Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1381

Published online 7 March 2012 in Wiley Online Library

Optimal Scheduling of Track Maintenance on a Railway Network

Tao Zhang,[a,b*,†] John Andrews[b] and Rui Wang[a,c]

To ensure the safety and continued operation of the railway network system, many maintenance and renewal activities are performed on the track every month. Unplanned maintenance activities are expensive and would cause low service quality. Therefore, the track condition should be monitored, and when it has degraded beyond some acceptable limit, it should be scheduled for maintenance before failure. An optimal timetable of the maintenance activities is needed to be scheduled, planning the monthly workload, to reduce the effect on the transportation service and to reduce the potential costs. Considering the uncertainties of the deterioration process, the safety of transportation service, the lifetime loss of the replaced track, the maintenance cost and the travel cost, this article advances an optimisation model for the maintenance scheduling of a regional railway network. An enhanced genetic algorithm approach is proposed to search for a solution producing maintenance schedule such that the overall cost is minimised in a finite planning horizon. A case study is given to demonstrate the application of the method. The case study results were derived by using an enhanced genetic algorithm method, which is specifically developed to deal with the characteristics of the railway maintenance problem. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: maintenance scheduling; railway network; track; genetic algorithm; optimisation

1. Introduction

In the UK, Network Rail now manages the maintenance of the entire railway network. Regions are the responsibility of different depots to maintain. The track in a region can be considered to be defined in terms of track sections, where a section is defined as a length of track between nodes where the traffic, ballast and sleeper types are constant. The track section is then broken down into small segments for data monitoring and reporting purposes. A segment can be considered as a basic unit for maintenance. In the UK, the typical length of a segment is 1/8 of a mile (200 m),[1] which is different from the USA where it is 0.1 mile (160 m).[2] The ride quality, as determinate by track geometry measurements, for segments of a railway network will not fail suddenly but will gradually deteriorate from a good condition to an unacceptable condition. The current condition (track geometry) of each segment indicates whether we need to do maintenance (such as tamping or stone blowing) and the priority with which this work must be carried out to avoid unacceptable conditions when speed restrictions or line closure is imposed until the situation is rectified. For example, over time the constant movement of traffic over the track causes gaps to form in the ballast structure, known as *voids*. Often, from the trackside, sleepers can be clearly seen bouncing up and down as the wheels pass over them. Excessive movement is dangerous, of course, so the voids need to be filled to give a firm base for each sleeper. This work is called *tamping*, which has been performed in the past directly by manual labour but today is performed by the tamping machine. The tamping machine works by vibrating the ballast and forcing it under the sleeper. These combined actions cause the ballast to form a close matrix that can support the track effectively. Such maintenance can correct the alignment of the rails to make them parallel and level, to achieve a more comfortable and safer ride for passengers and freight and to reduce the mechanical strain applied to the rails by passing trains. The condition of track segment can be detected by existing methods and devices such as the special measuring trains. Usually, a measurement train passes along the network about once per month in the UK.

The condition-based maintenance scheduling of a railway network is very important. Unplanned maintenance activities, resulting from the occurrence of unacceptable or unsafe failures, would have a great effect on the service of passengers and goods transportation. The resulting possession time during which the repair was carried out would interrupt and delay the normal transportation. In 2002, the train extra delay time caused by the infrastructure maintenance in Sweden is 20% of all extra delay time.[3] There are

[a]College of Information Systems and Management, National University of Defense Technology, Changsha 410073, China
[b]Nottingham Transportation Engineering Centre, University of Nottingham, Nottingham NG7 2RD, UK
[c]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK
*Correspondence to: Tao Zhang, College of Information Systems and Management, National University of Defense Technology, Changsha 410073, China.
†E-mail: zhangtao@nudt.edu.cn

commonly several different lines (up/down, fast/slow) in a track segment, and many track maintenance activities, such as renewal, may cause the entire segment to be unusable because of the size of the maintenance machines and the need to guarantee the safety of the workforce because of the close proximity between adjacent lines. To keep the serviceability of the entire railway network, large sums of money are spent on the rail track maintenance and renewal. In the Netherlands, there was approximately 295 million and 150 million euro for maintenance and renewal, respectively, in 2003.[4] In North America, billions of dollars on rail track maintenance are spent every year.[5] In the UK, track maintenance and railway station modernisation was projected to cost more than 22.7 billion pounds for 5 years from 2006.[6] Hence, it is critical to plan the coming maintenance activities in advance.

Because the railway network and its components have a range of performance levels—from perfect functioning to complete failure, it is a typical deteriorating and multistate system. Levitin *et al*.[7] generalised a replacement schedule optimisation problem to multistate systems. The multistate system reliability is defined as the ability to satisfy a demand that is represented as a required system performance level. The reliability of system elements is characterised by their lifetime distributions with hazard rates increasing in time. The optimal number of element replacements during the study period is defined as that which provides the desired level of the system reliability by the minimum sum of maintenance cost and cost of unsupplied demand caused by failures. To evaluate multistate system reliability, a universal generating function technique is applied. A genetic algorithm (GA) is used as an optimisation technique. Zhang *et al*.[8] presented a new reliability-based optimal maintenance scheduling method that considers the effect of maintenance in reducing costs. An ordering list of element maintenance effects with various maintenance-interval types is constructed. Using this ordering list, reliability-based optimal maintenance scheduling for simple reliability structures and composite reliability systems is then carried out. Mosheiov *et al*.[9] studied a single machine scheduling problem. The processor needs to go through a maintenance activity, which has to be completed before a given deadline. The objective function is the minimum total weighted completion time. An efficient heuristic, which is shown numerically to perform well, is presented. Wang *et al*.[10] considered identical parallel machine scheduling problems with a deteriorating maintenance activity. In this model, each machine has a deteriorating maintenance activity, that is, delaying the maintenance increases the time required to perform it. Two goals were separately considered: minimising the total absolute differences in completion times and minimising the total absolute differences in waiting times. Cheng *et al*.[11] studied the problem of unrelated parallel-machine scheduling with deteriorating maintenance activities. Each machine has at most one maintenance activity, which can be performed at any time throughout the planning horizon. The length of the maintenance activity increases linearly with its starting time. The objective is to minimise the total completion time or the total machine load. However, the previously mentioned advanced models are not suitable for a railway network system.

