

Cite this article

Burkhalter M and Adey BT (2022)
Required accuracy of information when determining optimal railway intervention programmes.
Infrastructure Asset Management 9(1): 18–27,
<https://doi.org/10.1680/jinam.20.00032>

Research Article

Paper 2000032
Received 25/11/2020; Accepted 12/05/2021
Published online 19/06/2021
Published with permission by the ICE under the
CC-BY 4.0 license.
(<http://creativecommons.org/licenses/by/4.0/>)

Keywords: infrastructure planning/
railway systems/transport management

Required accuracy of information when determining optimal railway intervention programmes

Marcel Burkhalter MSc Civil Eng.

Research Assistant, Institute of Construction and Infrastructure Management,
ETH Zurich, Zurich, Switzerland (Orcid:0000-0003-3017-1751)
(corresponding author: burkhalter@ibi.baug.ethz.ch)

Bryan T. Adey BEng, MSc, PhD

Professor and Head, Infrastructure Management Group, Institute of
Construction and Infrastructure Management, ETH Zurich, Zurich, Switzerland

Over the past decade, there has been a notable increase in mathematical models for determining optimal intervention programmes – that is, those with the highest net benefits. These increasingly sophisticated models rely on an increasing amount of accurate infrastructure information, which is often not available for infrastructure managers. They are, therefore, confronted with the question of which data they should collect and what accuracy these have to have in order that intervention programmes determined with sophisticated optimisation model are reliably optimal. In this paper, it is shown how infrastructure managers can estimate the required information accuracy when determining optimal railway intervention programmes. The required accuracy of the input information is defined by determining the ranges of values over which (a) the optimal intervention programme does not change, (b) the optimal intervention programmes can be considered similar and (c) the net benefit obtained by an intervention programme can be considered near optimal. The ranges are determined for one variable at a time considering the estimated values of the infrastructure manager as default values. The method is illustrated for a railway line in Switzerland using a constrained network flow model to determine the optimal intervention programmes.

Notation

I	set of interventions
IP_x	reference intervention programme determined using the value x
$IP_{x'}$	optimal intervention programme determined using the varied value x'
J	similarity coefficient
NB	net benefit
O	optimality coefficient
x	default value
x'	varied value
$\delta_{i,x}$	binary variable that is 1 if intervention i is selected with value x
$\delta_{i,x'}$	binary variable that is 1 if intervention is selected with value x'

Introduction

One of the tasks of railway infrastructure managers is the development of intervention programmes – that is, the interventions to be executed on the railway infrastructure in the next planning period and how they are to be combined (Adey, 2019). This is a challenging task for large railway networks, with a large diversity of the assets, complex relationships between assets and a large number of ways that service can be affected. In recent years, more and more sophisticated mathematical models have been developed that enable the consideration of the complex structure of railway infrastructure networks when determining optimal intervention programmes (Budai-Balke, 2009; Burkhalter and Adey, 2018; Caetano and Teixeira, 2015; Consilvio *et al.*, 2019; Dao *et al.*, 2019; Pargar *et al.*, 2017). These models rely on

an increased amount of accurate infrastructure information – for example, asset condition, deterioration, intervention costs, duration and effectiveness, traffic volume and risks of asset failure. This information is often either inaccurate or not available for infrastructure managers facing the task of determining intervention programmes.

Fellow researchers have made significant steps to improve the accuracy of input information (Artagan *et al.*, 2020), to consider the uncertainties in the input information (Andrews *et al.*, 2014; Sasidharan *et al.*, 2020) and to deal with the limited availability of input information (Zhang and Wang, 2014). The first enables infrastructure managers to collect and predict information that is more accurate. The others offer models able to consider the existing uncertainties or gaps in the available information. All of them improve single aspects of the problem that arise due to the increasing amount of required and accurate information. They, however, do not offer guidance to the infrastructure manager on how to decide which data they should collect and what accuracy these need to have to determine reliable optimal intervention programmes when using sophisticated mathematical models.

This paper shows how infrastructure managers can obtain the required accuracy of input information so that the optimal intervention programme determined by them is reliably optimal and is not significantly affected when the actual values of variables differ from the assumed ones within the defined accuracy. The method defines the required accuracy by the ranges of the values of input variables (a) over which the optimal intervention programme does not change, (b) over which the

optimal intervention programmes can still be considered similar and (*c*) over which a determined intervention programme can still be considered near optimal. The similarity coefficient is based on the Jaccard coefficient comparing the decisions of intervention programmes, while the optimality coefficient considers the net benefit reached by intervention programmes to estimate the quality of the decisions. The ranges for each input variable are defined considering the variation of the value away from its default estimated value for one variable at a time.

The method is illustrated on an example railway network in Switzerland, where the optimal intervention programme is determined using the optimisation model in the paper by Burkhalter and Adey (2018). The default values of all input variables are estimated based on the information provided by the infrastructure manager and expert knowledge. Based on these estimates, the ranges for the variables are found using 0.95 and 0.98 thresholds for the similarity and optimality criteria.

The paper is structured as follows. First, the problem situation of intervention programmes and their optimisation and input information is presented. Second, the method proposed to determine the required accuracy is described. Third, the method for determining the required accuracy is applied on a real-world railway network. Fourth, the method used is discussed. Lastly, the conclusion of the paper is cited.

