are pre-processed into video compilations where  
178 consecutive frames have a few millimetres of overlap and the effects of variations in the train speed are  
179 removed. As a result, recordings of roughly a thousand kilometre of rail amount to producing hundreds of  
180 Gigabytes of video data.

181 Deep convolutional neural networks (DCNNs) has been applied for different problems in the area of  
182 classification due to their algorithmic advantages (Krizhevsky et al., 2012; LeCun et al., 2015). We use a  
183 DCNN model in order to automatically estimate from the ABA signals the defect severity throughout the tracks

184 based on a big data analysis. For training the DCNN, based on previous results (Jamshidi et al., 2017; Faghih-  
185 Roohi et al., 2016), we obtain a set of labelled images with their severity. The labels used from the images  
186 samples are on a scale from 0 to above 4 according to the severity level of the defects visible in the squats  
187 found by analysis of rail images. Non-defect track images are assigned a value of zero and defects are assigned  
188 from 1 and above. The severity of the squat  $s$  can be used to represent the health conditions of the rail,  $H_s(t)$ , at  
189 the time instant  $t$  of the measurement as follows:

$$190 \quad H_s(t) = \begin{cases} S_1 & \text{if } 0 < L_s(t) \leq 1 \\ S_2 & \text{if } 1 < L_s(t) \leq 2 \\ S_3 & \text{if } 2 < L_s(t) \leq 3 \\ S_4 & \text{if } 3 < L_s(t) \leq 4 \\ S_5 & \text{if } 4 < L_s(t) \leq 5 \end{cases} \quad (1)$$

191 where  $L_s(t)$  is the measured level of severity,  $S_1$  refers to a seed squat,  $S_2$  is a light squat,  $S_3$  is a moderate squat,  
192  $S_4$  is a severe squat, and  $S_5$  is a squat with risk of rail break. The images and their severity are matched with  
193 their corresponding ABA signal. To do so, a window of the ABA samples is defined with length of 3036  
194 samples, covering full responses to local defects. This facilitates matching the signals with the video frames.  
195 Figure 3 shows two samples of image data used for the severity analysis associated with the corresponding  
196 ABA signals. The labelled data is therefore split into two parts for training and testing. To keep consistency in  
197 the defect detection, the labelled samples are collected from different locations over the measured track  
198 and they cover all the types of squats. They are compiled into a training set for each of the classes. The  
199 dataset was obtained by manual labelling of the images by an expert. The labelled sample defects are then  
200 divided into a training set and a testing set. The sample size was 125 squats. The distribution of the squats  
201 classes in terms of severity set is 70 samples for  $S_1$ , 8 for  $S_2$ , 6 for  $S_3$ , 8 for  $S_4$ , and 33 for  $S_5$ . 75% of the  
202 data is assigned for training and 25% for validating of the network performance. The samples of the  
203 labelled images are composed of 125 different squats collected from different locations of the track. We  
204 train a convolutional neural network regression model using the samples. The average binary accuracy (defect  
205 vs. non-defect) of the network on all tested samples is taken into account. Although putting a high acceptance  
206 threshold on the network output response might increase the rate of false positive detection, we use the  
207 threshold to detect the correct classes of the defects, seed (trivial) defects, and the normal classes. Once the  
208 DCNN for the image data is trained, defects in the large pool of previously unseen samples can be found.

209

210

“place Fig. 3 about here”

211 Using a set of convolutional layers, the defect features are included in the DCNN model as filters to recognize  
212 distinguishing features and to create a feature map. A Rectified Linear Unit (ReLU) is used as activation  
213 function after the convolution steps, as well as max-pooling layers in order to down-sample the outcome of  
214 each layer (Srivastava et al., 2014). The convolutional and pooling layer are finally attached to a sequence of  
215 three fully-connected layers to get class predictions (see in Figure 4).

216 "place Fig. 4 about here"

217  
218 The separating rail observations (detecting squats using DCNNs) from track characteristics (determined by  
219 influential factors) is one of the major contributions of this paper. On the one hand, the DCNN is used to  
220 estimate the severity of the defects according to the ABA and image sources. This just gives the defect analysis  
221 (the rail observation in the Step 1) and not the rail health condition. On the other hand, track prior knowledge  
222 containing the influential factors can impact the rail health condition (Step 2) as those factors affect the quality  
223 of rail use over time (rail degradation). Thus, influential factors are collected to contribute the track  
224 characteristics for the estimation of rail health condition. For instance, a piece of rail positioned on a rail curve  
225 can get degraded faster than the same rail piece on a straight rail. To include track characteristics effects, the  
226 interdependency between the rail observations (the DCNNs) and the track prior knowledge is investigated in  
227 Step 3.

228

### 229 **3. Step 2: Prior knowledge of the track**

230 General characteristics of the railway track system can have a large influence in the initiation and growth of the  
231 rail defects. A list of some generic track characteristics that are potentially relevant to the appearance of rail  
232 defects are discussed next. The idea in this paper is to take seven factors into account as "general characteristics  
233 of track" as according to the literature survey, they are proved to be significantly influencing in the initiation  
234 and growth of the rail defects. In particular, we classified the seven influential factors based on Step 2 into three  
235 categories: (1) track profiles, (2) track irregularities, and (3) operational speed profile and tractive efforts.  
236 However, there are other factors that can affect the track. As an example, train traffic can be influential and has  
237 an important role in the actual rail health conditions. In this paper, we assume that the influence of the traffic  
238 tonnage, which increases the amount of contact force between wheel-rail, can be seen in the defect severities  
239 (the rail observation). Furthermore, tonnage will be an influential factor when predicting defect evolution over  
240 time. During the same time period, the rail defects in segments with a higher tonnage evolved faster than the  
241 defects in segments with a low tonnage. Additionally, observations indicate that a higher number of defects will

242 be found in tracks with a higher tonnage. These are two possible ways to include the effect of the traffic  
243 tonnage in the proposed approach: (1) Indirectly via the effect of the tonnage in the rail observation. Condition  
244 monitoring measurements will automatically update both the appearance of new defects and the severity of the  
245 defects. (2) Directly via the inclusion of tonnage as influential factor. This case is most suitable when the  
246 infrastructure manager wants to predict the evolution of the defects; as tonnage will indicate how fast detected  
247 defects will evolve. In this paper, we do not include the prediction of the defect evolution, so in this case the  
248 indirect method via rail observation is conducted. Part of the future research is to consider the effect of tonnage  
249 within a predictive approach.

250 We employ various sources of information to obtain the prior knowledge of track using a big data analysis.

### 251 *3.1. Track profiles*

252 Track profiles are design features. Deviations of the track alignments (vertical, lateral, etc.) with respect to the  
253 nominal alignment can lead to track irregularities (Wang et al., 2012; Kawaguchi et al., 2005). Mutton et al.  
254 (1991) analyse the wheel-rail contact conditions in the curved and tangent track to investigate the influence of  
255 the lateral profile of the track on the initiation and growth of rail defects. Grassie (2012) reviews the research on  
256 squats and squat-type defects. The author concludes that squats are associated with straight tracks and gentle  
257 curves, but not with tight curves. Likewise, Li et al. (2008) report that squats in the Netherlands occur mainly  
258 on straight tracks and gentle curves. On the contrary, head checks occur mostly on the curved tracks of radii  
259 less than 3000 m (Li, 2010). In the current paper, the horizontal curvature of the track is taken into account.  
260 Furthermore, the rail segments are defined based on the rail curvature. In this way, only one influential factor  
261 for the horizontal curvature is considered for one segment. The vertical profile is ignored as the corresponding  
262 changes in the Dutch railway network are small.

### 263 *3.2 Track irregularities*

264 The track geometry changes from the design geometry due to trains passing over the track. More passing trains  
265 could worsen the track geometry conditions. In the literature, the irregularity amplitude and its wavelengths are  
266 mostly used as the controlling factors of the track quality. The limits for those controlling factors are typically  
267 analysed using measurements and dynamic simulations. The presence of track irregularities was found to have  
268 an influential effect on RCF defect appearance (Nielsen et al., 2005). Track geometry problems are widely  
269 explained as one of the influential factors considering wheel-rail interactions, maintenance planning, and life of  
270 railway tracks. Irregularities have an impact on ride comfort and the traffic safety level. All those influences are  
271 therefore very critical in railway dynamics. Nonetheless, the critical level is directly related to track usage. In  
272 the literature, there are also different studies about the influence of track geometry on the track conditions and  
273 the track degradation. Thus, by considering the significant contribution of the track geometry in the track

274 conditions and then subsequent maintenance plans, control of track irregularities plays an important role on  
275 facilitating condition-based maintenance planning (Andrade and Teixeira, 2011; Andrade and Teixeira, 2012).

276 Using geometry measurements for the rail maintenance planning is of important considerations for the  
277 infrastructure manager (Veit, 2007; Sharma et al., 2018). A maintained track geometry considerably contributes  
278 not only to train safety but also track health conditions. Furthermore, track geometry monitoring could help to  
279 prolong the effective track life time by managing the track degradation, the track health conditions, and  
280 subsequently the cost of the maintenance operations (Kawaguchi et al., 2005).

281 The measurement data has been used to develop statistical modelling of railway track irregularities in the last  
282 three decades. Track safety and ride comfort are among the first track irregularities analysis using field data.  
283 Hamid and Gross (1981) discuss the impact of track quality on track maintenance decisions and performance-  
284 based analysis of track geometry using a statistical model for a long track. The paper develops a degradation-  
285 based track conditions model to explain interaction between rail defects and performance indicators. A similar  
286 investigation has been carried out using linear models to capture the track response to a train load in terms of  
287 track irregularities and potential appearance of rail defects (Corbin and Fazio, 1981). Bing and Gross (1983)  
288 use a comprehensive model to predict the track quality for maintenance operations. They employed multiple  
289 data of traffic and train speed, track structure, and maintenance time slots to predict the track quality over time.  
290 In the current paper, based on the available data, we select three sets of irregularity-related influential factors  
291 including (1) the vehicle effect, which is a signal indicating the train ride quality based on several geometry  
292 measurements and operating trains characteristics, (2) track geometry, which is an indicator estimated based on  
293 a combination of different track geometry measurements such as horizontal alignment, the vertical alignment,  
294 and cant differences, and (3) track superelevation, which is the difference between the designed cant and the  
295 measured cant.

### 296 3.3. *Operational speed profile and tractive efforts*

297 Tractive effort and curving in the track are found to be potentially responsible for RCF-type rail damages  
298 (Grassie and Elkins, 2005). The review of the squat defects by Grassie (2012) reveals that these defects are  
299 associated with driving traction i.e. locomotives and power cars. Observations by Li et al. (2011) show the  
300 relationship between braking and squat occurrence in the Dutch railway network. The authors conclude that the  
301 traction performance of the rolling stock has a large influence on the initiation and growth of squats. They  
302 found many squats at pieces of a track where the gradient of the speed was the highest and the speed was low.  
303 Moreover, the low speed was also influential, as more frequent activation of the Anti-lock Brake System (ABS)  
304 system occurs at lower speeds. Tractive and braking efforts, which differ by the types of locomotives, can also  
305 influence the occurrence of RCF defects. A wide range of the Direct Current (DC) or the Alternating Current  
306 (AC) drive systems are used in different countries to provide the required power supply of the trains. A

307 comparison is made between AC versus DC locomotives under diverse operational conditions in Australia to  
308 investigate the possible development of squats in the rail head (Vo et al., 2015). Scott et al. (2014) find that the  
309 most susceptible locations to the squat defects are those where low-speed running occurs with high wheel slip  
310 and low adhesion. They investigated the traction characteristics of the typical AC traction motors to find the  
311 potential link between the generation of defects and the rolling stock type. In the current paper, the speed  
312 profile of the typical rolling stock is investigated to determine its potential correlations with the occurrence of  
313 defect. The related effects considered in this paper includes: (1) train speed profile, which is the speed of the  
314 measurement train in km/h, (2) train acceleration profile, which is the acceleration of the measurement train in  
315  $m/s^2$ , and (3) rail head wear, which estimates the difference between the measured height of the railhead and  
316 the nominal height of a new rail railhead in mm. The measurements are obtained with tacho signals,  
317 accelerometers, and scanning laser sensors mounted on the measurement train.

#### 318 **4. Step 3: Interdependency analysis**

319 According to the track prior knowledge explained in Step 2, those track factors that are observed to be  
320 influential on rail conditions in the Dutch railway network are considered. We use the data available in the  
321 Dutch railway infrastructure monitoring system, BBMS ("Branche Breed Monitoring Systeem"), to acquire the  
322 signals of the influential factors. In this paper, we use both dynamic and static measurements to obtain the  
323 influential factors. After processing the measurements, the influential factors are calculated for a single  
324 measurement campaign. Part of the further research includes the use of historical measurements to study the  
325 evolution of the influential factors over time. Seven signals are chosen as influential factors that might  
326 significantly affect the rail conditions including (1) train speed profile, (2) train acceleration profile, (3) track  
327 horizontal curvature, (4) track geometry parameter, (5) rail head wear, (6) vehicle effect, and (7) track  
328 superelevation. In Figure 5, a map is employed to show the track including all the seven influential factors. The  
329 data are captured over the whole track to analyse the dynamics of the track influential factors.

330

331

332 "place Fig. 5 about here"

333

334 Hence, on the one hand we have a set of data over the track representing the track knowledge and on the other  
335 hand, squats are detected along the track with their severity and location using the ABA signals and the image  
336 data. The interdependency is defined by investigating how to match the location and the severity of a certain  
337 defect with the signals of the track influential factors. To do so, the track is partitioned into different segments  
338 and the interdependency is investigated per segment.