To date, many studies have been carried out to investigate the problem of maintenance scheduling for railway systems. The systems considered include a single track system,[12] a track system with several segments,[4] a regional railway system[13–15] and a large-scale railway network.[3] From these studies, it is evident that there are some effective ways to improve the service quality and to reduce the maintenance costs. As discussed earlier, it is possible to do maintenance activities before the components deteriorate into the unacceptable condition. Also, maintenance activities can be carried out in the interval between train operations (e.g. at night), which will cut down on the effect on the transportation service. For example, Higgins *et al*.[12] developed a model designed to help resolve the conflicts between train operations and the schedule of maintenance activities by using a tabu heuristic search. Opportunistic maintenance is possible where some maintenance activities may be carried out concurrently to reduce the possession time. For example, we can combine the maintenance activities of different components in the same track segment or the maintenance activities of two adjacent track segments. Initially, Budai *et al*.[13] started to study the preventive maintenance scheduling problem for railway infrastructure. The aim was to obtain a timetable for preventive maintenance activities such that the possession and maintenance costs are minimised. The possession costs are mainly determined by the possession time. In a later article,[14] an improved approach using a GA, a memetic algorithm and a two-phase heuristic based on opportunities was given. Zhao *et al*.[4] described a procedure that may be used to help synchronise the renewal of railway track components so that the costs of the renewal processes are minimised. Three types of combinations were considered to reduce the costs. A GA-based approach was used to find the solution. Another means of carrying out effective maintenance is to assign the resources to reduce the travel costs of the maintenance teams and machines. Peng *et al*.[3] presented a time–space network model to solve the track maintenance scheduling problem. The objective is to minimise the total travel costs of the maintenance teams as well as the effect of maintenance projects on railroad operation, which are formulated by three types of side constraint: mutually exclusive, time window and precedence constraints. An iterative heuristic solution approach is proposed to solve this problem.

In the studies mentioned earlier, the duration, the earliest and the latest possible starting times of each project are known. However, in fact, each segment has an uncertain deterioration process because of various factors such as environment and usage. Lapa *et al*.[16] presented a methodology for preventive maintenance planning of a pressurised water reactor on the basis of a cost-reliability model, which takes into account both the probability of needing corrective maintenance represented by a Weibull distribution and the probability of doing bad maintenance. A GA-based approach was used to find the best solution.

Taking into account the deterioration processes, the safety of transportation service, the cost of losing useful life, the maintenance cost and the travel cost, an overall cost model is advanced to evaluate a given solution in this study. The enhanced GA-based approach is proposed to obtain a good solution. The rest of the article is organised as follows. Section 2 gives the description of this problem and model formulation. Section 3 describes the steps of GA, which is used for searching a satisfying solution. Section 4 presents a case study of a regional railway network with more than 30 segments awaiting maintenance. Section 5 draws conclusions from the research.

2. Model formulation

The measurement train passes along the network approximately once per period (normally 1 month in the UK). After that, a schedule of the maintenance activities for the next month is planned.

Let $|\kappa|$ be the cardinality of set κ. Let $G=(V,A)$ be a regional railway network, where $V=\{v_1,v_2,\ldots,v_{|V|},\}$ represents the set of junctions or stations (nodes) and $A=\{a_1,a_2,\ldots,a_{|A|},\}$ is the set of sections between two nodes. Here, ?v_i and a_i represents the ith node and the ith section, respectively. The track section is then broken down into small segments whose length is 1/8 of a mile in the UK. A segment can be regarded as a component of the track system. Let $S=\{s_1,s_2,\ldots,s_{|S|}\}$ be the set of all segments in the network. We define $Sec(s_i)$ as the section that segment s_i locates in $Sec(\cdot)\in A$.

Because the locations of sections in the network are different, their corresponding importance to the network is also different. They will experience different traffic conditions as defined by the different sizes of trains and passenger loadings. There are also freight trains that are not on the timetable. In this study, considering the factors mentioned earlier, the average number of trains passing on a section per day is used to indicate a section of importance. Let I_i be the importance of the ith section,

$$I_i=\sum_{k=1}^{K}\alpha_k N_{i,k}/Max\left(\sum_{k=1}^{K}\alpha_k N_{1,k},\sum_{k=1}^{K}\alpha_k N_{2,k},\ldots,\sum_{k=1}^{K}\alpha_k N_{|A|,k}\right) \tag{1}$$

where α_k is the weight of the kth train type, $N_{i,k}$ is the average number of trains of kth type passing by the ith section per day and K is the total number of train types. Hence, the importance values will be in the range $0<I_i\le 1$.

This study takes into account the deterioration process, which can be divided into three phases: a good condition, a degraded condition awaiting maintenance and an unacceptable condition, as shown in Figure 1. Let d_T *and* d_U be the maintenance trigger point value and the unacceptable point value of the condition, respectively. The segment would be functional in the good and the degraded awaiting maintenance conditions. It can be used until it deteriorates to the unacceptable condition. The segments whose conditions are worse than the trigger point d_T need to be taken into account for maintenance scheduling in the next planning horizon. It will be very dangerous when any segment is operating in an unacceptable condition, and the cost for traffic interruption of an ensuing failure will be excessive. Hence, it is better to repair or to renew the segments before they deteriorate to an unacceptable condition resulting in service disruption. Normally, the period when the segment is awaiting maintenance condition is called the *time window*, which is the best period to start maintenance activity.