Optimal intervention programmes

An intervention programme defines the interventions to be executed within a specific planning period – for example, in the next 4–12 years (Adey, 2019). When determining it, infrastructure managers have to make a set of decisions, including the intervention to be executed on specific assets; the time of the intervention execution – for example, in which year; how multiple interventions are grouped; and with which condition interventions are executed – that is, how traffic is disrupted by the interventions. In their decisions, they have to consider the different types of assets and interventions, their complex relationships, how their execution affects the service provided by the infrastructure and the available resources. Optimisation models, such as the one presented in the paper by Burkhalter and Adey (2018), enable the determination of the optimal intervention programme considering all these aspects, where the optimal intervention programme is the one with the maximal net benefit. The net benefit considers thereby the costs related to and the benefits achieved by the execution of the interventions.

Such optimisation models require a large amount of infrastructure information that needs to be estimated and defined by the infrastructure manager, which is always coupled with uncertainties, meaning that the actual value of input variable may vary from its estimated value. To determine reliable optimal intervention programmes, infrastructure managers need to know the required accuracy that their input information needs to have.

Determination of required information accuracy

Ranges of the values of input variables

The required accuracy of the values of the input variables depends on the effect of these values on the intervention programme and the net benefit obtained from the intervention programme. A variation in the value of an input variable may affect the net benefit value of an intervention programme, the decision set of the optimal intervention programme and whether an optimal intervention programme remains optimal. Figure 1 shows the effect of a variable variation from the default value x to a varied value x' . The black column represents the decision set – for example, decisions represented by variables δ_i – and the net benefit obtained by the intervention programme that is optimal with value x – for example, IP_x . The light blue column represents the decisions – for example, $\delta_{i,x'}$ – and the net benefit obtained by the intervention programme that is optimal with value x' . The two differences shown in grey indicate how a variable variation affects the net benefit value estimated – that is, the sensitivity of the objective value of the optimisation on the variable. They, however, do not provide any information regarding the reliability of the intervention programme IP_x – that is, the sensitivity of the decision and the optimality on the variable. This can be seen only when considering the decision sets and the optimality gap shown in black.

Although the required accuracy will vary from situation to situation and infrastructure manager to infrastructure manager, three ranges of the values are of particular importance – namely

- the range over which the optimal intervention programme does not change – that is, how far can the value of a single variable be changed before the optimal intervention programme changes
- the range over which the optimal intervention programme can still be considered similar, using an agreed-on value of similarity coefficients
- the range over which the net benefit obtained from the intervention programme can still be considered near optimal, using an agreed-on value of the optimality coefficient.

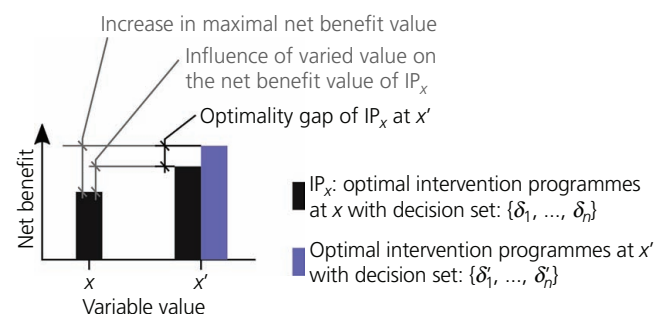


Figure 1. Effect of a variable variation on the optimal intervention programme

Similarity of intervention programmes

The similarity between two intervention programmes can be quantified using the Jaccard coefficient (Jaccard, 1912), which is the intersection over the union of two sets X and Y (Equation 1). The division by the union of the two sets means that the Jaccard coefficient neglects all elements that are not part of at least one of the two. This is favourable when the resulting value should be a function of the options selected, and not on the entire set, as done by other similarity coefficients – for example, the simple matching coefficient.

$$1. \quad J(X, Y) = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|}$$

The similarity of two intervention programmes can be estimated by dividing the number of interventions identically selected in both intervention programmes by the sum of all interventions selected in either of the two intervention programmes. This is shown in Equation 2, where I is the set of all interventions, $\delta_{i,x}$ and $\delta_{i,x'}$ are binary variables that are 1 if intervention $i \in I$ is selected in intervention programmes IP_x and $IP_{x'}$. IP_x refers to the reference intervention programme determined using the default values x , and $IP_{x'}$ refers to the optimal intervention programme determined using the varied values x' .

$$2. \quad J(IP_x, IP_{x'}) = \frac{\sum_{i=1}^I \delta_{i,x} \times \delta_{i,x'}}{\sum_{i=1}^I [1 - (1 - \delta_{i,x}) \times (1 - \delta_{i,x'})]}$$

Using such measures, the similarity of two intervention programmes is expressed by a number between 0 and 1, where 0 means that the two intervention programmes have no common interventions and 1 means that the two intervention programmes consist of the exact same interventions.

Optimality of an intervention programme

The optimality of an intervention programme can be defined as the percentage that its net benefit reaches the net benefit of the optimal intervention programme. With respect to the accuracy of input variables, this refers to the optimality reached by an intervention programme determined with the assumed default value of a variable – for example, value x – when the actual value of the variables differs – for example, value x' .

Equation 3 shows the formulation of the optimality coefficient. There, IP_x represents the reference intervention programme that is optimal with the variable value x . $IP_{x'}$ represents the intervention programme that is optimal with the variable value x' . The optimality of IP_x when the value of the variable is x' – that is, $O(IP_x, x')$ – is the division of the net benefits obtained by IP_x and $IP_{x'}$ estimated using value x' – that is, $NB(IP_x, x')$ and $NB(IP_{x'}, x')$.

$$3. \quad O(IP_x, x') = \frac{NB(IP_x, x')}{NB(IP_{x'}, x')}$$

Using such a measure, the optimality of an intervention programme given the value of a variable is expressed by a value smaller or equal to 1. A value of 1 means that the net benefit of the reference intervention programme is equal to the optimal net benefit – that is, the highest net benefit possible with the considered variable variation. A value between 0 and 1 – for example, 0.5 – means that the reference intervention programme reaches a percentage of the optimal net benefit – for example, 50%. A value of 0 means that the reference intervention programme has a net benefit of 0. Different from the similarity coefficient, the optimality can become negative with a value below 0, meaning that the net benefit of the reference intervention programme is negative – that is, the intervention programme is not worthwhile to be executed – with the considered variable variation.