339 To numerically represent the severity of a segment, we consider the average of the severities of all the squats  
 340 that are located in segment  $j$ :

341  
 342

$$343 \quad \mathfrak{N}_j(t) = \left( \frac{\sum_{s \text{ in segment } j} H_s(t)}{\sum_{s \text{ in segment } j} \delta_s(t)} \right) \quad (2)$$

344  
 345 where  $H_s(t)$  is the severity of the squat  $s$  provided by the ABA detection algorithm for the measurement time  
 346  $t$ . The function  $\delta_s(t)$  equals 1 when  $s$  is a squat, and equals 0 otherwise.

347 Regarding the processing of the datasets, once all the data sets (signals) over the track are acquired, the  
 348 signals are processed according to equations (3) and (4). First, signals are normalized using (3), and then  
 349 the influential factor is obtained by the average of the signal as in (4). The influential factor is then a  
 350 "representative" value of the measured signal for that segment. So, the signals should all be normalized  
 351 between  $L_{\text{int}}$  and  $L_{\text{end}}$  which are respectively the upper bound and lower bound of the interval selected for the  
 352 normalization. The function can be expressed as:

353

$$354 \quad \gamma_{j,\text{Nor}}^k(x,t) = \frac{(\gamma_j^k(x,t) - \gamma_{j,\text{min}}^k(t))(L_{\text{end}} - L_{\text{int}})}{\gamma_{j,\text{max}}^k(t) - \gamma_{j,\text{min}}^k(t)} + L_{\text{int}} \quad (3)$$

355  
 356 where  $\gamma_j^k(x,t)$  is the data for the  $k$ -th influential factor at the location  $x$  and time instant  $t$ ,  $\gamma_{j,\text{min}}^k(t)$  and  
 357  $\gamma_{j,\text{max}}^k(t)$  are minimum and maximum values of the signal at the segment  $j$ . By considering  $x_{j,\text{avg}}^k$  as the  
 358 location where average value of the data occurs (as a representative of each segment), the data value for the  
 359 segment  $j$  is calculated according to:

360

$$361 \quad \gamma_j^k(t) = \gamma_{j,\text{Nor}}^k(x_{j,\text{avg}}^k, t) \quad (4)$$

362 where  $\gamma_j^k(t)$  is the influential factor for the segment  $j$  and the time step  $t$ .

363 By considering a matrix containing  $\gamma_j^k(t) = \left[ (\gamma_1(t))^T, \dots, (\gamma_{N_s}(t))^T \right]$  where  $N_s$  is the total number of segments

364 and  $\gamma_j(t) = [\gamma_j^1(t), \dots, \gamma_j^n(t)]$ , a clustering model is proposed as follows. We have selected the method called  
 365 Fuzzy C-Means due to its simplicity. Based on the fuzzy clustering approach, a data point will belong to all the  
 366 clusters but with a different membership degree. The closer to the centre of the cluster, the membership will be  
 367 near to one. Points far away from a cluster will have a membership degree near to 0 (Ma et al., 2015). Just for

368 illustration, three clusters are defined over the influential factors. The membership degree to the cluster  
369 determines how much a segment belongs to the cluster. The track is partitioned into 15 segments. Figure 6  
370 shows a schematic view of the clusters. As seen in the figure, segment 5 is highlighted by a rectangle indicating  
371 a high membership degree of the cluster 2 in the segment indicated by an arrow. Rail segment 4 has the higher  
372 membership to cluster 1; however, it does not belong to the cluster 1 as much as segments 1, 2, 3, 6, 14, and 15,  
373 which they all have membership values near to one. The results are used in order to obtain rail health  
374 conditions decision rules.

375

376 "place Fig. 6 about here"

377

378 In this paper, five levels are defined including very low ( $L_1$ ), low ( $L_2$ ), medium ( $L_3$ ), high ( $L_4$ ), and very high  
379 ( $L_5$ ) to represent the interdependency between the defect severity and the influential factors (for simplicity and  
380 interpretability of the data, linguistic terms such as very low, low, medium, high and very high are used).

381

382

### 383 5. Step 4: Fuzzy inference model

384

385 In this paper, a fuzzy inference system is used to develop rules about the rail health condition-based on the  
386 influential factors  $\gamma_j^k(t)$ . The Mamdani fuzzy system approach is considered due to its interpretability  
387 and simplicity (Camastra et al., 2015; Tosun et al., 2011). To explicitly express the inference system, the  
388 Mamdani inference can be defined as follows:

389

$$390 Y_j^m(t) = f_{\text{Mamdani}}(\gamma_j^1(t), \gamma_j^2(t), \dots, \gamma_j^k(t), \dots, \gamma_j^n(t)) \quad (5)$$

391

392 where  $Y_j^m(t)$  is the rail health conditions in section  $j$  and  $\gamma_j^k(t)$  the  $k$  influential factor in section  $j$ . Figure  
393 7 shows the architecture of the inference model. In the first layer, the values of input variables, e.g.  $\gamma_j^k(t)$   
394 are used. The membership degrees of the inputs to the fuzzy values are obtained in layer 2 and employed  
395 to compute the rule truth values in layer 3. At the layer 4, according to the truth values of the various  
396 rules, the rail health conditions of each rule in the segment  $j$  is estimated.

397

398 "place Fig. 7 about here"

399

400 The  $R$  fuzzy if-then rules are generated based on (5) to capture combinations of the influential factors. The  
 401 purpose is to assign a membership degree to each  $\gamma_j^k(t)$ . Gaussian membership functions are used to  
 402 fuzzify the inputs. The Gaussian type of membership function has been used because it is smooth and  
 403 nonzero at all points (Markowski & Mannan, 2008; Xie, 2003). The Gaussian membership function is  
 404 based on two parameters and can be represented as:

$$405 \quad \text{Gaussian}(x; c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \quad (6)$$

406  
 407 where for each membership function,  $c$  and  $\sigma$  are the parameters of the membership function. The  
 408 parameters are tuned so that every membership function has around 30 percent overlapping with the  
 409 neighboring functions. The rule  $r_i$  can be expressed as:

$$411 \quad r_i: \text{If } \gamma_j^1(t) \text{ is } V_1^i \text{ and } \dots \gamma_j^k(t) \text{ is } V_k^i \text{ and } \dots \text{ and } \gamma_j^n(t) \text{ is } V_n^i \text{ then } Y_j^m(t) \text{ is } G_m^i \quad (7)$$

412 where  $V_k^i$  is the fuzzy set related to input variable  $\gamma_j^k(t)$  and  $G_m^i$  is the fuzzy set of the rail health  
 413 conditions selected based on the expert judgment for rule  $r_i$ . The minimum of the fuzzified input values is  
 414 given as the rule truth value of each rule as seen in Figure 7. The fuzzy set of the output is obtained by the  
 415 Mandami union operator over all the rules. To defuzzify the output, the center of gravity approach is  
 416 applied so as to obtain a crisp value. The fuzzy inference system (Mamdani) is to map the inputs (the  
 417 influential factors) to the output (the rail health condition) using a set of fuzzy rules. Thus, the fuzzy rules  
 418 are components of the fuzzy inference system. To set the fuzzy rules, a questionnaire is provided to  
 419 systematically analyse the combinations of possible inputs. As the judgment relies on the expert  
 420 knowledge, it is prone to bias. Thus, the investigation is used to support the experts on the validation of  
 421 the judgements. The inclusion of the investigation results in the questionnaire, helps the expert to visualize  
 422 the effect of  $\gamma_j^k(t)$  over the segment  $j$  on the actual rail health conditions. Furthermore, as the  
 423 questionnaire will lead to a model of the rail condition using the knowledge of expert, the expert qualified  
 424 to fill out the questionnaire is a rail maintenance engineer or a rail inspection expert. The expert should  
 425 have experience with both rail monitoring and rail maintenance. By using the proposed methodology, the  
 426 infrastructure management company will benefit from systematically keeping the knowledge of rail  
 427 experts in the company. So, in case a rail expert is not available, the railway company can still use the  
 428 previously developed rules or update them according to new infrastructure requirements. In the  
 429 questionnaire, two options are given including "influential" "non-influential". Then, the experts are asked  
 430 to rank between 1 and 2 the effect of the combination of influential factors into the health conditions of the

431 rail. A major contribution of the fuzzy system is to include non-crisp values (fuzzy values) in the output  
432 (the rail health condition). Although a binary approach is used for the questionnaire, (1) we can capture  
433 the fuzzy dynamics on the rail health condition and (2) we cover all the rule combinations. Otherwise and  
434 with using five-level ranking, number of the rules created would be too much time consuming for the  
435 experts whereas some of those rules would be useless in the decision making. Moreover, the five-level  
436 ranking is used to just improve the visualization quality of the interdependency analysis. The  
437 questionnaire is converted into a fuzzy inference system, where the rules are given by the options of the  
438 questionnaire (two possible fuzzy sets per influential factor) and the output fuzzy sets of each rule are  
439 given by the answers of the experts (three possible fuzzy sets).

440

#### 441 **6. Step 5: Rail maintenance decisions**

442

443 After estimating the rail health conditions for each segment, the entire rail can be evaluated according to the  
444 estimated health conditions. The aim is to find the most critical pieces of the track for the condition-based  
445 planning of grinding operations. Squats can be treated by grinding completely when they are at an early stage  
446 of growth or they can effectively be kept at safe level (to avoid having disastrous consequences) when they are  
447 severe. In this paper, a clustering method is proposed to grind the most critical pieces of the track efficiently  
448 based on predefined maintenance time slots determined by the infrastructure manager. As different tracks have  
449 different maintenance time slots, it is important to consider the available time slots to carry out the grinding  
450 operation. In the Dutch railway network, the time slots vary from one railway station to the next railway  
451 station. This means that not all segments of a long track that include different railway stations have the same  
452 maintenance time available for doing grinding. The grinding planning is formulated as in Figure 8.

453 As depicted in Figure 8, if maintenance time is still available after the grinding, the clustering approach can be  
454 applied to the other critical track pieces to effectively utilize the whole available maintenance time slot. Hence,  
455 the infrastructure manager makes sure that the maintenance time is fully used to avoid inducing extra  
456 maintenance costs.

457 The clustering approach strives to cover as many severe defects using as few clusters as possible within the  
458 limited maintenance time slot, which usually takes 8-10 hours at night in the Dutch railway network (this  
459 depends on the type of operations, and it could change per day, week, and year). The proposed clustering  
460 approach assigns a defect, e.g. a squat, to a cluster. The model includes not only the squat position, but also the  
461 squat severities acquired by the ABA system measurement and rail image data. The proposed grinding model is  
462 elaborated in the previous work of the authors (Su et al., 2017). Table 1 presents the notations used in the  
463 model.

464

465 "place Fig. 8 about here"

466  
467

468 "place Table 1 about here"

469

470 We call  $\left[\underline{\xi}, \bar{\xi}\right]$  the physical range, and clusters located within the physical range are called active clusters.

471 Also, the setup time,  $T_s$ , typically includes the machine travelling time, preparation time and finishing time  
472 for a maintenance operation. The start and end locations of the  $g$ -th cluster are the decision variables of the  
473 clustering problem. Thus, the grinding model can be formulated as the following non-smooth optimization  
474 problem (Su et al., 2017):

475

$$476 \quad \max_{\{d_g^{\text{start}}, d_g^{\text{end}}\}_{g=1}^{N_c}} \sum_{g=1}^{N_c} \sum_{l=1}^{N_d} \omega_l I_{d_g^{\text{start}} \leq X_l \leq d_g^{\text{end}}} + \sum_{g=1}^{N_c} I_{d_g^{\text{end}} > \bar{\xi}} \quad (8)$$

477 subject to

$$I_{d_g^{\text{start}} \leq X_l \leq d_g^{\text{end}}} = \begin{cases} 1 & \text{if } d_g^{\text{start}} \leq X_l \leq d_g^{\text{end}} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

$$I_{d_g^{\text{end}} > \bar{\xi}} = \begin{cases} 1 & \text{if } d_g^{\text{end}} > \bar{\xi} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$d_1^{\text{start}} \geq \underline{\xi} \quad (11)$$

$$478 \quad d_{N_c}^{\text{end}} \leq \bar{\xi} + 2N_c(\Delta d_{\min} + \varepsilon) \quad (12)$$

$$\Delta d_{\min} \leq d_g^{\text{end}} - d_g^{\text{start}} \leq \Delta d_{\max} \quad \forall g \in \{1, \dots, N_c\} \quad (13)$$

$$d_{g+1}^{\text{start}} - d_g^{\text{end}} \geq \varepsilon \quad \forall g \in \{1, \dots, N_c - 1\} \quad (14)$$

$$d_g^{\text{start}} \leq \bar{\xi} \Rightarrow d_g^{\text{end}} \leq \bar{\xi} \quad \forall g \in \{1, \dots, N_c\} \quad (15)$$

$$\sum_{g=1}^{N_c} I_{d_g^{\text{end}} \leq \bar{\xi}} \cdot \left( \frac{d_g^{\text{end}} - d_g^{\text{start}}}{v_G^{\text{on}}} + T_G^{\text{on}} + T_G^{\text{off}} \right) + \sum_{g=1}^{N_c-1} I_{d_g^{\text{end}} \leq \bar{\xi}} \cdot \frac{d_{g+1}^{\text{start}} - d_g^{\text{end}}}{v_G^{\text{off}}} \leq T_t - T_s \quad (16)$$

479 The indicator function  $I_\sigma$  takes value 1 if the statement  $\sigma$  is true, and 0 otherwise. The first term in the objective  
480 function (8) rewards the squats covered by a cluster depending on their severities, while the second term serves  
481 to maximize the number of non-active clusters, i.e. minimize the number of active clusters. The second term in  
482 (8) counts the number of clusters outside the physical range, i.e. non-active squats. As the total number of  
483 available clusters  $N_c$  is fixed, maximizing the number of non-active clusters is equivalent to minimizing the