This study assumes that the deterioration process follows a kind of probability distribution derived from an analysis of history data. Let $f(d,t)$ be the deterioration time probability density function, which represents the segment deterioration to an unacceptable condition by time ?t, given it is currently in condition d, $0\le d\le 1$. $d=0$ represents that the segment is in the perfect condition, and $d=1$ represents that it is in the unacceptable condition. In this study, we assume that it has a Weibull distribution. $f(d,t)$ is given as follows:

$$f(d,t)=\frac{\beta(d)}{\eta(d)}\left(\frac{t}{\eta(d)}\right)^{\beta(d)-1}e^{-\left(\frac{t}{\eta(d)}\right)^{\beta(d)}} \tag{2}$$

where $\beta(d)=\beta_0+0.2d$ *and* $\eta(d)=\eta_0(1-d)$. β_0 and η_0 are the typical shape and scale values, respectively.

Let $S_T=\left\{s_1^T,s_2^T,\ldots,s_{|S_T|}^T\right\}$ be the set of the segments whose conditions are worse than the trigger point d_T, where $s_i^T\in S$, S_T is a subset of S, and $|S_T|$ is the number of the segments awaiting maintenance. Let $D=\left\{d_1,d_2,\ldots,d_{|S_T|}\right\}$ be the set of the current conditions of the segments awaiting maintenance, where d_i is the condition value of segment s_i^T, $i=1,2,\ldots,|S_T|$.

Let $P=\{p_1,p_2,\ldots,p_m,\ldots,p_M\}$ be the initial positions of all the M available maintenance teams, where $p_m=1,2,\ldots,|S|$ is the initial position of team m represented by the segment no. Define $q_{i,j}$ and $t_{i,j,m}$ as the travel distance and the travel time, respectively, for

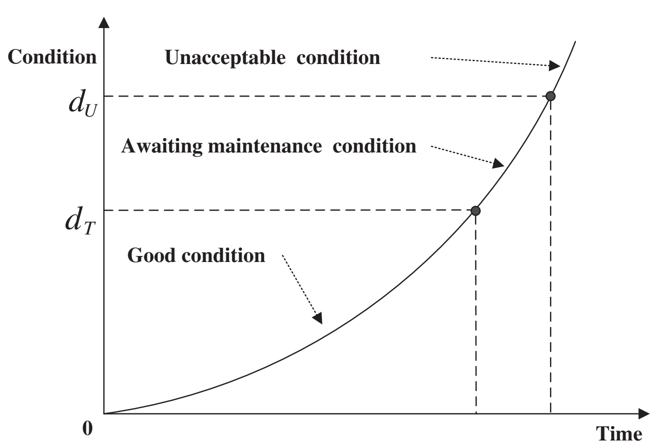

Figure 1. The deterioration process of segment

mth maintenance team to move from segment s_i to segment s_j, $i, j = 1, 2, \ldots, |S|$, which includes the average waiting time for avoiding the interruption of transportation.

Let $T = \{t_1, t_2, \ldots, t_{|S_T|}\}$ and $Team = \{m_1, m_2, \ldots, m_{|S_T|}\}$ be a solution for the schedule (unit: day) of all maintenance activities and a solution for assignment of maintenance teams, respectively, $m_i = 1, 2, \ldots, M$. We define $T_m = \{t_1^m, t_2^m, \ldots, t_{|T_m|}^m\}$ as the subset of the schedule of the maintenance activities, which will be performed by the team m. We define $Seg(\kappa, i)$ as the index number of the segment that the ith element of set κ stands for, $Seg(\kappa, i) = 1, 2, \ldots, |S|$.

In this study, four factors have to be considered for optimisation: the safety of transportation service, the cost of useful life loss of maintained segments, the maintenance cost and the travel cost. Hence, the optimisation model can be formulated as follows:

$$\text{Minimize} \quad C = c^{\text{unsafety}} \sum_{i=1}^{|S_T|} I_{\text{Sec}(s_i^T)} T_i^{\text{unsafety}} + c^{life} \sum_{i=1}^{L} T_i^{\text{life}} + \sum_{m=1}^{M} c_m^{\text{night}} Nights(T_m) + \sum_{m=1}^{M} c_m^{\text{travel}} D(T_m) \tag{3a}$$

Subject to

$$\forall t_i \in T, 0 \le t_i \le H, t_i \in N^+ \tag{3b}$$

$$\underset{m \in M}{\cup} T_m = T, \sum_{m=1}^{M} |T_m| = |T| \tag{3c}$$

$$\forall T_m \subset T, t_{p_m, \text{Seg}(T_m^{\uparrow}, 1), m} \le t_1^{m\uparrow} \tag{3d}$$

$$\forall T_m \subset T, \text{for any } i \text{ from 2 to } |T_m|, \text{if } t_i^{m\uparrow} > t_{i-1}^{m\uparrow}, t_{Seg(T_m^{\uparrow}, i-1), Seg(T_m^{\uparrow}, i), m} \le t_i^{m\uparrow} - t_{i-1}^{m\uparrow} \tag{3e}$$