Identification of ranges

The ranges of values of each input variable, which define the required accuracy, are determined by varying them one at a time (Saltelli *et al.*, 2000). For each variation, the similarity of the new optimal intervention programme is measured with respect to the reference optimal intervention programme, and the optimality of the reference intervention programme is estimated. The values of all variables are varied until the defined thresholds are reached. This is different from other similar types of analyses, such as the determination of which uncertainties to quantify (Papathanasiou and Adey, 2020), or the effect of variations in the values of input variables (Lautala and Pouryousef, 2011; Sharma *et al.*, 2018).

A global sensitivity analysis (Saltelli *et al.*, 2007), which would be able to analyse the entire parameter space, is not used because it would require the use of an output value that clearly indicates that a variation in an input variable is better or worse. This does not exist in the situation, where the variable sensitivity on the decisions and its optimality are to be evaluated and not the sensitivity of the variables on the net benefit value itself.

Example

Infrastructure network

The method for determining the required information accuracy is demonstrated on a real-world situation faced by a railway infrastructure manager in Switzerland. There, the infrastructure manager had to determine the intervention programme for a real-world railway network consisting of a 17 km long single-track line (Figure 2).

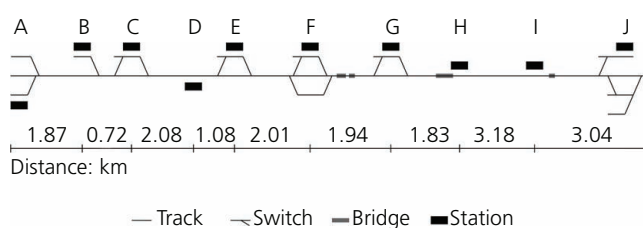


Figure 2. Railway network

Table 1. Interventions

Intervention	Applied condition	Condition following intervention	Unit	Unit cost: CHF/unit	Shared cost factor: %	Track possession: h/unit
Rail replacement	4	2	m	200	20	0.04
Track renewal	5	1	m	2350	20	0.10
Switch replacement	5	1	Assets	255 000	16	8.00
Bridge renewal	5	1	m ²	4000	0	0.50

The assets considered are 101 track segments, 23 switches and four bridges. The characteristics of these assets cover the most important aspects of assets on a railway network when developing intervention programmes. For example, the track has a longitudinal characteristic with interventions executed continuously along the track. The switches represent smaller, local assets with higher failure probabilities and shorter intervention intervals, while bridges represent larger infrastructure assets that have long lifetimes and that build the foundation for other assets.

The conditions of the assets are categorised between 1 (like new) and 5 (insufficient). The deterioration of the assets is considered with deterioration rates of 0.16 for main tracks, 0.075 for track sidings, 0.12 for switches, 0.025 for bridges between condition states 1 and 3 and 0.1 for bridges between condition states 3 and 5. The conditions are improved by the interventions considered for this analysis, which are shown in Table 1 with their estimated unit cost and unit track possession.

The user costs are estimated by using the passenger volume of 7200 passengers per day, the value of time for railway passengers of 14.75 CHF/h (1 CHF = US\$1.11) (VSS, 2009) and the additional travel time dependent on the exact closure – that is, up to 13 min. The interventions can be executed during a night break, an extended night shift, a single day shift, a weekend closure or a 24 h closure.

Table 2 shows the failure probabilities considered for each asset type and condition state. In Table 3, the probabilities of an accident in case of a failure; the probabilities of injuries and fatalities in case of an accident; and the consequences of an accident, injury and fatality are given. For corrective interventions, it is assumed that a corrective intervention costs 10% more than the corresponding preventive intervention in Table 1. Further, reaction times of 4, 2 and 24 h are assumed between the times that a failure occurs and a corrective intervention is started for track, switch and bridge assets, respectively.

Table 2. Probability of failure per condition state (CS) and asset type

Asset type	Probability of failure				
	CS 1	CS 2	CS 3	CS 4	CS 5
Track	1.0×10^{-3}	2.5×10^{-3}	6.3×10^{-3}	1.6×10^{-2}	8.0×10^{-2}
Switch	8.0×10^{-3}	1.4×10^{-2}	2.5×10^{-2}	4.5×10^{-2}	1.6×10^{-1}
Bridge	4.0×10^{-5}	1.2×10^{-4}	3.4×10^{-4}	1.0×10^{-3}	6.0×10^{-3}

All information are based either on information from the infrastructure manager or on assumptions derived from the literature.

Optimisation model and input information

The optimisation model proposed in the paper by Burkhalter and Adey (2018) is used to determine the optimal intervention programme. The optimal intervention programme is defined as the one with the highest net benefit – that is, the benefit minus the costs of the intervention programme. The benefit is defined as the reduction in costs in the future due to the improved condition of the infrastructure – that is, reduction in the risk due to a failure and reduction in the cost due to future interventions. The costs consist of all costs related to the execution of the interventions of the intervention programme. The model considers the costs on the owner – that is, the infrastructure manager – and the user – that is, the railway passenger. The reader is referred to the paper by Burkhalter and Adey (2018) for a more detailed description.