484 number of active clusters. The active cluster is defined via the kilometre positions of its start and end points  
485 within the physical range  $[\underline{\xi}, \bar{\xi}]$ . A non-active cluster is outside the physical range and has no physical  
486 meaning. We use the idea of non-active cluster to be able to have idle clusters. Also,  $X_l$  is the kilometre location  
487 of the  $l$ -th squat. Constraints (11)-(12) set the distance range of the clusters. Note that the upper bound  $\xi_{\max}$  is  
488 set as indicated to allow the situation of non-active cluster, i.e. all clusters are located outside the physical range  
489  $[\underline{\xi}, \bar{\xi}]$ . The term  $\varepsilon$  in(12) is included to avoid the overlapping of clusters. To determine  $\varepsilon$ , we suggest just to  
490 take a tiny positive value (like  $\varepsilon = 0.001$  m). When  $\varepsilon$  is high, the distance between clusters will be higher and  
491 interesting rail pieces might not get covered by a cluster. Constraint (13) restricts the size of each cluster. The  
492 minimal and maximal size of a cluster is indeed determined by operational considerations of the grinding  
493 machine. The minimal size of a cluster is usually set to be the shortest length that the grinding machine can  
494 manipulate. The maximal size of a cluster should be less than the length of the rail considered. Constraint (13)  
495 restricts the size of each cluster. Constraint (14) ensures that the clusters are not overlapping, where the small  
496 positive parameter  $\varepsilon$  is the minimum distance between two clusters. So, there may be track sections between  
497 clusters that will be not included in the grinding planning. The constraint (15) forbids fractional clusters. The  
498 fractional cluster means that the start and end points of a cluster must both be inside or outside the physical  
499 range. We only allow to use active clusters (start and end points are both inside the physical range) and non-  
500 active clusters (start and end points both outside the physical range). The constraint (16) is the time limit  
501 constraint to ensure that the resulting clusters can be processed within the given maintenance time slot. The  
502 left-hand side of constraint (16) computes the total maintenance time, including the time to grind the active  
503 clusters (first term), the time for the machine to travel between the clusters (second term), and the setup time  $T_s$ .  
504 Constraint (16) guarantees that the total maintenance time to execute the clustering plan is less than the  
505 duration of the maintenance time slot  $T_s$ . The non-smooth optimization problem (8)-(16) can either be solved  
506 by gradient-free algorithms like pattern search and genetic algorithms, or transformed into an MILP problem  
507 following the standard procedure described in (Bemporad and Morari, 1999). In Su et al. (2017), the clustering  
508 method was employed as part of the low-level optimization, in a setup where the decisions are based on  
509 prediction including uncertainties via a scenario-based chance-constrained approach.

510

## 511 7. Numerical results

512 The track Amersfoort-Weert in the Netherlands is selected as a case study (nearly 125 kilometers of track). The  
513 track passes through Utrecht, Geldermalsen, 's-Hertogenbosch, and Eindhoven to reach the destination (Weert)  
514 (Figure 9). The whole track is partitioned into 15 segments to take all the signals of the influential factors per

515 segment into account. Also, the definition of the segments is based on track curvature, which means that each  
516 curve is included into one segment regardless the segment sizes.

517 The squat problem is aimed in the case study due to the fact that: (1) squats are one of the most commonly  
518 observed defects on rails, (2) squat-related costs are more than 5000 euro/km per year in The Netherlands.  
519 Although the rail grinding helps to treat all type of rail defects, e.g. corrugation, head checks and wheel burn,  
520 the optimal maintenance decisions proposed in the current paper focus on the squat problem and for the other  
521 rail defect types, it is crucial to take the effect of those defects in the maintenance decisions into account. For  
522 the estimation of the actual rail conditions, as explained in Section 2, the images are analysed using image  
523 processing to detect the ones including squats.

524

525 "place Fig. 9 about here"

526

527 The rail image analysis is defined based on the input images that are down-scaled to  $375 \times 275$  pixels and  
528 converted into gray-scale images. The sequence of three fully-connected layers translates the extracted high-  
529 level features from the previous layers into 3 classes representing the normal rail, trivial defects (seed squats),  
530 and squats. The normal class includes all the components in a healthy state, including plain rails, switches,  
531 welds, possible non-defect contaminations, etc. Trivial defects appear in the form of indentations or small  
532 damages to the rail head, while squats are usually defects that are fully grown deformations cracking the rail  
533 surface. The overlap between different rail images can cause mismatch between rail images and the ABA  
534 signal and might affect the estimation of the rail health conditions. To avoid this, first the video frames are pre-  
535 processed and the overlaps are removed. Then, we align images with the ABA signal using GPS tags and  
536 different reference points of the rail infrastructure (such as switches, crossings, joints, etc.). Figure 10 shows the  
537 mean absolute error of the detection algorithm as a function of the training epoch of the network for both  
538 training and validation data. The 75% of the data is assigned for the training and 25% is to validate the network  
539 performance. The samples of the labelled images are composed of 125 different squats collected from different  
540 locations of the track. Figure 11 shows the comparative predictions and the ground truth values for all samples  
541 in the test set. Thus, although the number of samples is limited, as the samples were picked up from different  
542 locations and vary from light to severe squats, one can argue that the dataset covers all the interesting cases.

543 "place Fig. 10 about here"

544 "place Fig. 11 about here"

545

546 Finally, the trained model is used with the new samples provided from the target track and predictions based  
547 purely on ABA are calculated. Figure 12 shows a sample plot of the results by the detection algorithm, which  
548 are used as the rail actual health conditions, and shows the position of the defects and their severity.

549 "place Fig. 12 about here"

550 The time needed for training is 40 hours per 1500 examples. Once the network is trained, it is used to find  
551 squats in the large pool of previously unseen samples (prediction). Unlike the training time, the prediction time  
552 is insignificant (30 seconds per 15000 examples). The prediction result then has an average binary accuracy  
553 (squat vs. normal) of 96.9 %. The detected squats are then analyzed in terms of the severity according to Step  
554 2.

555 In Figure 13, to perform the interdependency analysis we have compared the defect severities within a segment  
556 with each influential factor (track characteristic). This information can be used to guide the design of fuzzy  
557 rules created from interviews with experts about the relation between health conditions and influential factors.  
558 Based on the interdependencies, a set of fuzzy rules is defined to estimate the health conditions based on the  
559 influential factors as obtained in (3). All the rules are given the same weight. Moreover, all the input variables  
560 are combined through the rules. In this paper, 127 fuzzy if-then rules are generated in order to meet the possible  
561 interdependencies. Furthermore, based on the fuzzy rules, the sensitivities of the health conditions to the  
562 influential factors are captured as shown in Figure 14. This figure presents how the influential factors model the  
563 rail health conditions, varying from fully healthy (severity equal to zero) to completely unhealthy (severity  
564 equal to one), while all the other influential factors are assumed to be fully healthy (equal to zero). Three plots  
565 are used to show the sensitivity.

566 Variation of the inputs of an expert in the questionnaire can lead to different final maintenance  
567 decision results. Several experts are asked to fill out the questionnaire so that variations cause by  
568 single expert are reduced. Among all the influential factors, train speed has the highest effect on  
569 the grinding decision and superelevation has less influence. An increase of 20% in the train speed  
570 related influential factor gives an 8% increase on the rail health condition, whereas an increase of  
571 20% in the superelevation related influential factor gives 5%. A misestimation of 20% in a single  
572 factor gives at most 8% difference in final results error in the case of changing train speed related  
573 influential factor with superelevation related influential factor.

574  
575 "place Fig. 13 about here"

576

577 As an example in Figure 14(a), the effect of two input variables namely train speed,  $\gamma_j^1(t)$ , and train  
578 acceleration profile,  $\gamma_j^2(t)$ , respectively, is presented. As shown in the figure, the train speed changes over  
579 the track affect the rail health conditions more in comparison with the train acceleration profile. This is an  
580 indication of the importance of the train speed for maintenance decisions. Figure 14(b) depicts the  
581 influence of the speed profile versus the superelevation,  $\gamma_j^7(t)$ . The plot shows that the rail health  
582 conditions cannot get excited by the influence of the superelevation as much as the effect of the train  
583 speed profile. In Figure 14(c), the vehicle effect,  $\gamma_j^6(t)$ , is compared with the superelevation  $\gamma_j^7(t)$ . As  
584 can be seen in the plot, the both factors are not as influential as the train speed and the train acceleration  
585 on the rail health conditions. However, the vehicle effect can influence the health conditions more in  
586 comparison with the superelevation. Therefore, an increase in the most influencing factors, i. e.  $\gamma_j^1(t)$  and  
587  $\gamma_j^2(t)$  can increase the criticality of the segment up to requiring maintenance. If this criticality goes  
588 beyond the given rail health conditions of other rail segments of the track, then the grinding decision  
589 changes directly. Therefore, the infrastructure manager should take the segments with higher train speed  
590 profile and train acceleration into account in the maintenance plan.

591 "place Fig. 14 about here"

592 Relying on the fuzzy model, the rail health conditions is estimated. Each segment is evaluated based on the  
593 health conditions as shown in Figure 15 and Table 2. Table 2 presents the results of the case study. Given the  
594 influential factors, the rail health conditions based on the fuzzy inference system is estimated. Although some  
595 rules might not be needed as they might never apply in practice, we aimed to develop a questionnaire that  
596 captures all the possibilities to have a full coverage of inputs. Using the proposed inference system, any rail  
597 segment can be evaluated with given influential factors. Table 2 gives an example on how the fuzzy inference  
598 system performs. The influential factors are obtained from rail field measurements and the last column is  
599 calculated using the fuzzy rules.

600  
601 "place Table 2 about here"

602  
603 Figure 15 also indicates that  $j_7, j_8, j_9, j_{10}$  obtain the highest values of the rail health conditions. It means that  
604 those segments have a critical health conditions compared to other rail segments. These segments highlighted  
605 by the red line in the figure belong to the track between railway stations Geldermalsen and 's Hertogenbosch.  
606 Furthermore, the rail actual conditions (rail observation) is depicted in Figure 15. The figure shows the number

607 of squats over the full track from Amersfoort to Weert. The defects are detected based on the proposed  
608 detection model described in Section 2.

609 The segments with the most severe squats are distinguished by two different arrows in Figure 16. As seen in the  
610 figure, the segments 7 and 10 include the highest number of squats. Thus, the segments 7, 8, 9 and 10 are  
611 selected as the critical segments to be maintained. Depending on the available maintenance time slots, the track  
612 can be ground. If after grinding the above-mentioned segments, there is time to perform maintenance in the rest  
613 of the rail network, the segments 13 and 15 are candidates to be maintained (marked by black arrows). From  
614 the figure,  $j_{13}$  has more squats than  $j_{15}$ , but according to Figure 15,  $j_{13}$  is more critical in terms of the health  
615 conditions and also it is shorter in length, which increases the squat density distributed on the track. Therefore,  
616 segment 15 is chosen as the alternative option.

617 "place Fig. 15 about here"

618 "place Fig. 16 about here"

619  
620 Once the critical segments are determined, the optimal clustering model is used to cover squats subject to the  
621 time limit imposed by the maintenance slot. The proposed clustering model is able to treat the most important  
622 squats. In this paper, the length of the maintenance time slot is set to 8 hour. This 8 hour time period includes  
623 the setup time which covers the time required for transportation, machinery, personnel, etc. The most relevant  
624 squats are covered by a cluster, as the clustering model penalizes a squat outside any cluster by its severity.  
625 Hence, the most important squats are treated by grinding, even when the maintenance slot is not long enough.  
626 This is not normally the case for cyclic grinding, which is currently the most used method in the Dutch network  
627 in which the full track is ground. For cyclic grinding, the grinding machine starts grinding from the start point  
628 going towards to the end of the track without any guarantee to capture the most important squats. Figure 17  
629 shows the clustering result between the stations Geldermalsen and 's Hertogenbosch covering the critical  
630 segments, i.e.  $j_7, j_8, j_9, j_{10}$ . The target track is around 20 km as shown in the x-axis of the figure. According to  
631 the proposed detection model, the squat severity is estimated as indicated in the y-axis. The grinding model  
632 proposes two clusters within the maintenance time slot capturing the most severe squats by considering the  
633 density of the squats. In this way, the grinding machine starts grinding from the beginning of the track to reach  
634 the 52.42 kilometre, then the machine stops working to drive to cluster two (the transfer time is supposed  
635 negligible), which starts at the track position 60.51 km until the end of the track. Moreover, the number of the  
636 defects between (remaining track piece) 52.46 km and 60.51 km is much less (43 defects and average severity  
637 2.10) than in the first cluster between 46 km until 52.4 km (77 defects and average severity 2.15) and the  
638 second cluster between 60.51 km until 66 km (187 defects and average severity 2.25). Thus, although we have

639 defects between 52.46 and 60.51, by considering (1) the maintenance time slot limitation and (2) maintenance  
640 priority of the segments 7 and 10 (Figure 17) in terms of higher value of health condition, the defects between  
641 52.42 and 60.51 remain with no maintenance intervention until the next maintenance time slot. Without the  
642 proposed clustering model, the grinding machine will not be able to capture the most important squats, either at  
643 the beginning of the track or at the end of the track. Some severe squats would therefore remain untreated,  
644 which would increase maintenance costs and the probability of rail failure.

645  
646 The cost to employ the grinding machine is 35k euro for one night considering 10 hours. Note that 10 hour is  
647 fixed meaning that for shorter maintenance time slots (shorter than 10 hours) the cost is the same. Thus, the  
648 infrastructure manager will be charged the same amount of money, although the machine is used for less than  
649 10 hours. Thus, in case there would be 2 hours extra time available after finishing the grinding of the critical  
650 rail pieces, the infrastructure manager has the chance to fill all the available time to keep the grinding machine  
651 running. In that case, according to the proposed methodology, the grinding machine can be transferred to the  
652 segment 10,  $j_{10}$ , which has a more critical health conditions compared to the rest of the target track. Then, the  
653 infrastructure manager can ensure that traffic-free hours have been used to treat all the most important squats  
654 over a long track.