where c^{unsafety} is the cost of an unsafe track segment per unit time due to the segment being in an unacceptable condition, c^{life} is the cost of the loss of the remaining useful life due to the opportunistic maintenance activities starting before the condition reaches the unacceptable point, T_i^{unsafety} is the mean unsafe working time of the ith segment, T_i^{life} is the mean life lost on the ith segment, c_m^{night} is the maintenance cost per night of the team m, and $Nights(T_m)$ is the number of working days of team m in the solution T_m. Hence, the fewer the working days, the better the solution would be due to the benefit of opportunistic maintenance. c_m^{travel} is the travel cost per mile of team m. $T_m^{\uparrow} = \{t_1^{m\uparrow}, t_2^{m\uparrow}, \ldots, t_{|T_m|}^{m\uparrow}\}$ represents the ascending set of T_m, $t_1^{m\uparrow} \le t_2^{m\uparrow} \le \ldots \le t_{|T_m|}^{m\uparrow}$. $D(T_m)$ is the total travel distance of team m under the solution T_m. It is given by

$$D(T_m) = q_{P_m, \text{Seg}(T_m^{\uparrow}, 1)} + \sum_{i=1}^{|T_m|-1} q_{\text{Seg}(T_m^{\uparrow}, i), Seg(T_m^{\uparrow}, i+1)} \tag{4}$$

Equation (3a) minimises the total costs, including the unsafe cost, the lost lifetime cost, the maintenance cost and the travel cost. Equation (3b) ensures that all maintenance activities are scheduled in the planning horizon. Equation (3c) ensures that each maintenance activity is assigned to one maintenance team. Equations (3d) and (3e) ensure that there is enough time left for one team to move to the next project.

The condition of the track geometry will only be known at the point that the measurement train passes along the network. Therefore, it is only at these discrete points in time that will be discovered that problems, such as the segments being in the unacceptable condition, will occur. If the ith segment deteriorates to an unacceptable condition by the time t, $t < t_i$, t_i is its maintenance schedule, then the unsafe working time is $t_i - t$. Hence, the mean unsafe working time T_i^{unsafety} can be given as

$$T_i^{\text{unsafety}} = \int_0^{t_i} (t_i - t) f(d_i, t) dt \tag{5}$$

Similarly, the mean life loss time T_i^{life} can be given as

$$T_i^{life} = \int_{t_i}^{\infty} (t - t_i) f(d_i, t) dt \tag{6}$$

In reality, there would be more jobs than could be carried out by the teams, and some would not be completed by the end of the month. In which case, the conditions would be monitored again and a new schedule would be formulated on the basis of the current conditions of all segments.

3. Solution methodology

It is known that the solution varies nonlinearly with the continuous variables and that the size of the problem dramatically increases with the number of segments considered. Such a scheduling problem has been proven to be an non-deterministic polynomial-time hard (NP-hard) problem. The standard methods of nonlinear programming are therefore not suitable, particularly when the number of segments considered is large. The GA-based approach has been proven to be effective on solving NP-hard problems and has been successfully applied to many engineering optimisation problems, including maintenance scheduling. Hence, the GA is used and enhanced to solve this problem. The flowchart of the enhanced GA is presented in Figure 2.

3.1. Population initialisation

The planning horizon, H, represents the maximum permitted value of each maintenance planned in the starting time. To be easily dealt with, we defined a vector $P = [p_1, \ldots, p_i, \ldots, p_N, p_{N+1}, \ldots, p_{N+j}, \ldots, p_{2N}]$ as an individual of the population also called a chromosome in GA terms, p_i is an integer, $p_i \in \{1, 2, 3, \cdots, M\}$, M is the number of maintenance teams, p_{N+j} is a floating point @number and $N = |S_T|$ is the total number of segments that are in the awaiting maintenance condition. To be specific, a chromosome $[p_1, \ldots, p_i, \ldots, p_N, p_{N+1}, \ldots, p_{N+j}, \ldots, p_{2N}]$ stands for the solution shown in the next paragraph.

Assuming that there are three maintenance teams (denoted as 1, 2 and 3) and five awaiting maintenance segments, the chromosome [1, 2, 1, 3, 2, 4.6, 2.1, 4.1, 3.1, 2.4] represents maintenance tasks of segments 1 and 3 assigned to the team 1. The schedule is on the fourth day because 4.1 is smaller than 4.6; hence, team 1 will repair segment 3 first and then go on to repair segment 1. Note that the float number 4.1 does not mean the maintenance started at 4.1; the floating point number is only used to determine which segment is to be maintained first. Similarly, team 2 initially goes to repair segment 2 and then segment 5 at the same second day. Team 3 is assigned to repair segment 4 on the third day.

To speed up the optimisation process, we initialised the population by the approach of orthogonal experimental design,[17] which belongs to a sophisticated branch of statistics and can generate better scattered solutions than random design. It has been proven by the evolutionary community that evolutionary algorithms with initial population generated by orthogonal experimental design can achieve the optimal solution with few function evaluations.[18,19] For completeness, a brief description of this approach is provided in the following paragraphs (for more details, see Montgomery[17] and Leung and Wang[20]).

We defined $L_R(Q^C)$ as an orthogonal array with R rows and C columns, where Q is the level, which is an odd number. R and C must satisfy the following equation[17]:

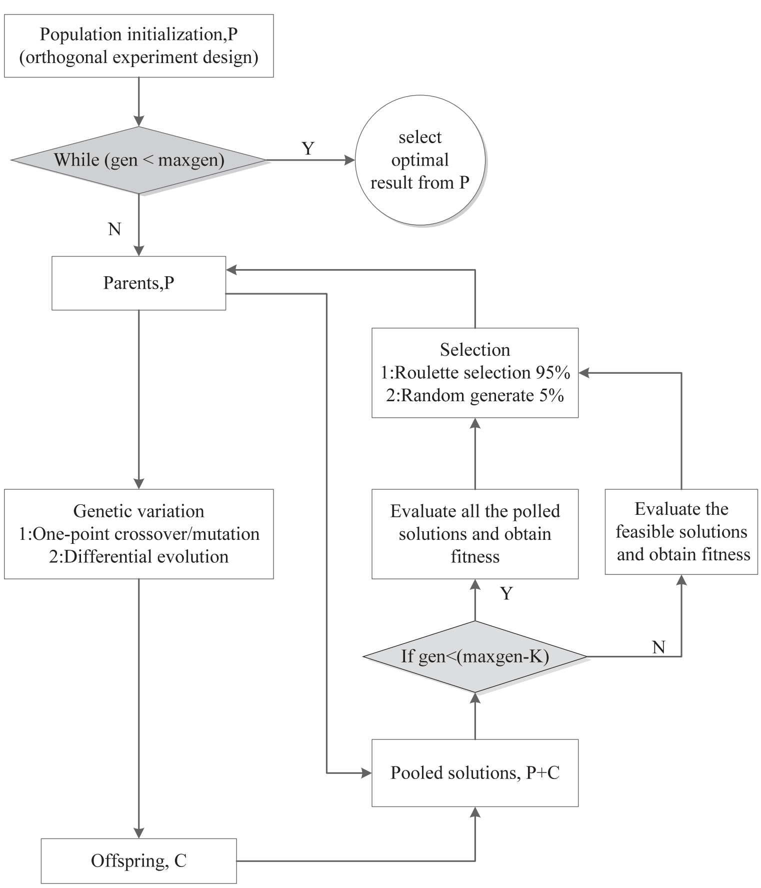

Figure 2. Flowchart of the enhanced GA used in this study

$$\begin{cases} R = Q^J \geq N_{\text{pop}} \\ C = \dfrac{Q^J - 1}{Q - 1} \geq N_{\text{var}} \end{cases} \tag{7}$$

where J is a positive integer, N_{pop} is the population size and N_{var} is the number of decision variables (here it is equal to $2N$). In this study, we used $P(R, C)$ to represent the orthogonal array (the initial population) and $a_{i,j}$ to denote the level of the jth factor in the ith combination in $P(R, C)$; if $C > N_{\text{var}}$, we only choose the first N_{var} columns. The characteristics of orthogonal experimental design guarantee that the selected array is still orthogonal. The specific algorithm to generate an orthogonal array is taken from the study of Leung and Wang.[20]

Algorithm: A = orthogonal (Q,J); input: the number of levels Q and integer J; output: orthogonal array A

for k = 1 to J do

$$j = \frac{Q^{k-1} - 1}{Q - 1} + 1$$

for i = 1 to R do

$$a_{i,j} = \left\lfloor \frac{i-1}{Q^{J-k}} \right\rfloor \bmod Q$$

end for
end for
for k = 2 to J do

$$j = \frac{Q^{k-1} - 1}{Q - 1} + 1$$

for s = 1 to j −1 do
for t = 1 to Q −1 do
for i = 1 to R do

$$a_{i,(j+(s-1)(Q-1)+t)} = \left(a_{i,s} \times t + a_{i,j}\right) \bmod Q$$

end for
end for
end for
end for
For all i ∈ [1, R] and j ∈ [1, C], ai, j = ai, j + 1. $\lfloor v \rfloor$ is the maximal integer smaller than v.

As for the configuration of Q and J, the only requirement is that the value of should satisfy Equation (7). In our case, we set $Q = M$. The reason is that in chromosome $P = [p_1, \ldots, p_i, \ldots, p_N, p_{N+1}, \ldots, p_{N+j}, \ldots, p_{2N}]$, the first N variable $p_i \in \{1, 2, 3, \cdots, M\}$; that is, each variable has M choices, so the level is set to M. For convenience, Q for the last N variable p_{N+j} is also set to M. With Q and N_{pop}, J can be set by Equation (7). Hence, each chromosome $P^{(i)} = \left[p_1^{(i)}, p_2^{(i)}, \ldots, p_j^{(i)}, \ldots, p_{2N}^{(i)}\right]$, $i = 1, 2, \ldots, N_{pop}$ of the initial population can be given by

$$p_j^{(i)} = \begin{cases} a_{i,j} & j = 1, 2, \ldots, N \\ \dfrac{a_{i,j}}{Q} H & j = N+1, N+2, \ldots, 2N \end{cases} \tag{8}$$

The population size should be large enough to search for a satisfying solution. However, the larger the population size, the greater the computing time required.

3.2. Calculation of fitness

With respect to the calculation of chromosome fitness, it is obvious that the smaller the cost $_?C$, the better the solution $_?P$ and then the higher the fitness F. First, the cost of each individual is computed by Equation (3a). Then, we check all the individuals and find the individuals that do not satisfy the constraints, denoted as PS^-. To obtain a global optimal solution, F is computed according to

$$F = \begin{cases} 1/C, & if \;\; p \notin PS^- \\ 1/\left(C \times G^k\right), & if \;\; p \in PS^- \end{cases} \tag{9}$$

where G is the current generation number and k is a user-defined integer (k=1, in this study). It represents the extent of the penalty for the solutions in PS^-. (To guarantee that all the final solutions are feasible, we only evolve the solutions that satisfy the constraints in the last K generations (K=5, in this study), and the solutions in PS^- are disregarded.)

3.3. Roulette selection

The well-known roulette selection is applied to choose the chromosomes to act as parents to perform the subsequent genetic variation process. The variation process is described in Section 3.4. After variation, the new population is generated to continue the evolution process. In this study, the probability of a chromosome being chosen is proportional to its fitness. The higher the fitness value, the greater chance the chromosome is selected. Hence, the next generation is likely to be better than the previous one. Again, to avoid the case of solutions converging to the local optimal point, in each generation, only $P_g (0 < P_g < 1)$ solutions are selected by the roulette selection, the rest of solutions $(1 - P_g)$ are newly initialised ($P_g = 95\%$ in this study).