Figure 3 shows the main cost elements of the net benefit estimation and the input information required by the model as input variables. The lines represent how the variables influence the cost elements. For example, the preventive intervention costs affect the intervention costs of the intervention programme and the intervention costs of future interventions. The three variables discount rate, deterioration rate and asset condition affect all cost elements, which are not shown in the figure in favour of readability. The variables are coloured according to the type of information into intervention-related information (green), traffic-related information (blue), asset-related information (yellow), accident-related information (red) and modelling information (violet). Infrastructure information such as the network topology and the asset location and extent are not represented in the figure, as they are considered to be fixed for a given infrastructure. The focus of Figure 3 is to indicate the information that may be inaccurate or currently unavailable for infrastructure managers.

Table 3. Probabilities related to accidents

Category	Probability of an accident in case of a failure	Probability of injury per passenger in case of an accident	Probability of fatality per passenger in case of an accident
Track	0.0023	0.18	0.035
Switches	0.0002	0.18	0.035
Bridge	0.1000	0.70	0.200

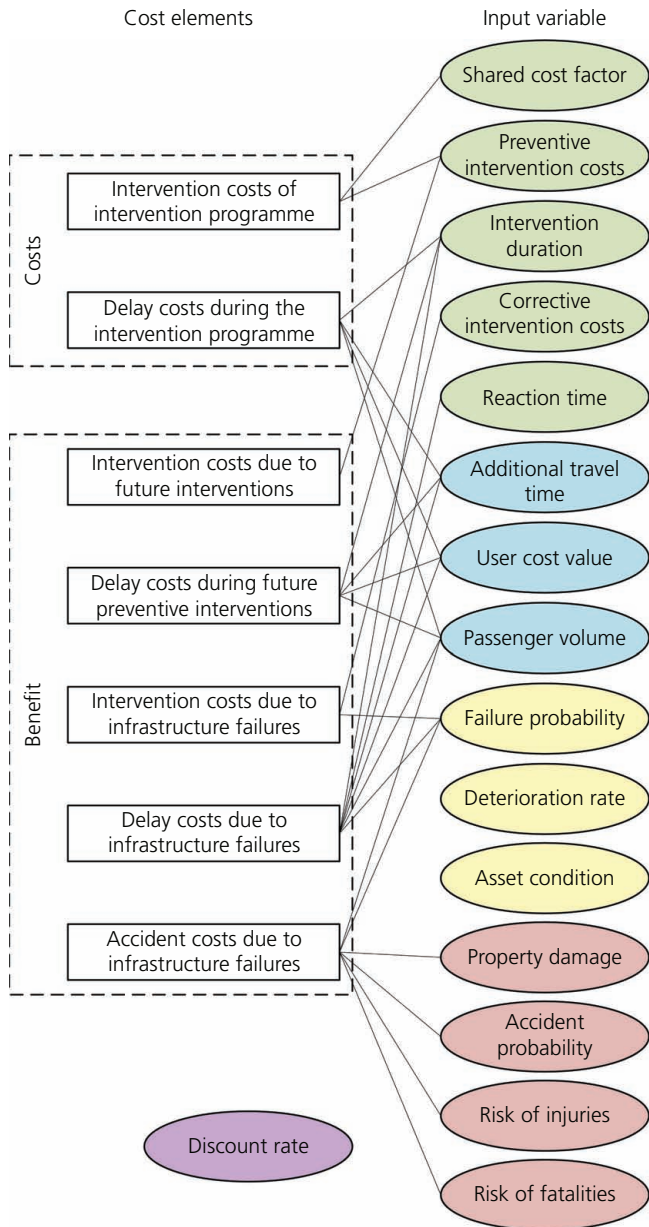


Figure 3. Input data for estimating the cost and benefit of an intervention programme

Definition of thresholds

Before the ranges can be determined that they are required for accurate input information, the thresholds for the accuracy measures have to be defined. The range over which the optimal intervention programme does not change is defined by the range over which both the similarity and optimality coefficients are equal to 1. The range over which the optimal intervention programme can still be considered similar is defined with a threshold on the similarity coefficient of 0.95 – that is, $J(IP_x, IP_{x'}) \geq 0.95$. This means that an optimal intervention programme is still considered to be similar when at least 95% of

the interventions are equal to the reference intervention programme. This means that out of 100 selected interventions, at most five interventions are either selected in addition or not selected. The range over which the reference intervention programme is still considered to be near optimal is defined with a threshold on the optimality coefficient of 0.98 – that is, $O(IP_x, value = x') \leq 0.98$. This means that the reference intervention programme is still considered optimal when it obtains a net benefit of at least 98% of the optimal net benefit with a given variation of a variable. This threshold is in the similar range as the accuracy obtained when the optimisation model is applied on large railway networks using a genetic algorithm (Burkhalter and Adey, 2019).

Ranges of values of input variables

Figures 4–7 show the determined ranges for the input variables over which they are within the defined threshold for the unchanged, similarity and optimality ranges. The figures show the ranges in percentage variations (x -axis) of the variable from the default value, which is highlighted at variation 0. This is favourable particularly for the variables that take different values for different categories. For example, the intervention costs are different for each intervention on each type of object. The use of percentage variation enables showing the ranges for the intervention costs of all interventions together. In some graphs, the entire optimality range is not shown in favour of readability. There, the range is indicated by the value at which the threshold is reached. The four figures each show a group of variables with similar ranges.