655

656 "place Fig. 17 about here"

657

658

## 659 **8. Conclusions**

660 In this paper, we have proposed an integrated approach for maintenance decision system of the railway  
661 infrastructures. The methodology includes infrastructure conditions monitoring and maintenance decision  
662 making. The proposed approach is applied to the condition-based treatment of squats, with big data information  
663 coming from a track in the Dutch railway network. The algorithm makes use of both axle box acceleration  
664 signals and rail video images, which contribute a huge amount of data. The use of both rail data sources  
665 reduces the detection error of the surface defects. Moreover, we have used the track characteristics of the Dutch  
666 railway network, enabling the infrastructure manager to interconnect the track influential factors with the actual  
667 rail health conditions. We therefore investigated how to define a list of decision actions to support the decisions  
668 regarding the maintenance plan by analysing the above-mentioned interdependency. The results propose a  
669 maintenance decision approach based on the actual conditions of the rails but together with the insights  
670 resulting from the influential factors. We proposed a partitioning in 15 different segments for a track that  
671 can be considered quite long (105 km). The maintenance decision system is proposed using a clustering  
672 model to perform grinding over the critical pieces of the rail. The results include the most severe squats covered

673 by the maintenance clusters. Thus, although not all the squats are treated, the infrastructure manager can make  
674 sure that there is considerably less safety risk or high maintenance cost until the next rail measurement  
675 campaign. To include possible practical limitations. Then, we include the maintenance time slot as a constraint  
676 problem. Different pass numbers of the grinding machine, resulting in different grinding depths, have an impact  
677 on the rail defect risk after grinding. Different pass numbers also lead to different grinding speeds. The current  
678 clustering model considers only one grinding depth, meaning one fixed pass number and grinding speed.

679 While this paper is focused on the analysis of squats, the results are applicable to the analysis of other types of  
680 rail defects like corrugations, damaged insulated joints, welds and other types of RCF defects. To apply the  
681 proposed methodology to all those defects, the infrastructure manager will need to analyse the rail observations  
682 in terms of that specific type of defect versus the track characteristics to define the list of decision rules. The  
683 methodology for the design of the rules is flexible, so they can be adapted to different railway networks. In the  
684 further research the interdependency analysis can be conducted at a more detailed level, for instance at every  
685 kilometre or even at meter of track. In future research, based on the influential factors it will be possible to  
686 anticipate the rail conditions much better, so predictive maintenance could be achieved. The maintenance  
687 operations could be different from one type of defect to another, but the general methodology can be adapted,  
688 as far as the defects can be grouped into different clusters. In addition, the proposed methodology can be linked  
689 to a rail maintenance cost analysis to reduce life cycle cost (LCC). Also, by having different measurement sets  
690 of rail data, a prediction model of how the defects can grow over time could be added to the methodology,  
691 correlated to the influential factors. This will help the infrastructure manager to predict the rail health  
692 conditions in advance and also to prolong the maintenance decision time horizon. In the future work, we will  
693 consider a flexible number of passes of the grinding machine to obtain more efficient clustering plans. Another  
694 topic for further research is to evaluate the methodology for different regions to investigate the influence of  
695 exogenous factors like environmental factors to the decision rules and consequently the maintenance decision  
696 rules.

697

## 698 **ACKNOWLEDGEMENTS**

699 This research is part of the NWO/ProRail project "Multi-party risk management and key performance indicator  
700 design at the whole system level (PYRAMIDS)", project code 438-12-300, and the STW/ProRail project  
701 "Advanced monitoring of intelligent rail infrastructure (ADMIRE)", project 12235, which are partly funded by  
702 the Ministry of Economic Affairs. The authors also would like to thank INSPECTION for providing us with  
703 image data and technical support.

704

705 **REFERENCES**

- 706 1. Åhrén, T. and Parida, A., 2009. Maintenance performance indicators (MPIs) for benchmarking the railway  
707 infrastructure: a case study. *Benchmarking: An International Journal*, 16(2), pp.247-258.
- 708 2. Andrade, A.R. and Teixeira, P.F., 2011. Uncertainty in rail-track geometry degradation: Lisbon-Oporto line case  
709 study. *Journal of transportation engineering*, 137(3), pp.193-200.
- 710 3. Andrade, A.R. and Teixeira, P.F., 2012. A Bayesian model to assess rail track geometry degradation through its  
711 life-cycle. *Research in transportation economics*, 36(1), pp.1-8.
- 712 4. Attoh-Okine, N.O., 2017. *Big Data and Differential Privacy: Analysis Strategies for Railway Track Engineering*.  
713 John Wiley & Sons.
- 714 5. Bemporad, A. and Morari, M., 1999. Control of systems integrating logic, dynamics, and constraints.  
715 *Automatica*, 35(3), pp.407-427.
- 716 6. Bing, A.J. and Gross, A., 1983. Development of railroad track degradation models. *Transportation research*  
717 *record*, (939) , pp. 27-31.
- 718 7. Budai, G., Huisman, D. and Dekker, R., 2006. Scheduling preventive railway maintenance activities. *Journal of*  
719 *the Operational Research Society*, 57(9), pp.1035-1044.
- 720 8. Caetano, L.F. and Teixeira, P.F., 2016. Strategic model to optimize railway-track renewal operations at a network  
721 level. *Journal of Infrastructure Systems*, 22(2), p.04016002.
- 722 9. Camastra, F., Ciaramella, A., Giovannelli, V., Lener, M., Rastelli, V., Staiano, A., Staiano, G. and Starace, A.,  
723 2015. A fuzzy decision system for genetically modified plant environmental risk assessment using Mamdani  
724 inference. *Expert Systems with Applications*, 42(3), pp.1710-1716.
- 725 10. Corbin, J.C. and Fazio, A.E., 1981. Performance-based track-quality measures and their application to  
726 maintenance-of-way planning. *Transportation Research Record*, (802), pp.19-27.
- 727 11. Faghih-Roohi, S., Hajizadeh, S., Núñez, A., Babuska, R. and De Schutter, B., 2016, July. Deep convolutional  
728 neural networks for detection of rail surface defects. *International Joint Conference on Neural Networks (IJCNN)*,  
729 (pp. 2584-2589).
- 730 12. Fan, Y., Dixon, S., Edwards, R.S. and Jian, X., 2007. Ultrasonic surface wave propagation and interaction with  
731 surface defects on rail track head. *Ndt & E International*, 40(6), pp.471-477.
- 732 13. Fumeo, E., Oneto, L. and Anguita, D., 2015. Condition based maintenance in railway transportation systems  
733 based on big data streaming analysis. *Procedia Computer Science*, 53, pp.437-446.
- 734 14. Gandomi, A. and Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International*  
735 *Journal of Information Management*, 35(2), pp.137-144.
- 736 15. Ghofrani, F., He, Q., Goverde, R.M. and Liu, X., 2018. Recent applications of big data analytics in railway  
737 transportation systems: A survey. *Transportation Research Part C: Emerging Technologies*, 90, pp.226-246.
- 738 16. Grassie, S.L. and Elkins, J.A., 2005. Tractive effort, curving and surface damage of rails: Part 1. Forces exerted  
739 on the rails. *Wear*, 258(7-8), pp.1235-1244.
- 740 17. Grassie, S.L., 2012. Squats and squat-type defects in rails: the understanding to date. *Proceedings of the*  
741 *Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226(3), pp.235-242.

- 742 18. Hajizadeh, S., Li, Z., Dollevoet, R.P. and Tax, D.M., 2014, August. Evaluating Classification Performance with  
743 only Positive and Unlabeled Samples. In *Joint IAPR International Workshops on Statistical Techniques in Pattern*  
744 *Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (pp. 233-242). Springer, Berlin,  
745 Heidelberg.
- 746 19. Hamid, A. and Gross, A., 1981. Track-quality indices and track degradation models for maintenance-of-way  
747 planning. *Transportation Research Board*, 802, pp.2-8.
- 748 20. He, Q., Kamarianakis, Y., Jintanakul, K. and Wynter, L., 2013. Incident duration prediction with hybrid  
749 tree-based quantile regression. In *Advances in Dynamic Network Modeling in Complex Transportation*  
750 *Systems* (pp. 287-305). Springer, New York, NY.
- 751 21. He, Q., Li, H., Bhattacharjya, D., Parikh, D.P. and Hampapur, A., 2015. Track geometry defect rectification  
752 based on track deterioration modelling and derailment risk assessment. *Journal of the Operational Research*  
753 *Society*, 66(3), pp.392-404.
- 754 22. Heinicke, F., Simroth, A., Scheithauer, G. and Fischer, A., 2015. A railway maintenance scheduling problem with  
755 customer costs. *EURO Journal on Transportation and Logistics*, 4(1), pp.113-137.
- 756 23. Islam, D.M.Z., Laparidou, K. and Burgess, A., 2016. Cost effective future derailment mitigation techniques for  
757 rail freight traffic management in Europe. *Transportation Research Part C: Emerging Technologies*, 70, pp.185-  
758 196.
- 759 24. Jamshidi, A., Faghih-Roohi, S., Hajizadeh, S., Núñez, A., Babuska, R., Dollevoet, R., Li, Z. and Schutter, B.,  
760 2017. A big data analysis approach for rail failure risk assessment. *Risk analysis*, 37(8), pp.1495-1507.
- 761 25. Jamshidi, A., Núñez, A., Dollevoet, R. and Li, Z., 2017. Robust and predictive fuzzy key performance indicators  
762 for condition-based treatment of squats in railway infrastructures. *Journal of Infrastructure Systems*, 23(3),  
763 p.04017006.
- 764 26. Kawaguchi, A., Miwa, M. and Terada, K., 2005. Actual data analysis of alignment irregularity growth and its  
765 prediction model. *Quarterly report of RTRI*, 46(4), pp.262-268.
- 766 27. Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural  
767 networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- 768 28. Lasisi, A. and Attoh-Okine, N., 2018. Principal components analysis and track quality index: A machine learning  
769 approach. *Transportation Research Part C: Emerging Technologies*, 91, pp.230-248.
- 770 29. LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, 521(7553), p.436.
- 771 30. Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D. and Hampapur, A., 2014. Improving rail network  
772 velocity: A machine learning approach to predictive maintenance. *Transportation Research Part C:*  
773 *Emerging Technologies*, 45, pp.17-26.
- 774 31. Li, Z., 2009. Squats on railway rails. In *Wheel–Rail Interface Handbook*, Woodhead Publishing Books , pp. 409-  
775 436.
- 776 32. Li, Z., Dollevoet, R., Molodova, M. and Zhao, X., 2011. Squat growth—Some observations and the validation of  
777 numerical predictions. *Wear*, 271(1-2), pp.148-157.

- 778 33. Li, Z., Molodova, M., Núñez, A. and Dollevoet, R., 2015. Improvements in axle box acceleration measurements  
779 for the detection of light squats in railway infrastructure. *IEEE Transactions on Industrial Electronics*, 62(7),  
780 pp.4385-4397.
- 781 34. Li, Z., Zhao, X., Esveld, C., Dollevoet, R. and Molodova, M., 2008. An investigation into the causes of squats—  
782 Correlation analysis and numerical modeling. *Wear*, 265(9-10), pp.1349-1355.
- 783 35. Liu, X. and Dick, C.T., 2016. Risk-based optimization of rail defect inspection frequency for petroleum crude oil  
784 transportation. *Transportation Research Record: Journal of the Transportation Research Board*, (2545), pp.27-35.
- 785 36. Liu, X., Barkan, C. and Saat, M., 2011. Analysis of derailments by accident cause: evaluating railroad track  
786 upgrades to reduce transportation risk. *Transportation Research Record: Journal of the Transportation Research*  
787 *Board*, (2261), pp.178-185.
- 788 37. Liu, X., Saat, M. and Barkan, C., 2012. Analysis of causes of major train derailment and their effect on accident  
789 rates. *Transportation Research Record: Journal of the Transportation Research Board*, (2289), pp.154-163.
- 790 38. Ma, F., Wu, Q., Yan, X., Chu, X. and Zhang, D., 2015. Classification of automatic radar plotting aid targets based  
791 on improved fuzzy C-means. *Transportation research part c: emerging technologies*, 51, pp.180-195.
- 792 39. Makino, T., Kato, T. and Hirakawa, K., 2012. The effect of slip ratio on the rolling contact fatigue property of  
793 railway wheel steel. *International Journal of Fatigue*, 36(1), pp.68-79.
- 794 40. Makino, T., Neishi, Y., Shiozawa, D., Kikuchi, S., Okada, S., Kajiwara, K. and Nakai, Y., 2016. Effect of defect  
795 shape on rolling contact fatigue crack initiation and propagation in high strength steel. *International Journal of*  
796 *Fatigue*, 92, pp.507-516.
- 797 41. Mariani, S., Nguyen, T., Phillips, R.R., Kijanka, P., Lanza di Scalea, F., Staszewski, W.J., Fateh, M. and Carr, G.,  
798 2013. Noncontact ultrasonic guided wave inspection of rails. *Structural Health Monitoring*, 12(5-6), pp.539-548.
- 799 42. Markowski, A.S. and Mannan, M.S., 2009. Fuzzy logic for piping risk assessment (pFLOPA). *Journal of loss*  
800 *prevention in the process industries*, 22(6), pp.921-927.
- 801 43. Molodova, M., Li, Z., Núñez, A. and Dollevoet, R., 2014. Automatic detection of squats in railway infrastructure.  
802 *IEEE Transactions on Intelligent Transportation Systems*, 15(5), pp.1980-1990.
- 803 44. Mutton, P.J., Epp, C.J. and Dudek, J., 1991. Rolling contact fatigue in railway wheels under high axle loads. In  
804 *Mechanics and Fatigue in Wheel/Rail Contact* (pp. 139-152).
- 805 45. Nielsen, J.C., Ekberg, A. and Lundén, R., 2005. Influence of short-pitch wheel/rail corrugation on rolling contact  
806 fatigue of railway wheels. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and*  
807 *Rapid Transit*, 219(3), pp.177-187.
- 808 46. Parida, A. and Chattopadhyay, G., 2007. Development of a multi-criteria hierarchical framework for maintenance  
809 performance measurement (MPM). *Journal of Quality in maintenance Engineering*, 13(3), pp.241-258.
- 810 47. Peng, F. and Ouyang, Y., 2012. Track maintenance production team scheduling in railroad networks.  
811 *Transportation Research Part B: Methodological*, 46(10), pp.1474-1488.
- 812 48. Peng, F. and Ouyang, Y., 2014. Optimal clustering of railroad track maintenance jobs. *Computer-Aided Civil and*  
813 *Infrastructure Engineering*, 29(4), pp.235-247.
- 814 49. Popović, Z., Radović, V., Lazarević, L., Vukadinović, V. and Tepić, G., 2013. Rail inspection of RCF defects.  
815 *Metalurgija*, 52(4), pp.537-540.