3.4. Variation operation

Often the variation operation includes two processes, that is, crossover and mutation. Many strategies have been proposed to perform crossover and mutation. The most commonly used ones [21,22] are differential evolution, one-point crossover/mutation, uniform crossover/mutation, simulated binary crossover, Gaussian mutation, and so forth. In this study, the individual is formed by two parts: the first part is an integer vector, which is about the assignment plan of which team to maintain which segment, and the second part is a real vector of the schedule. Hence, for the first part, the one-point crossover/mutation, is applied, whereas for the second part, differential evolution (DE), which is designed for the use in real number encoding, is applied to perform variation process.

With respect to one-point crossover/mutation, as shown in Figure 3, it first selects the crossover point i randomly, dividing each parent chromosome into two parts, swapping the corresponding part with the other chromosome in the pair to produce two new child chromosomes. Second, every value in the chromosomes mutate with a probability P_m. The mutation sets is an integer vector $[1, 2, \cdots M]$, where M is the number of maintenance team.

Details for DE can be found in the studies of Price,[23] Storm and Price[24] and Das and Suganthan.[25] However, in the interest of describing a coherent methodology, a brief description is also provided here. In each generation G, DE goes through each decision vector $P_{i,G} = [p_{N+1,G}, p_{N+2,G}, \cdots, p_{i,G}, \cdots, p_{2N,G}]$ of the population and creates the corresponding trial vector $P'_{i,G}$ according to

$$p'_{i,G} = \mathrm{p}_{r_1,G} + F \times \left(\mathrm{p}_{r_2,G} - \mathrm{p}_{r_3,G}\right) \tag{10}$$

where integers $r_1, r_2, r_3 \in [N+1, N+2, \cdots 2N]$ and mutually different, and $F > 0$. The integers r_1, r_2, r_3 are chosen randomly from the interval $[N+1, N+2, \cdots 2N]$ and are different from i. F is a scale factor of the perturbation (generally,$F \in [0, 2]$, F is set to 0.5 in this study), which controls the amplification of the difference $\left(\mathrm{p}_{r_2,G} - \mathrm{p}_{r_3,G}\right)$. The higher the value of F, the less probability of solutions would be located in the local optimal area. If the $p'_{i,G}$ obtained goes out of the bounds, we then set it to the boundary value. After this, the variation is executed using the following scheme to yield the final new solution:

$$P'_{i,G+1} = \left(p'_{N+1,G+1}, p'_{N+2,G+1}, \cdots, p'_{i,G+1}, \cdots, p'_{2N,G+1}\right) \tag{11}$$

where

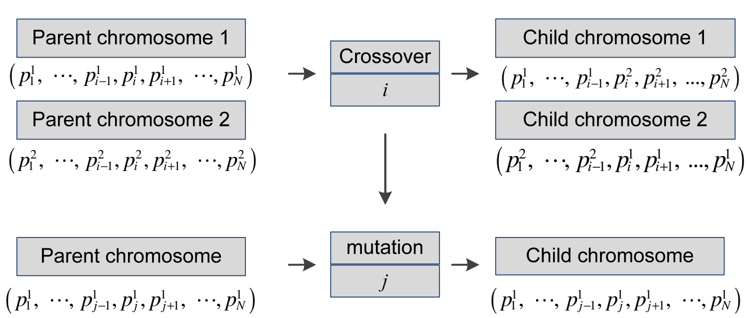

Figure 3. One-point crossover/mutation of the first part of chromosome

$$p'_{i,G+1} = \begin{cases} p'_{i,G} & \text{if } \mathrm{rand}_1 < P_c \ or \ i = \mathrm{rand}_2 \\ p_{i,G} & \text{else} \end{cases}$$

$rand_1$ is the evaluation of a uniform random generator number, and $rand_1 \in (0, 1)$. P_c is the crossover rate. It stands that if $rand_1$ is smaller than P_c, then $p'_{i,G+1}$ is substituted by $p_{i,G}$. $rand_2$ is a randomly chosen index, which ensures that $P'_{i,G+1}$ has at least one element that is different from the $P_{i,G}$.

3.5. *Replacement*

After variation, the old individual will be replaced by the new one. However, there is the case that none of the newly generated solutions is better than the existing best solution. Therefore, to guarantee that the best solution is not disregarded in the evolutionary process, the GA is applied within an $\mu+\mu$ elitist strategy.[26] It means that in each generation, the newly generated offspring N_{pop} and the parent N_{pop} are firstly pooled together, and then selection is executed on the pooled solutions, $N_{\mathrm{pop}}+N_{\mathrm{pop}}$. N_{pop} good solutions are selected as the new parents for the next generation. For example, in generation G, the fitness values of the solutions (A, B, C and D, $N_{\mathrm{pop}}=4$) are 12, 8, 6 and 4, respectively. The fitness values of the related offspring (A1, B1, C1 and D1) are 10 8 7 and 6, respectively, assuming that the higher the values of fitness, the better the solution. In the general process, the offspring will be the new parents of generation $G+1$, and so the best solution A found so far is disregarded. However, with the $\mu+\mu$ elitist strategy, solutions A, B, C and D and their offspring A1, B1, C1 and D1 are first pooled together, and then the good solutions are selected on the basis of the pooled solutions. That is to say, the solutions A B, A1 and B1 will be selected as the parents of generation $G+1$.