Figure 4 shows the ranges for the condition of the assets, the deterioration rate and the shared cost factor. The range over which the optimal intervention programme is not changed is for all three

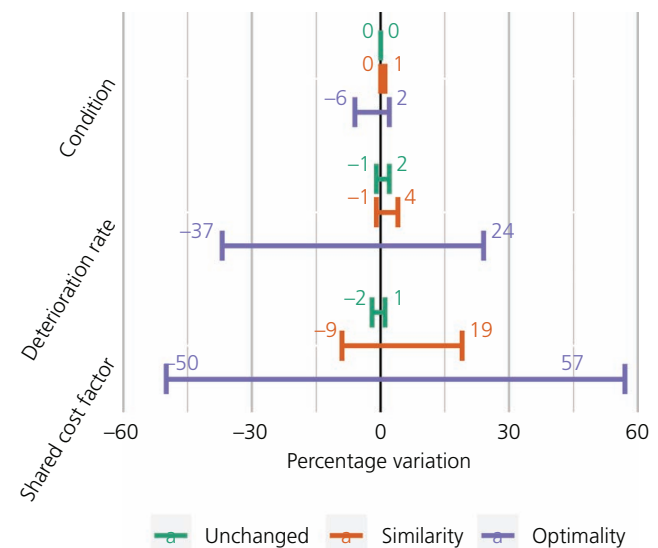


Figure 4. Range of variables of group 1

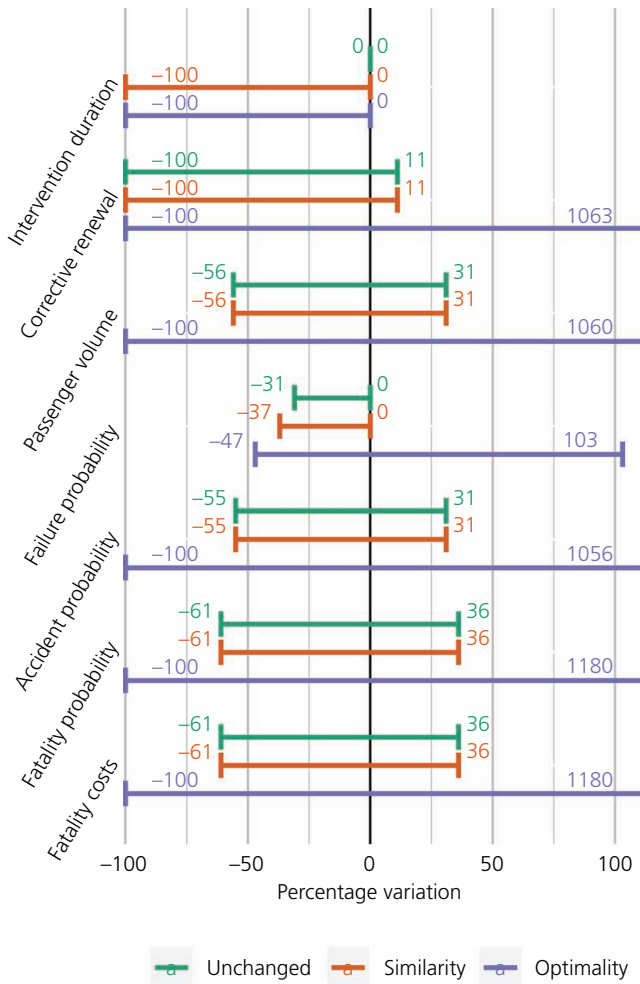


Figure 5. Range of variables of group 2

variables within $\pm 2\%$, where any change in the condition of the assets leads to a change in the optimal intervention programme. The percentage variation for the condition of the asset refers thereby to the percentage of assets with overestimated (negative variation) or underestimated (positive variation) condition states. For example, a variation of 2% refers to the situation where the condition states of 2% of the assets are underestimated. The ranges over which the optimal intervention programme are considered to be similar are 0 to +1% for the condition of the asset, -1 to +4% for the deterioration rate and -9 to +19% for the shared cost factor. The range over which the reference intervention programme is near-optimal, are higher with -6 to +2% for the condition of the assets, -37 to +24% for the deterioration rate and -50 to +57% for the shared cost factor. This means that the reference intervention programme determined with the assumed value of the shared cost factor (see Table 1) obtains a near-optimal net benefit when the actual value of the shared cost factor is 30% higher, even though the optimal intervention programme differs significantly from the reference intervention programme – that is, similarity < 0.95.