- 816 50. Santos, R. and Teixeira, P.F., 2011. Heuristic analysis of the effective range of a track tamping machine. *Journal*  
817 *of Infrastructure Systems*, 18(4), pp.314-322.
- 818 51. Sciammarella, C.A., Chen, R.J.S., Gallo, P., Berto, F. and Lamberti, L., 2016. Experimental evaluation of rolling  
819 contact fatigue in railroad wheels. *International Journal of Fatigue*, 91, pp.158-170.
- 820 52. Scott, D., Fletcher, D.I. and Cardwell, B.J., 2014. Simulation study of thermally initiated rail defects. *Proceedings*  
821 *of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 228(2), pp.113-127.
- 822 53. Sharma, S., Cui, Y., He, Q., Mohammadi, R. and Li, Z., 2018. Data-driven optimization of Railway  
823 maintenance for track geometry. *Transportation Research Part C: Emerging Technologies*, 90, pp.34-58.
- 824 54. Song, Z., Yamada, T., Shitara, H. and Takemura, Y., 2011. Detection of damage and crack in railhead by using  
825 eddy current testing. *Journal of Electromagnetic Analysis and Applications*, 3(12), p.546- 550.
- 826 55. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: A simple way to  
827 prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), pp.1929-1958.
- 828 56. Su, Z., Jamshidi, A., Núñez, A., Baldi, S. and De Schutter, B., 2017. Multi-level condition-based maintenance  
829 planning for railway infrastructures—A scenario-based chance-constrained approach. *Transportation Research Part*  
830 *C: Emerging Technologies*, 84, pp.92-123.
- 831 57. Tosun, M., Dincer, K. and Baskaya, S., 2011. Rule-based Mamdani-type fuzzy modelling of thermal  
832 performance of multi-layer precast concrete panels used in residential buildings in Turkey. *Expert Systems with*  
833 *Applications*, 38(5), pp.5553-5560.
- 834 58. Veit, P., 2007. Track quality—luxury or necessity. *Railway Technical Review Special: Maintenance and Renewal*,  
835 pp.8-12.
- 836 59. Vo, K.D., Zhu, H.T., Tieu, A.K. and Kosasih, P.B., 2015. Comparisons of stress, heat and wear generated by AC  
837 versus DC locomotives under diverse operational conditions. *Wear*, 328, pp.186-196.
- 838 60. Wang, L., Wang, P., Quan, S. and Chen, R., 2012. The effect of track alignment irregularity on wheel-rail contact  
839 geometry relationship in a turnout zone. *Journal of Modern Transportation*, 20(3), pp.148-152.
- 840 61. Wen, M., Li, R. and Salling, K.B., 2016. Optimization of preventive condition-based tamping for railway tracks.  
841 *European Journal of Operational Research*, 252(2), pp.455-465.
- 842 62. Xie, M., 2003. *Fundamentals of robotics: linking perception to action*, World Scientific Publishing Company, 54.
- 843 63. Xu, L. and Zhai, W., 2017. A novel model for determining the amplitude-wavelength limits of track irregularities  
844 accompanied by a reliability assessment in railway vehicle-track dynamics. *Mechanical Systems and Signal*  
845 *Processing*, 86, pp.260-277.
- 846 64. Zhuang, L., Wang, L., Zhang, Z. and Tsui, K.L., 2018. Automated vision inspection of rail surface cracks: A  
847 double-layer data-driven framework. *Transportation Research Part C: Emerging Technologies*, 92, pp.258-277.
- 848 65. Zoeteman, A., Dollevoet, R. and Li, Z., 2014. Dutch research results on wheel/rail interface management: 2001–  
849 2013 and beyond. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid*  
850 *Transit*, 228(6), pp.642-651.
- 851 66. Zywił, J. and Oberlechner, G., 2001. Innovative measuring system unveiled. *International Railway Journal*,  
852 41(9).
- 853

854 **Appendix 1:** To generate the fuzzy rules, a questionnaire was filled out by an expert. In the questionnaire we asked  
855 to use linguistic terms e.g. non-influential=0 and influential=1 for the factors (column 1 to column 7). Then a score  
856 between 0 to 2 is given to the rail health conditions (the last column) considering the combination of situations from  
857 the influential factors. Healthy = 0, Average =1, Unhealthy=2.

858

859

“place Table 3 about here”

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888 **Appendix 2:** Transformation of the non-smooth optimization problem into an MILP problem, (8)-(16),  
 889 according to the standard procedure described in Bemporad and Morari (1999).

890

891 First, we introduce the following binary variables:

$$\delta_{g,l}^{\text{end}} = 1 \Leftrightarrow d_g^{\text{end}} - x_l \leq 0 \quad (17)$$

$$\delta_{g,l}^{\text{start}} = 1 \Leftrightarrow d_g^{\text{start}} - x_l \leq 0 \quad (18)$$

892 
$$\bar{\delta}_g = 1 \Leftrightarrow d_g^{\text{end}} - \bar{\xi} \leq 0 \quad (19)$$

$$\underline{\delta}_g = 1 \Leftrightarrow d_g^{\text{start}} - \underline{\xi} \leq 0 \quad (20)$$

$$\forall g \in \{1, \dots, N_c\}, \forall l \in \{1, \dots, N_d\}.$$

893 Then we introduce the following variables:

$$z_{1,g} = \underline{\delta}_g d_g^{\text{start}} \quad (21)$$

894 
$$z_{2,g} = \bar{\delta}_g d_g^{\text{end}} \quad (22)$$

$$z_{3,g} = \underline{\delta}_g d_{g+1}^{\text{start}} \quad (23)$$

$$\forall i \in \{1, \dots, N_c\}.$$

895 Define

$$\delta = \left[ \underbrace{\delta_{1,1}^{\text{start}} \dots \delta_{N_c, N_d}^{\text{start}} \delta_{1,1}^{\text{end}} \dots \delta_{N_c, N_d}^{\text{end}}}_{(\delta^{\text{start}})^T \text{ and } (\delta^{\text{end}})^T} \quad \underbrace{\underline{\delta}_1 \dots \underline{\delta}_{N_c}}_{\underline{\delta}^T} \quad \underbrace{\bar{\delta}_1 \dots \bar{\delta}_{N_c}}_{\bar{\delta}^T} \right]^T$$

896

$$z = \left[ \underbrace{z_{1,1} \dots z_{1, N_c}}_{z_1^T} \quad \underbrace{z_{2,1} \dots z_{2, N_c}}_{z_2^T} \quad \underbrace{z_{3,1} \dots z_{3, N_c}}_{z_3^T} \right]^T$$

897 These equations can be transformed into equivalent mixed integer linear model expressed as:

898 
$$\min_{d, \delta, z} \sum_{g=1}^{N_c} \left( \sum_{l=1}^{N_d} \omega_l (\delta_{g,l}^{\text{end}} - \delta_{g,l}^{\text{start}}) + \bar{\delta}_g \right) \quad (24)$$

899 subject to constraint (11) -(14) and

900 
$$\underline{\delta}_g - \bar{\delta}_g = 0 \quad \forall g \in \{1, \dots, N_c\} \quad (25)$$

901 
$$\sum_{i=1}^{N_c-1} \left( \left( \frac{1}{v_G^{\text{on}}} - \frac{1}{v_G^{\text{off}}} \right) z_{2,g} - \frac{1}{v_G^{\text{on}}} z_{1,g} + \frac{1}{v_G^{\text{off}}} z_{3,g} + (T_G^{\text{on}} + T_G^{\text{off}}) \bar{\delta}_g \right) \quad (26)$$

$$+ \frac{1}{v_G^{\text{on}}} (z_{2, N_c} - z_{1, N_c}) + (T_G^{\text{on}} + T_G^{\text{off}}) \bar{\delta}_{N_c} \leq T_t - T_s$$

902

903

904

905

906 **Figure captions**

907 Figure 1: Flowchart of the proposed methodology.

908 Figure 2: Defect severity analysis via ABA signals and image data using on-board train  
909 measurement. In the scheme, a severe squat is shown ( $S_4$ ). The actual measurements were  
910 obtained from two different trains: the CTO Train for the ABA signals, and the Inspection  
911 measurement train for the video images.

912 Figure 3: ABA signals including acceleration matched with rail video frames.

913 Figure 4: The prediction model consists of 3 one-dimensional convolutional layers each  
914 followed by a max-pooling operation, and a ReLU activated dense layer on top, which results  
915 in the final scalar estimation of the severity.

916 Figure 5: A schematic representative of the map of the track influential factors in a piece of  
917 the track.

918 Figure 6: Membership degree for all the segments based on their influential factors. The  
919 highlighted rectangle indicates that rail segment number 5 belongs highly to the cluster 2  
920 (membership degree almost one, indicated by the arrow).

921 Figure 7: Generic structure of the fuzzy decision model to compute the rail health conditions.

922 Figure 8: The proposed simplified grinding planning scheme.

923 Figure 9: Schematic track map between two stations, Amersfoort and Weert.

924 Figure 10: Mean absolute error (MAE) between the ground truth severities and the  
925 predictions. The network is trained using 75% of the data and is validated on the remaining  
926 25%.

927 Figure 11: Ground truth values provided by a human expert by estimating the defect severities  
928 from defect images vs prediction of the severity level from the ABA signal.

929 Figure 12: A sample of defect locations versus defect severity between kilometre 33 and 33.9  
930 in the track Amersfoort-Weert.

931 Figure 13: interdependency analysis between defect severity and track influential factors over  
932 15 track segments where  $\gamma_j^1(t)$ ,  $\gamma_j^2(t)$ ,  $\gamma_j^3(t)$ ,  $\gamma_j^4(t)$ ,  $\gamma_j^5(t)$ ,  $\gamma_j^6(t)$  and  $\gamma_j^7(t)$  are the train  
933 speed profile (m/s), train acceleration profile (m/s<sup>2</sup>), track horizontal curve (mm), track  
934 geometry parameter (measured at 40 km/h), rail head wear (mm), vehicle effect, and track  
935 superelevation (mm), respectively.

936

937 Figure 14: Examples of how the fuzzy rail conditions rules are defined according to the  
938 interdependency analysis.

939 Figure 15: Rail health conditions over the track segment showing the most critical pieces of  
940 the track highlighted with a red line on the map.

941 Figure 16: Number of moderate and severe squats within different segments. Red arrows  
942 show the target segments including the most severe squats and black arrows indicate the  
943 alternative options for grinding in case the enough time remains within the maintenance time  
944 a lot after maintaining the target segments.

945 Figure 17: Result of the grinding model that determines the optimal clustering of the target  
946 track. Squats are marked by either a colored square or a colored circle if they are covered by a  
947 cluster.

948

949            **Table captions**

950            Table 1: A list of the notations of the clustering model.

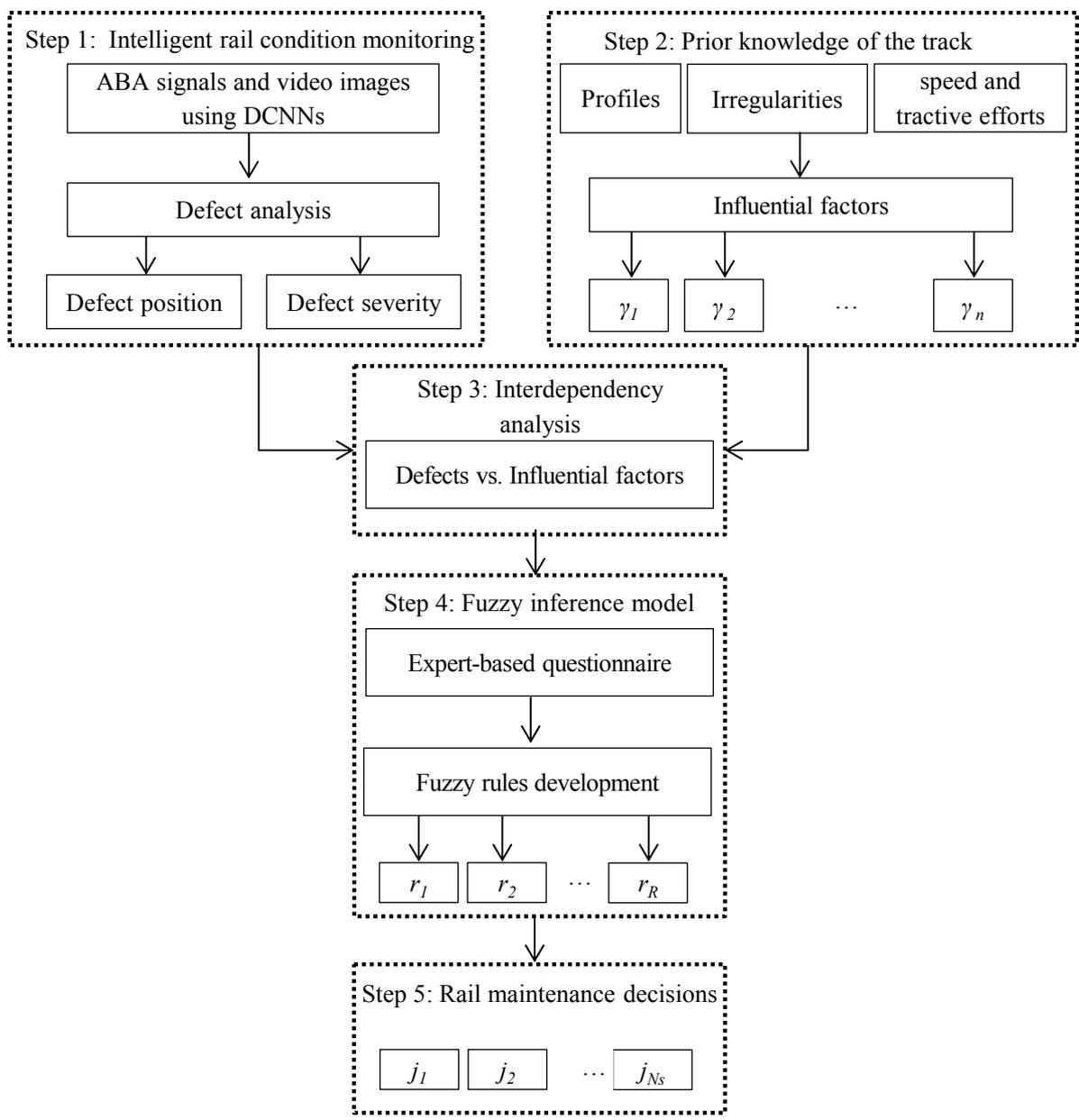
951            Table 2: Calculated influential factors per segment and estimated rail health conditions using  
952            the proposed fuzzy inference system.