3.6. *Advantages of the enhanced GA*

The optimisation algorithm used in this study is an enhanced GA, which is custom designed to deal with the specific characteristics of the maintenance scheduling problem. It is more suited than the general GA for this problem in the following aspects:

(1) The orthogonal experimental design method is used to initialise population. Therefore, the initial population is uniformly distributed in the entire decision space, which can enhance the possibility of searching better solutions and improve the converging speed.
(2) The DE operator, a simple yet powerful variation operator, is used, which aids the algorithm to find a global optimal solution.
(3) The application of the $\mu+\mu$ elitist strategy[26] guarantees that the best solution would not be disregarded and the evolution is always based on the best solutions; therefore, the convergence speed is improved.
(4) In each generation, some new populations are generated to increase the diversity of solutions, which avoid the algorithm falling into the local optimal area.

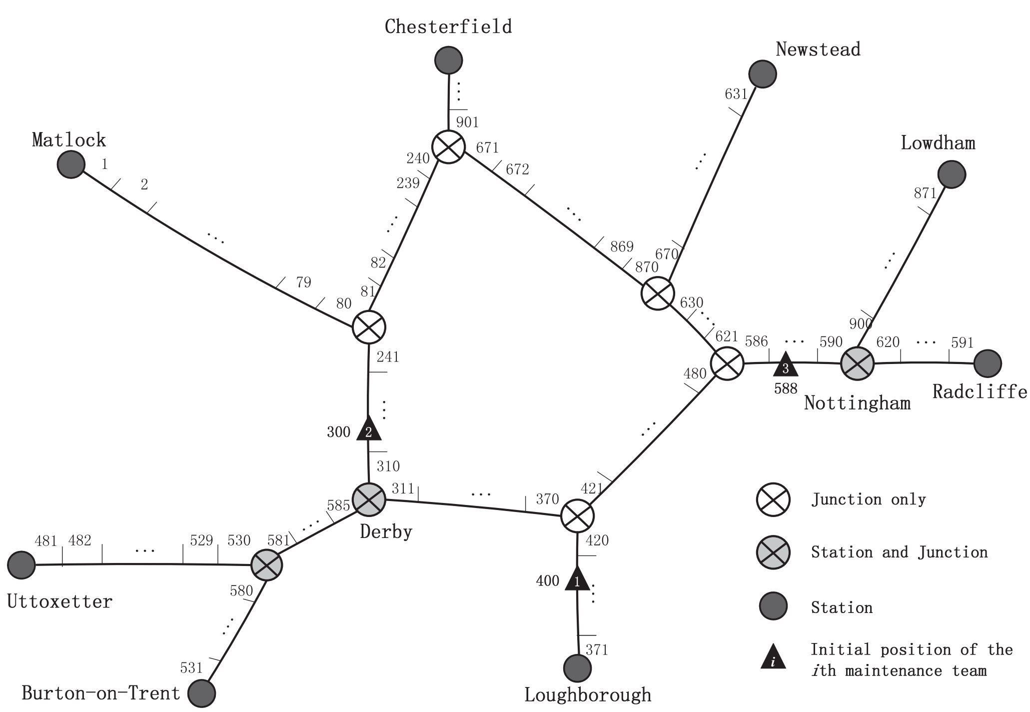

Figure 4. A regional railway network around Nottingham and Derby area in the UK

Table I. Position, importance and condition data of 30 segments awaiting maintenance

Segment no.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Position	573	71	893	531	562	894	495	520	665	18
Importance	0.8	0.5	0.7	0.8	0.8	0.7	0.9	0.9	0.8	0.5
Condition	0.853	0.867	0.915	0.868	0.915	0.914	0.944	0.855	0.910	0.966
Segment no.	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Position	361	163	101	215	354	322	49	865	571	525
Importance	1.0	0.85	0.85	0.85	1.0	1.0	0.5	0.85	0.8	0.9
Condition	0.937	0.935	0.901	0.977	0.969	0.924	0.960	0.927	0.864	0.958
Segment no.	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Position	849	644	511	30	29	486	355	629	704	590
Importance	0.85	0.8	0.9	0.5	0.5	0.9	1.0	0.95	0.85	1.0
Condition	0.869	0.971	0.958	0.893	0.987	0.970	0.951	0.901	0.950	0.889

4. Regional network case study

In this study, the proposed formulation and solution approach is applied to the maintenance scheduling of a regional railway network around the Nottingham and Derby area in the UK, as shown in Figure 4. There are six sections with more than 900 segments in this network.

Using the track geometry of the measurement train, the current condition of each segment is obtained, and there are 30 segments whose conditions are worse than the trigger condition, $d_T = 0.85$. These 30 segments need to be considered for tamping maintenance in the next planning horizon. Table I shows the position, importance and current condition data of them. The segment number indicates the unique label number of each segment in this network, as shown in Figure 4. The distance between each pair of segments is given. There are three available maintenance teams whose travelling speeds are 40, 30 and 30 mph, respectively. When they are tamping, it would take them approximately 30 min to do one segment. Every night, they have 4 h available for maintenance. The initial positions of three teams are 400, 300 and 588, respectively. The deterioration process of each segment follows a Weibull distribution whose typical values β_0 and η_0 are 1.2 and 900 days, respectively. The cost associated with unsafe segments and the lost lifetimes are 10 and 1.5 per day, respectively. The maintenance costs per night of three teams are 1.2, 1 and 1, respectively. The travel costs per mile of the three teams are 0.6, 0.5 and 0.5, respectively. It is assumed that each maintenance team does not need to return to its initial position or depot after maintenance every day.

The GA-based solution searching produces 100 generations, and the population size of each generation has 200 individuals. The selection, crossover and mutation rates are 0.95, 0.8 and 0.01, respectively.

If all three maintenance teams are available, the optimal solution is shown in Table II, where each number in black column stands for the planned day of the corresponding maintenance.