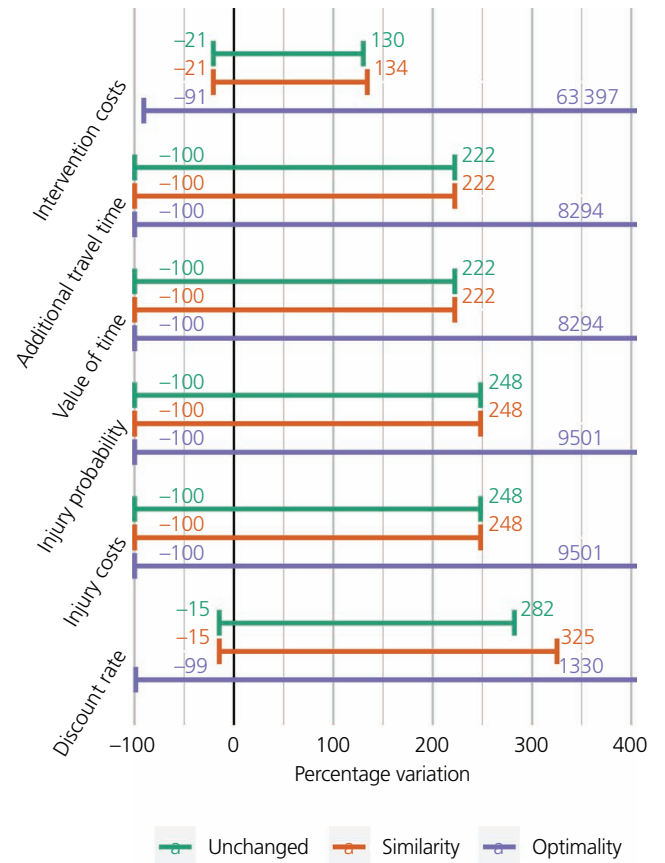


Figure 6. Range of variables of group 3

Figure 5 contains the variables for which the ranges over which the optimal intervention programme are unchanged and similar are within $\pm 100\%$. This includes the intervention duration, the costs for corrective interventions, the passenger volume, the failure probability, the accident probability, the fatality probability and the fatality costs. Regarding the intervention duration, any variation in

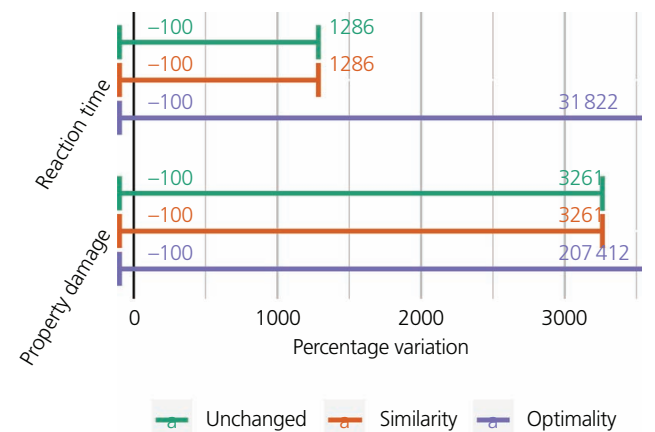


Figure 7. Range of variables of group 4

its value leads to changes in the optimal intervention programme. Over the range of -100 to 0%, the optimal intervention programme remains similar and the reference intervention programme near optimal. A variation of -100% refers thereby to a value of 0 for the variable – for example, an intervention duration equal to zero. The upper limits of the ranges for the intervention duration are at 0 because the reference intervention programme is impracticable with intervention durations that are longer than the assumed values. The unchanged ranges and similarity ranges for the costs of corrective interventions and passenger volume are -100 to +11 and -56 to +31%. Their optimality ranges are significantly higher with upper limits above +1000%, meaning that the reference intervention programme is near optimal even when the actual value of the variable is ten times higher than the assumed value. The failure, accident and fatality probabilities show increasing ranges, which correspond to their reduced influence in the risk quantification of the net benefit estimation.

Figure 6 contains the variables for which the optimal intervention programmes are unchanged within the range of -100 and +300%. The ranges related to the intervention costs are thereby the smallest ranges with approximately -21 to +130%. In contrast, the upper bound of the range over which the reference intervention programme can be considered near optimal is over +63 000%, which is much higher than for the other variables.

Variations in the additional travel time, value of time, injury probability and injury costs do not lead to changes in the optimal intervention programme over the range of -100 to +222 and -100

to +250%. The ranges for the discount rate show that rather small negative variations of the discount rate lead to significant changes in the optimal intervention programme – that is, lower bound at -15%, while the optimal intervention programme does not change over a larger positive variation – that is, upper bound at +282%.

Figure 7 shows the ranges for the reaction time and property damage. For both these variables, the optimal intervention programme does not change over the large range of -100 to +1286 and +3261%, respectively. The ranges over which the reference intervention programme is near optimal are even larger with reaching the 0.98 optimality threshold when the values are 318 and 2074 times higher than the values assumed.

Required accuracy of input variables

Table 4 summarises the ranges determined in the previous section. The ranges provide the required accuracy of the input variables for determining a reliable optimal intervention programme for the railway network considered in the example. The first range, the range over which the optimal intervention programme is unchanged, defines the required accuracy in order that the intervention programme determined by the infrastructure manager is truly optimal. The second range, the range over which the optimal intervention programme can be seen as similar, defines the required accuracy in order that the actual decision made by the infrastructure manager – that is, which interventions to execute – has at most marginal differences with the optimal intervention programme. The third range, the range over which the reference intervention programme is near optimal, defines the

Table 4. Required accuracy of input variables

Group	Input variable	Range over which the optimal intervention programme does not change: %		Range over which the optimal intervention programmes are similar: %		Range over which the reference intervention programme is near optimal: %	
		From	To	From	To	From	To
A	Condition			0	+1	-6	+2
	Deterioration rate			-1	+4	-37	+24
	Shared cost factor	-2	+2	-9	+19	-50	+57
	Intervention duration			-100	0	-100	0
B	Probability of failure	-31	0	-37	0	-47	+103
	Corrective intervention	-100	+11	-100	+11	-100	+1163
	Intervention costs	-21	+130	-21	+134	-91	+63 397
	Discount rate	-15	+282	-15	+325	-100	+1330
C	Passenger volume						
	Accident probability						
	Fatality probability						
D	Fatality cost	-60	+35	-60	+35	-100	+1100
	Additional travel time						
	Value of time						
	Injury probability						
E	Injury cost	-100	+250	-100	+250	-100	+9000
	Reaction time						
E	Property damage						
		-100	+1286	-100	+1286	-100	+31 822
		+3261		+3261		+207 412	

required accuracy in order that the intervention programme determined by the infrastructure manager obtains an at least near-optimal net benefit. Which required accuracy to choose depends on the current situation of information availability, the available resources for improving the accuracy of the input information and the aim of the infrastructure managers. The required accuracies based on the first range enable the infrastructure manager to determine the optimal intervention programme, but these require more effort and time to collect and estimate the input variables. The minimal required accuracies based on the second and third ranges enable infrastructure managers to determine reliable near-optimal intervention programme with less strict conditions on the required accuracy. This is of special interest for infrastructure managers when determining the intervention programme for large railway networks using heuristic optimisation algorithms, which enable finding a near-optimal intervention programme with less computational effort (Burkhalter and Adey, 2019).