953            Table 3: The questionnaire.

954

955  
956

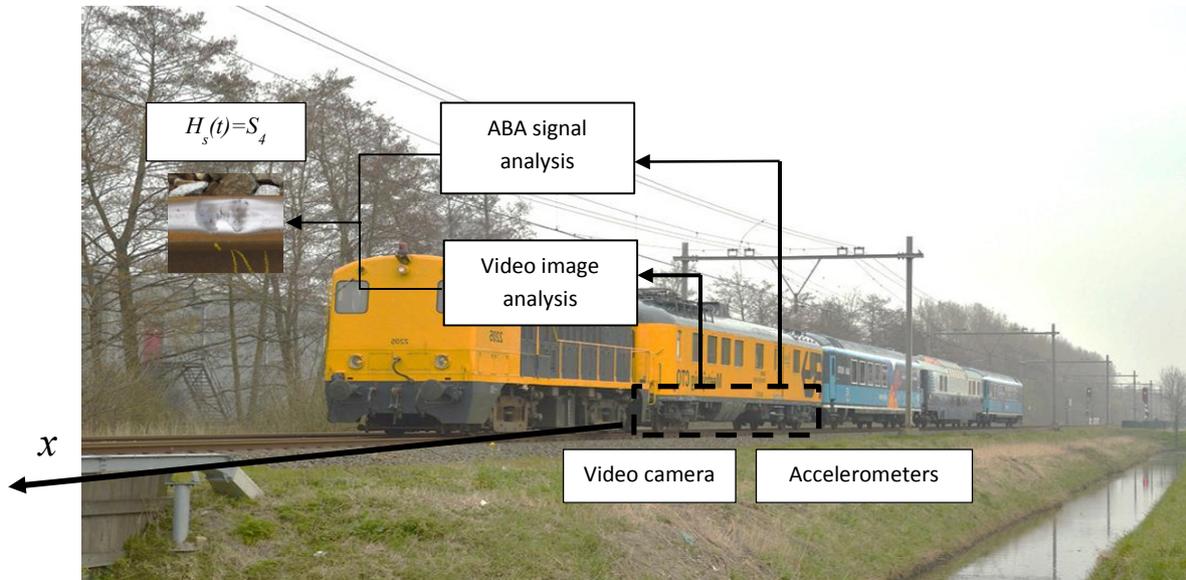
**Figure 1**



957  
958  
959  
960  
961  
962  
963  
964

965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998

**Figure 2**

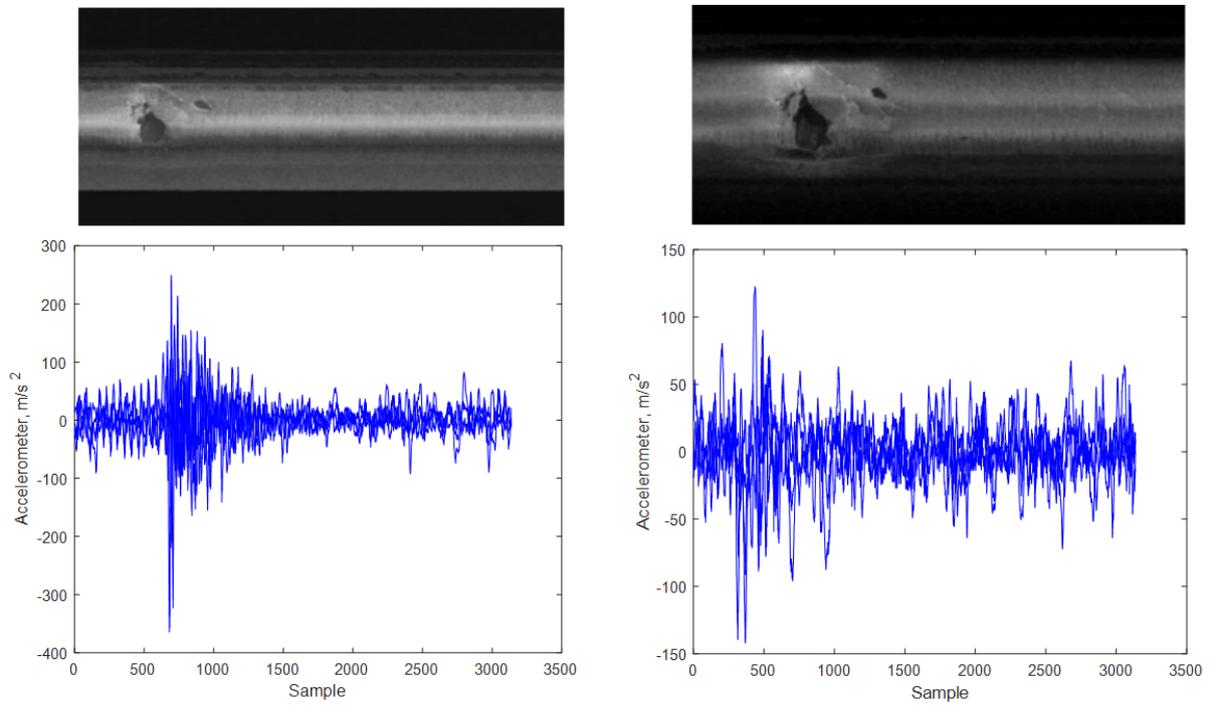


999

1000

1001

**Figure 3**



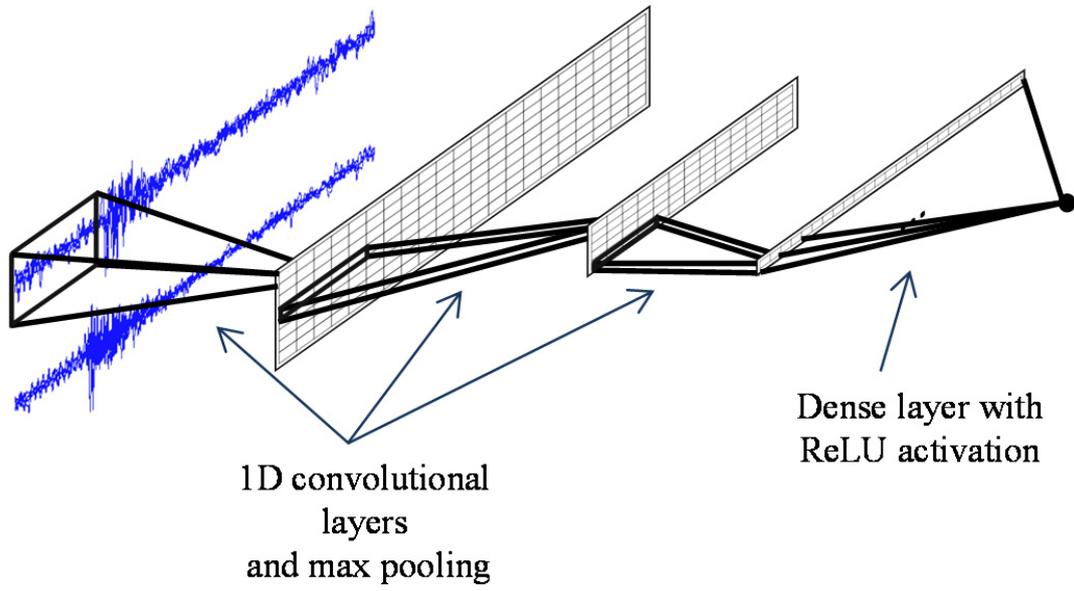
1002

1003

1004

1005  
1006  
1007

**Figure 4**

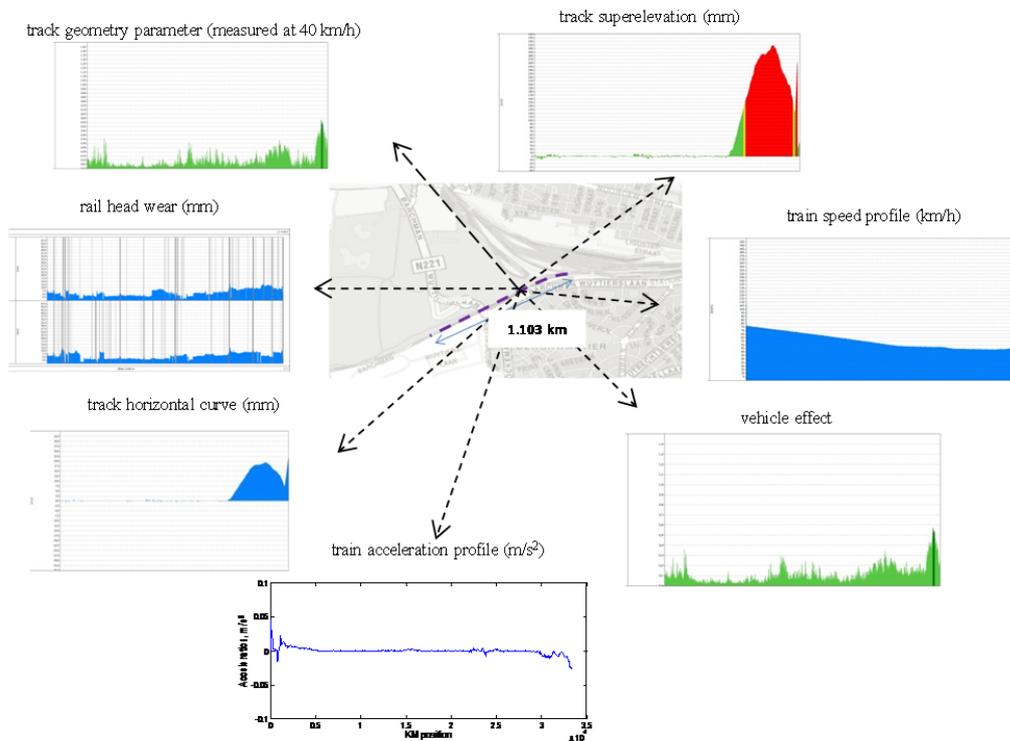


1008  
1009

1010

**Figure 5**

1011



1012

1013

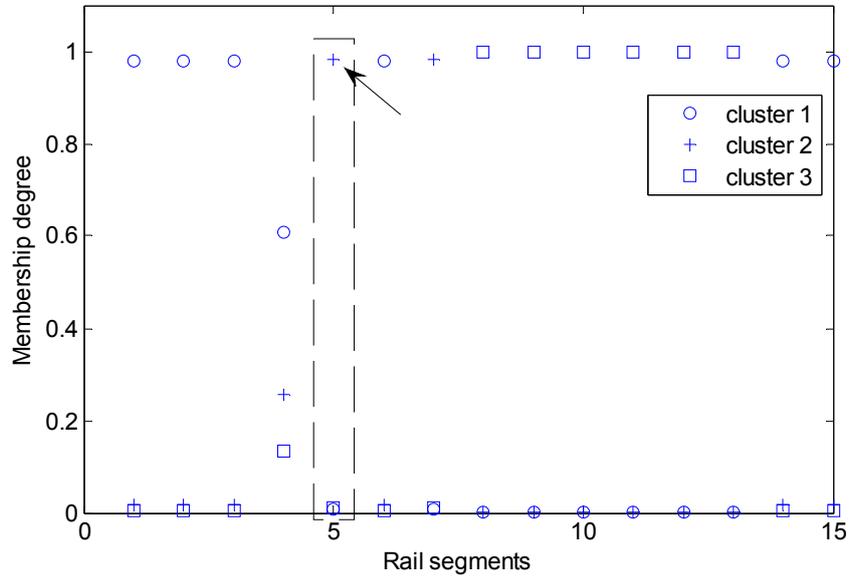
1014

1015

1016 **Figure 6**

1017

1018



1019

1020

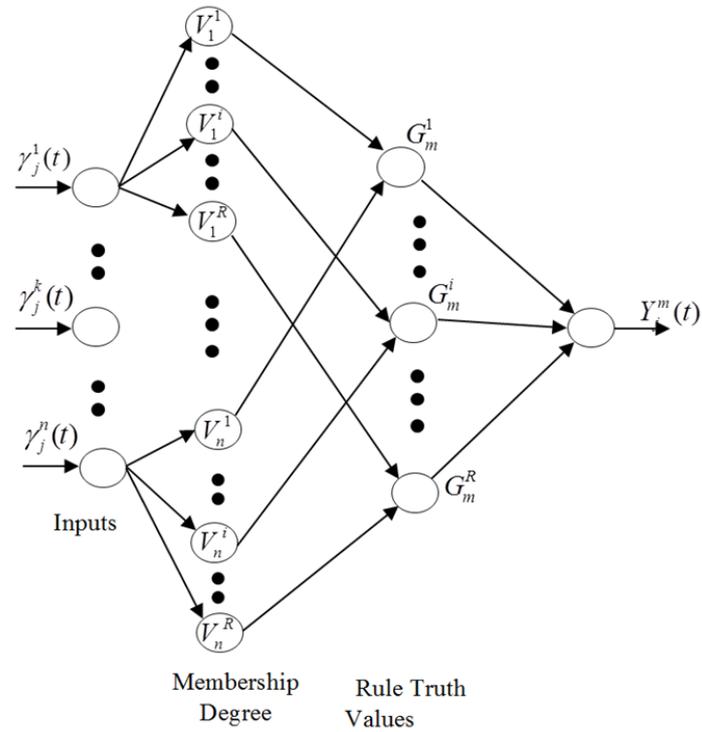
1021

1022

1023 **Figure 7**

1024

1025



1026

1027

1028

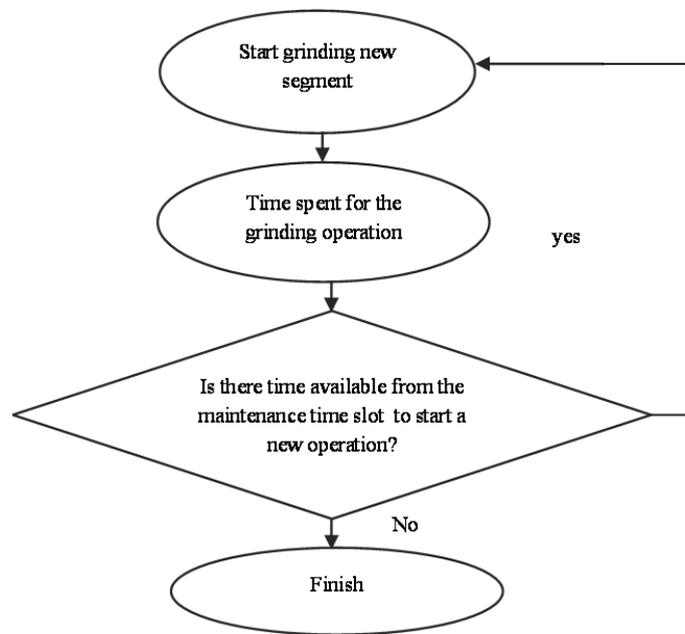
1029

1030

1031 **Figure 8**

1032

1033



1034

1035

1036

1037 **Figure 9**

1038

1039

1040



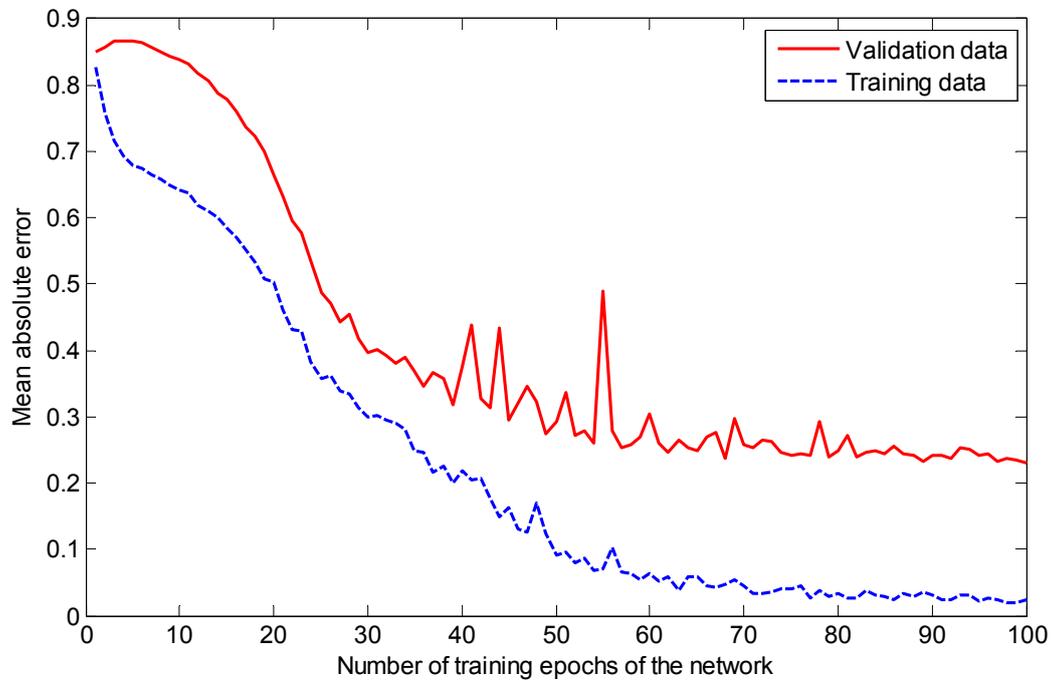
1041

1042

1043

1044  
1045  
1046

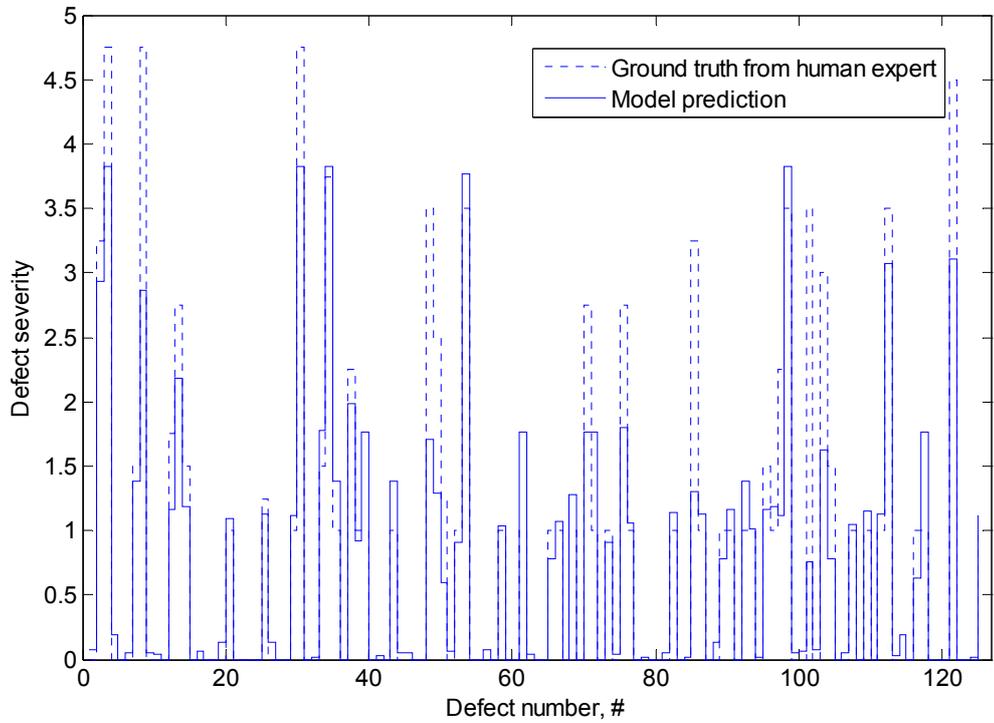
**Figure 10**



1047  
1048  
1049

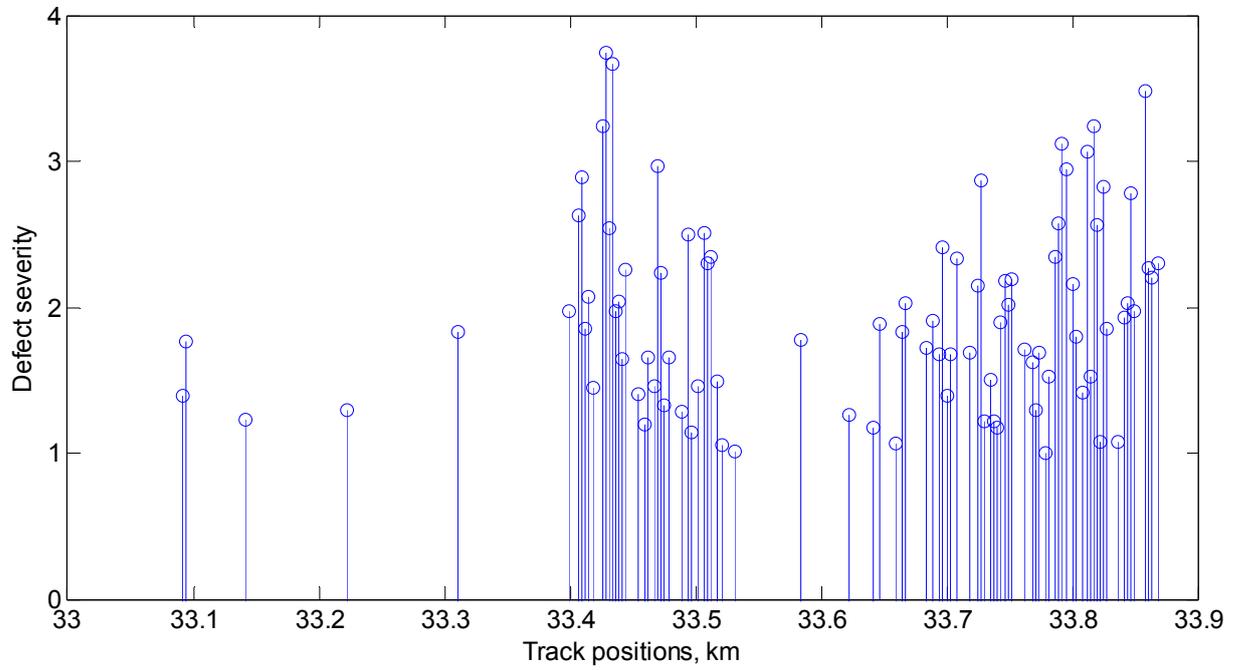
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070