Table 4 groups the variables into five groups – that is, A to E – according to their required accuracies when determining optimal intervention programmes for the railway network considered.

- Group A contains the condition of the assets, the deterioration rate, shared cost factor and the intervention duration. These variables need to be available with high accuracy, less than $\pm 2\%$, in order that the optimal intervention programme does not change. They also need to be available with required high accuracy in order that the optimal intervention programmes are similar and the reference intervention programme is near optimal. There, the required accuracy differs for each variable with all having ranges of up to 100%.
- Group B contains the failure probability, the costs for corrective and preventive interventions and the discount rate. The required accuracies of these variables differ from variable to variable, but all have either an upper or a lower limit of the range over which the optimal intervention does not change that requires the variable to be available with high accuracy. The upper limits of the required accuracy of the probability of failure and the costs for corrective interventions are at 0 and +11%. The lower limits of the required accuracy for the probability of failure, the costs for preventive interventions and the fatality costs are at -31, -21 and -15%.
- Group C contains the passenger volume, the accident probability and the fatality probability and cost. These variables need to be available with an accuracy of -60 and +35% in order to determine the optimal intervention programme.
- Group D contains the additional travel time, the value of time and the injury probability and costs. These variables need to be available with an accuracy of -100 and +250% in order to determine the optimal intervention programme.
- Group E contains the reaction time and the property damage. Both variables do not need to be available with high accuracy, as all values of the variables between 0 and at least 12 times higher than the assumed value do not change the optimal intervention programme.

All variables in groups B to E, except the probability of failure, enable determining a near-optimal intervention programme without having the variables available with high accuracy in the considered situation. All their ranges over which the reference intervention programme is near optimal are at least -91 to +1100%.

Discussion

A method for estimating the required accuracy of input information when determining optimal intervention programmes is proposed in this paper. The application of this method shows how the required accuracy of the input variables can be estimated to determine reliable optimal intervention programmes for a railway line. The results determine a high required accuracy for the condition of the assets, the deterioration rate and the shared cost factor referring to the intervention costs that can be reduced when multiple interventions are combined. The required accuracy of these variables is high for all ranges considered – that is, the range over which the optimal intervention programme does not change, the range over which it remains similar and the range over which the determined intervention programme is near optimal. Other variables such as the intervention costs, the costs for corrective interventions and the discount rate require information available with high accuracy in order that the optimal intervention programme does not change, but less high accuracy in order that the determined intervention programme is near optimal. This shows clearly that the required accuracy of input variables depends on the target optimality of the intervention programme set by the infrastructure manager. The estimation of the required accuracy requires a trade-off between the optimality and similarity obtained by a determined intervention programme and the effort required to estimate the input information with the required accuracy. An infrastructure manager may be satisfied with an intervention programme that is similar to the optimal intervention programme and whose net benefit obtained is near optimal, while another infrastructure manager and/or in another situation would want to spend the additional effort to estimate the higher required accuracy in order that the determined intervention programme is truly the optimal one.

The results of the example show the required accuracy of the input information to develop a reliable optimal intervention programme for the example network. They neglect thereby how accurate this information currently is in practice. For example, the optimal intervention programme is estimated to not change when the actual passenger volume lies within -60 and +35% of the assumed value. The available measuring techniques enable the estimation of the passenger volume with a higher accuracy. The information about the actual cost ratio that can be reduced by combining interventions is not as well known and underlays uncertainties that are higher than the required accuracy of $\pm 2\%$. So that infrastructure managers can decide where best to spend their time and money to improve the accuracy of their information, they need to compare the required accuracy determined with the available accuracy of the information.

Further, the results show a relatively low required accuracy for the additional travel time and the value of time, which are both related to user costs. This may be due to the considered situation with a railway line with a rather low traffic volume. In a situation with a high traffic volume – for example, an urban main line – the required accuracy of the additional travel time and value of time may be higher. It has, therefore, to be considered that the required accuracies should be determined for each specific situation – that is, they are difficult to generalise.

Conclusion

This paper proposes a method for estimating the required accuracy of input information when determining optimal intervention programmes for railway networks using a mathematical optimisation model. The method enables the determination of the required accuracy of the input information in order to ensure that the determined intervention programme is reliably optimal. The required accuracy is defined by the range of values of the input variables over which (a) the optimal intervention programme does not change, (b) the optimal intervention programme can be considered similar (c) and the net benefit obtained by an intervention programme can be considered near optimal.

The similarity of two intervention programmes is quantified using a similarity coefficient based on the Jaccard coefficient, which enables the comparison of the compositions of the two intervention programmes. The optimality of an optimal intervention programme is quantified by the division of the net benefit obtained by the reference intervention programme – that is, the one optimal with the default value of the variable – considering a variation in the input variables and the net benefit of the optimal intervention programme with this variation in the input variables.