**Figure 11**



1071  
1072  
1073

**Figure 12**

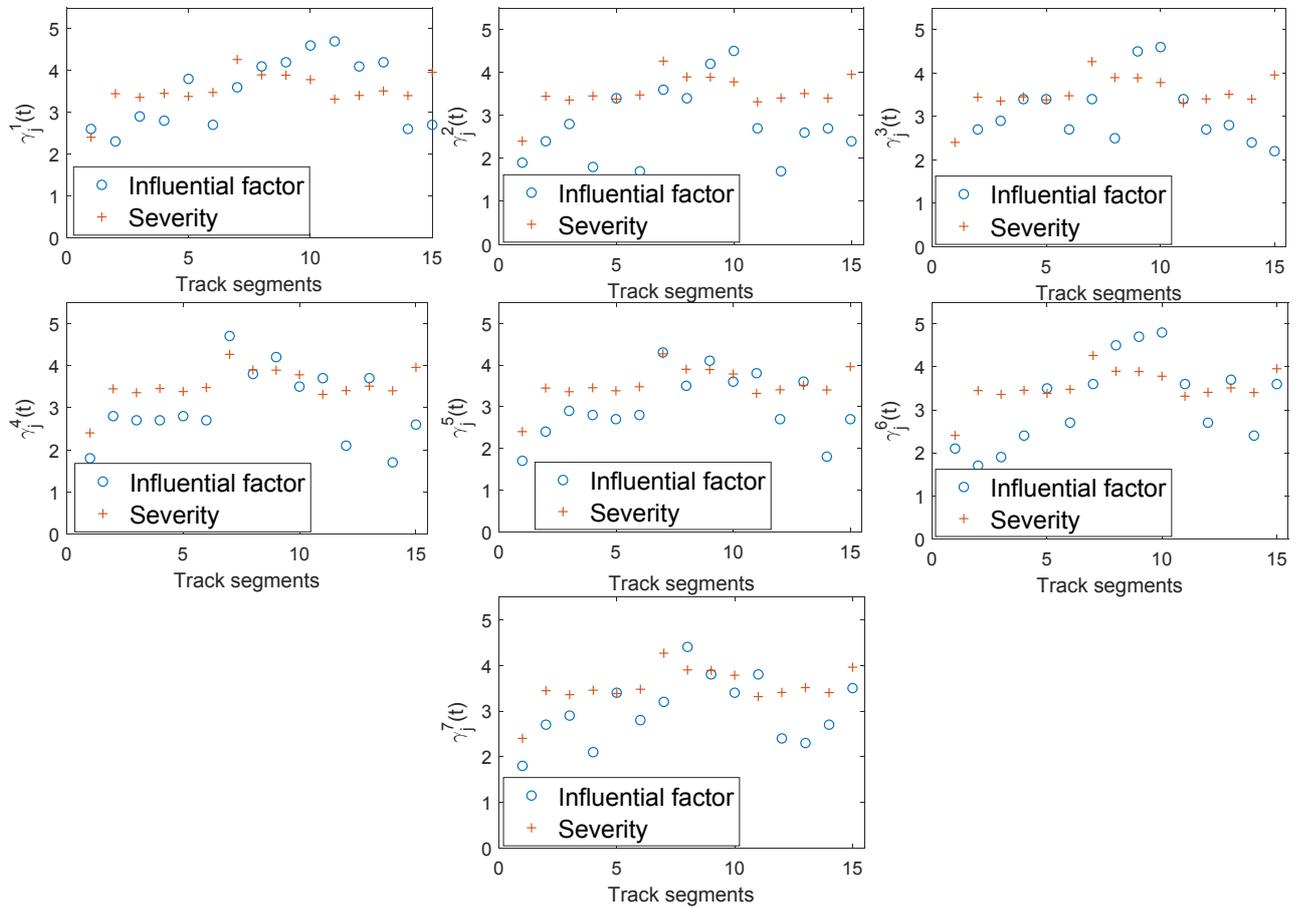


1074  
1075  
1076  
1077

1078

**Figure 13**

1079



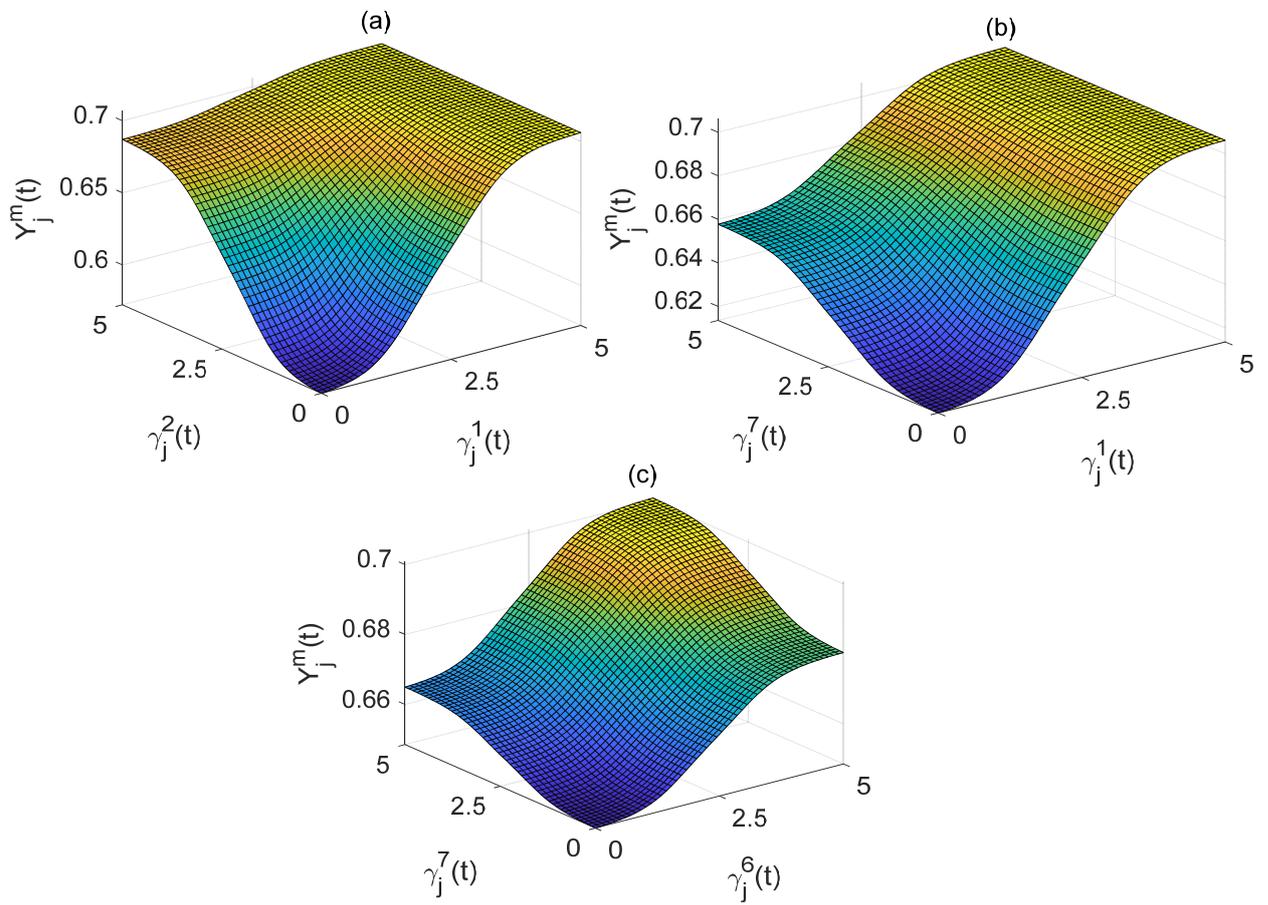
1080

1081

1082

1083  
1084  
1085

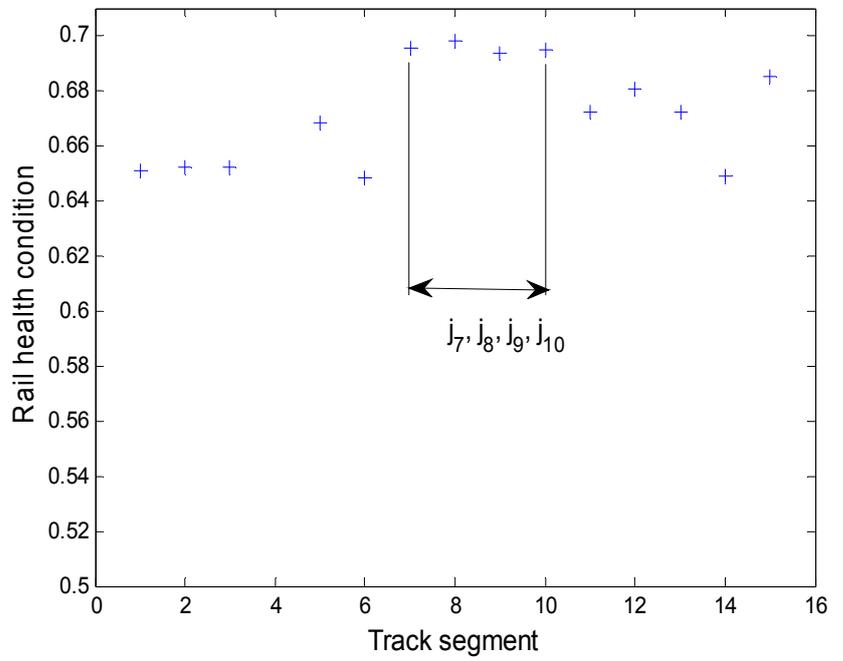
**Figure 14**



1086  
1087  
1088  
1089

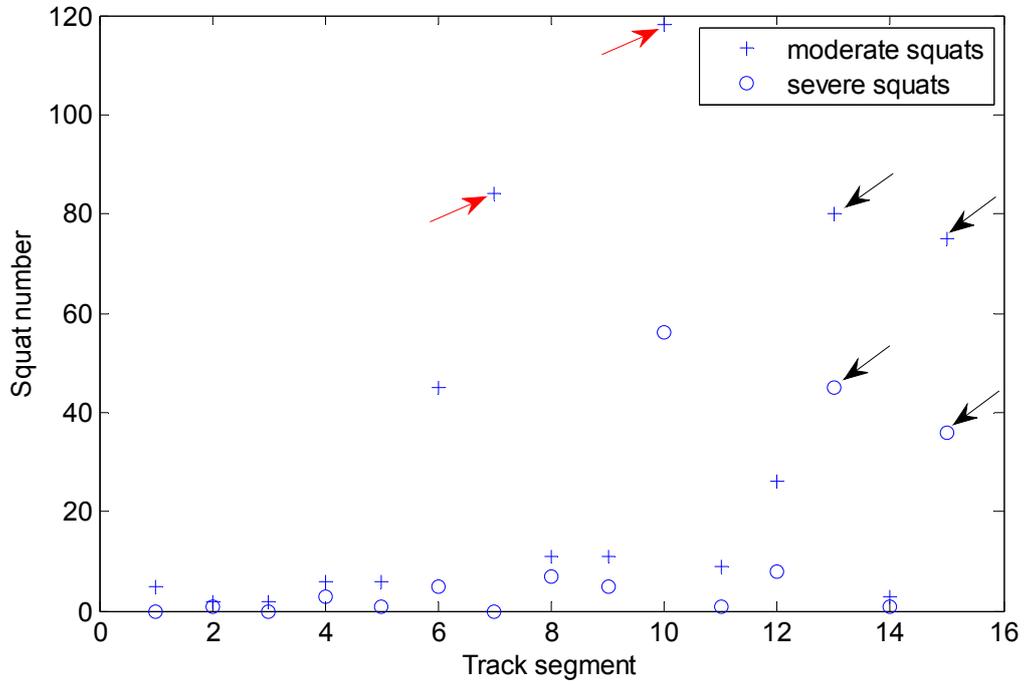
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111

**Figure 15**



1112  
1113

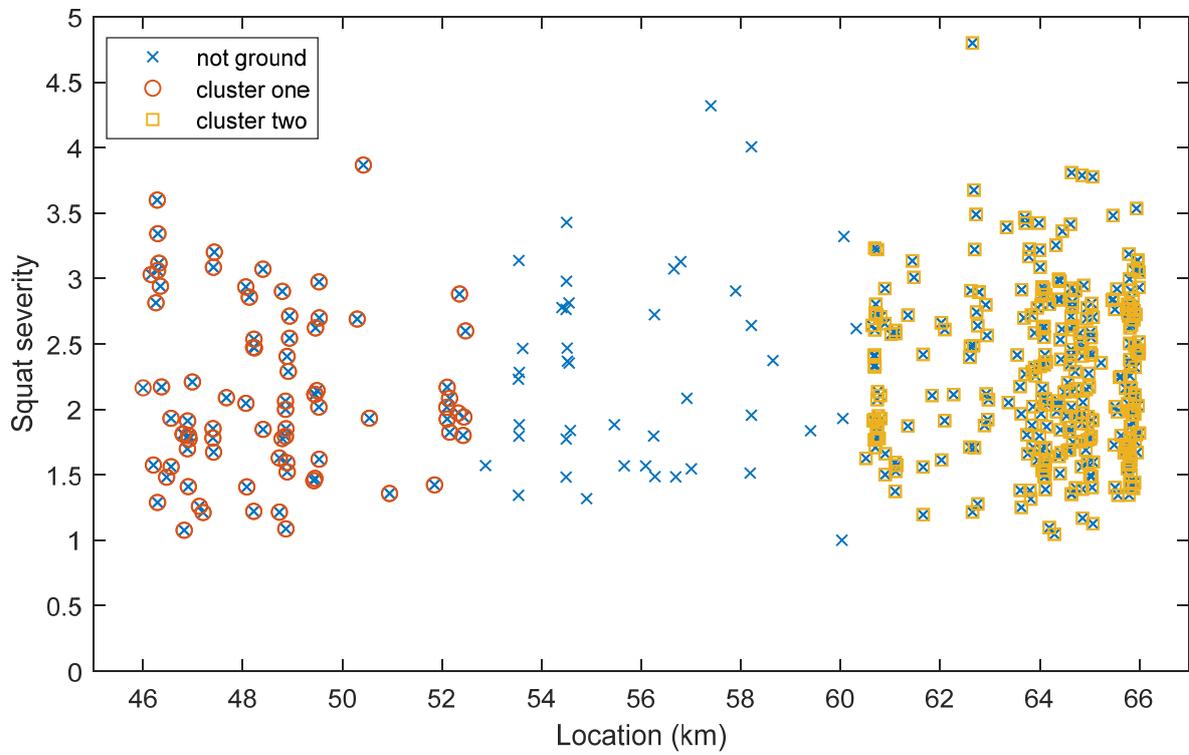
**Figure 16**



1114  
1115  
1116  
1117

1118  
1119

**Figure 17**



1120  
1121  
1122

1123  
1124  
1125

**Table 1**

<i>Notations</i>	<i>Definitions</i>
$N_c$	<i>Number of clusters</i>
$N_d$	<i>Number of squats</i>
$\underline{\xi}, \bar{\xi}$	<i>Track starting and ending positions, respectively</i>
$\omega$	<i>Defect severity</i>
$T_s$	<i>The setup time for a grinding maintenance operation.</i>
$T_t$	<i>Duration of maintenance time slot</i>
$d_g^{\text{start}}$ and $d_g^{\text{end}}$	<i>Track starting and ending positions of the g-th cluster</i>
$\Delta d_{\min}$ and $\Delta d_{\max}$	<i>Minimum and maximum size of each cluster</i>
$V_G^{\text{on}}$	<i>Grinding machine speed running over the track</i>
$V_G^{\text{off}}$	<i>Driving speed of the grinding machine when the machine is off</i>
$T_G^{\text{on}}$	<i>The time needed to switch from grinding to driving</i>
$T_G^{\text{off}}$	<i>The time needed to switch from driving to grinding</i>

1126  
1127

1128

**Table 2**

1129

Segments	$\gamma_j^1(t)$	$\gamma_j^2(t)$	$\gamma_j^3(t)$	$\gamma_j^4(t)$	$\gamma_j^5(t)$	$\gamma_j^6(t)$	$\gamma_j^7(t)$	$Y_j^m(t)$
1	2.6	1.9	1.1	1.8	1.7	2.1	1.8	0.6507
2	2.3	2.4	2.7	2.8	2.4	1.7	2.7	0.6522
3	2.9	2.8	2.9	2.7	2.9	1.9	2.9	0.6521
4	2.8	1.8	3.4	2.7	2.8	2.4	2.1	0.4929
5	3.8	3.4	3.4	2.8	2.7	3.5	3.4	0.6683
6	2.7	1.7	2.7	2.7	2.8	2.7	2.8	0.6481
7	3.6	3.6	3.4	4.7	4.3	3.6	3.2	0.6957
8	4.1	3.4	2.5	3.8	3.5	4.5	4.4	0.6982
9	4.2	4.2	4.5	4.2	4.1	4.7	3.8	0.6938
10	4.6	4.5	4.6	3.5	3.6	4.8	3.4	0.6949
11	4.7	2.7	3.4	3.7	3.8	3.6	3.8	0.6721
12	4.1	1.7	2.7	2.1	2.7	2.7	2.4	0.6803
13	4.2	2.6	2.8	3.7	3.6	3.7	2.3	0.6721
14	2.6	2.7	2.4	1.7	1.8	2.4	2.7	0.6489
15	2.7	2.4	2.2	2.6	2.7	3.6	3.5	0.6849

1130

1131