The method was demonstrated using a single-track railway line. The results show a high required accuracy for the condition of the assets, the deterioration rate, the factor of shared costs between interventions and the intervention duration. Already the slightest variations of the condition of the assets – that is, a wrong determined condition of an asset – and the intervention duration lead to changes in the optimal intervention programme, with required accuracy equal to $\pm 0\%$. The shared cost factor and the deterioration rate have required accuracies of $\pm 2\%$. Further, the failure probability, the intervention costs, the costs for corrective interventions and discount rate require accuracies that are within $+30$ to -30% . A slightly lower accuracy – that is, -60 to $+35\%$ – requires the passenger volume, the accident probability, the fatality probability and fatality costs.

The method proposed in this paper is useful to infrastructure managers to decide where best to spend their effort in order to improve the accuracy of the information. It enables estimating the required accuracy of information, which can be compared with the accuracy of the available information in order to determine the information to estimate or determine with higher accuracy.

Acknowledgement

The work presented here has received funding from Horizon 2020, the EU's Framework Programme for Research and Innovation, under grant agreement number 769373.

REFERENCES

- Adey BT (2019) A road infrastructure asset management process: gains in efficiency and effectiveness. *Infrastructure Asset Management* **6(1)**: 2–14, <https://doi.org/10.1680/jinam.17.00018>.
- Andrews JD, Prescott D and De Rozières F (2014) A stochastic model for railway track asset management. *Reliability Engineering and System Safety* **130**: 76–84, <https://doi.org/10.1016/j.res.2014.04.021>.
- Artagan SS, Bianchini Ciampoli L, D'Amico F, Calvi A and Tosti F (2020) Non-destructive assessment and health monitoring of railway infrastructures. *Surveys in Geophysics* **41**: 447–483, <https://doi.org/10.1007/s10712-019-09544-w>.
- Budai-Balke G (2009) *Operations Research Models for Scheduling Railway Infrastructure Maintenance*. PhD thesis, Erasmus University, Rotterdam, the Netherlands.
- Burkhalter M and Adey BT (2018) A network flow model approach to determining optimal intervention programs for railway infrastructure networks. *Infrastructures* **3(3)**: article 31, <https://doi.org/10.3390/infrastructures3030031>.
- Burkhalter M and Adey BT (2019) Determining optimal intervention programs for large railway infrastructure networks using a genetic algorithm. *Proceedings of the 12th World Congress on Railway Research, Tokyo, Japan*, pp. 1–6.
- Caetano LF and Teixeira PF (2015) Optimisation model to schedule railway track renewal operations: a life-cycle cost approach. *Structure and Infrastructure Engineering* **11(11)**: 1524–1536.
- Consilvio A, Sanetti P, Anguita D et al. (2019) Prescriptive maintenance of railway infrastructure: from data analytics to decision support. *Proceedings of the 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Cracow, Poland*, pp. 1–10.
- Dao CD, Hartmann A, Lamper A and Herbert P (2019) Scheduling infrastructure renewal for railway networks. *Journal of Infrastructure Systems* **25(4)**: article 04019027, [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000515](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000515).
- Jaccard P (1912) The distribution of the flora in the alpine zone. *New Phytologist* **11(2)**: 37–50, <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>.
- Lautala P and Pouryousef H (2011) Sensitivity analysis of track maintenance strategies for the High Speed Rail (HSR) services. In *Proceedings of the 2011 Joint Rail Conference*. American Society of Mechanical Engineers, New York, NY, USA, pp. 141–150.
- Papathanasiou N and Adey BT (2020) Identifying the input uncertainties to quantify when prioritizing railway assets for risk-reducing interventions. *CivilEng* **1(2)**: 106–131, <https://doi.org/10.3390/civileng1020008>.
- Pargar F, Kauppila O and Kujala J (2017) Integrated scheduling of preventive maintenance and renewal projects for multi-unit systems with grouping and balancing. *Computers & Industrial Engineering* **110**: 43–58, <https://doi.org/10.1016/j.cie.2017.05.024>.
- Saltelli A, Tarantola S and Campolongo F (2000) Sensitivity analysis as an ingredient of modeling. *Statistical Science* **15(4)**: 377–395, <https://doi.org/10.1214/ss/1009213004>.
- Saltelli A, Ratto M, Andres T et al. (2007) *Global Sensitivity Analysis: The Primer*. Wiley, Chichester, UK.
- Sasidharan M, Burrow MPN and Ghataora GS (2020) A whole life cycle approach under uncertainty for economically justifiable ballasted railway track maintenance. *Research in Transportation Economics* **80**: article 100815, <https://doi.org/10.1016/j.retrec.2020.100815>.
- Sharma S, Cui Y, He Q, Mohammadi R and Li Z (2018) Data-driven optimization of railway maintenance for track geometry.

Transportation Research Part C: Research in Transportation Economics **90**: 34–58, <https://doi.org/10.1016/j.trc.2018.02.019>.
VSS (Schweizerischer Verband der Strassen- und Verkehrsfachleute) (2009) SN 641 822a: Cost benefit analysis for road traffic: travel time costs for passenger traffic. VSS, Zurich, Switzerland.

Zhang W and Wang W (2014) Cost modelling in maintenance strategy optimisation for infrastructure assets with limited data. *Research in Transportation Economics* **130**: 33–41, <https://doi.org/10.1016/j.res.2014.04.025>.

How can you contribute?

To discuss this paper, please submit up to 500 words to the editor at journals@ice.org.uk. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial board, it will be published as a discussion in a future issue of the journal.