If there is only one available maintenance team that is assumed to be the first team, the optimal solution is shown in Table III. If several segments are planned for maintenance in a same day, the numbers in the black columns indicate the time sequence of maintenance. For example, if there are four segments planned to be repaired in the fifth day and the 25th segment will be done first, then the 17th, 15th and 11th segment will be the last one to be repaired.

Because there are 4 h (adjustable parameters of model) available for maintenance each night, each team can finish more than one maintenance projects in one night. Tables II and III show that the optimal solution attempts to reduce the total number of working nights so as to permit the maintenance teams to do other jobs (such as inspection) in the free nights. Hence, there are some gaps in the optimal schedule. For the reason of considering both the safety of transportation service and the loss of the remaining useful life in the optimisation model, the optimal solution would have a balance between them. Tables II and III show that the worse the condition of the track segment is, the earlier the planned maintenance time would be.

Table II. The optimal solution when all three teams are available

Segment no. / Day / Team no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1					**21**			**28**			**15**				**11**							**7**	**18**							
2	**20**	**29**		**29**			**19**			**4**		**22**	**9**	**6**		**17**	**9**		**29**	**17**	**27**				**4**	**6**	**4**		**4**	
3			**10**			**24**			**24**									**11**						**24**				**11**		**28**

Table III. The optimal solution when only the first team is available

Planned day Segment no. (position)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1 (573)																									**1**					
2 (71)																									**2**					
3 (893)																								**1**						
4 (531)																											**1**			
5 (562)																				**1**										
6 (894)																						**2**								
7 (495)						**2**																								
8 (520)																		**2**												
9 (665)																								**2**						
10 (18)									**2**																					
11 (361)					**4**																									
12 (163)												**2**																		
13 (101)																						**3**								
14 (215)									**1**																					
15 (354)					**3**																									
16 (322)																		**1**												
17 (49)					**2**																									
18 (865)																		**3**												
19 (571)																				**2**										
20 (525)												**1**																		
21 (849)																													**1**	
22 (644)						**4**																								
23 (511)									**3**																					
24 (30)																											**1**			
25 (29)					**1**																									
26 (486)						**1**																								
27 (355)						**3**																								
28 (629)																						**1**								
29 (704)												**3**																		
30 (590)																						**4**								

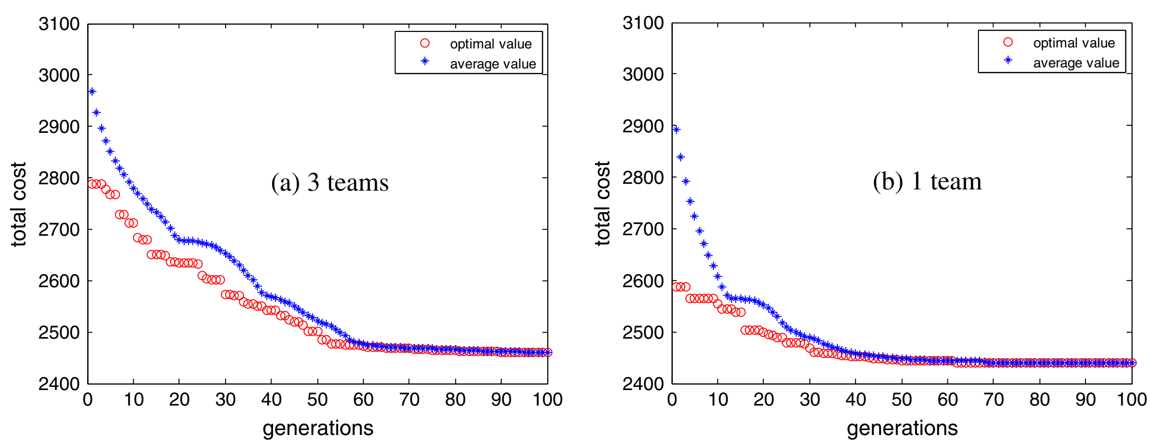

Figure 5. The change of cost reduction over changing generations

Figure 5 shows that both the average total cost and the optimal total cost is decreasing during the evolution process. Moreover, in the last generations, the average value is very close to the optimal value, which shows that most of the solutions reached the optimal point.

To show the advantages of our enhanced GA, this problem is also optimised by the general GA (see Figure 6), in which (i) the population is initialised randomly, (ii) a one-point crossover/mutation is applied for two parts of each chromosome as shown in Figure 3, (iii) the selection rate P_g is set to 1 (iv) and no elitist strategy is used. Figure 7 shows the related optimal results.

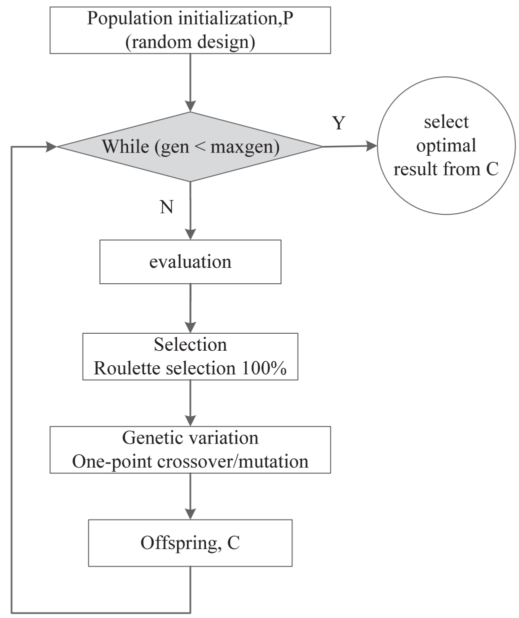

Figure 6. The flowchart of the general GA used in case study

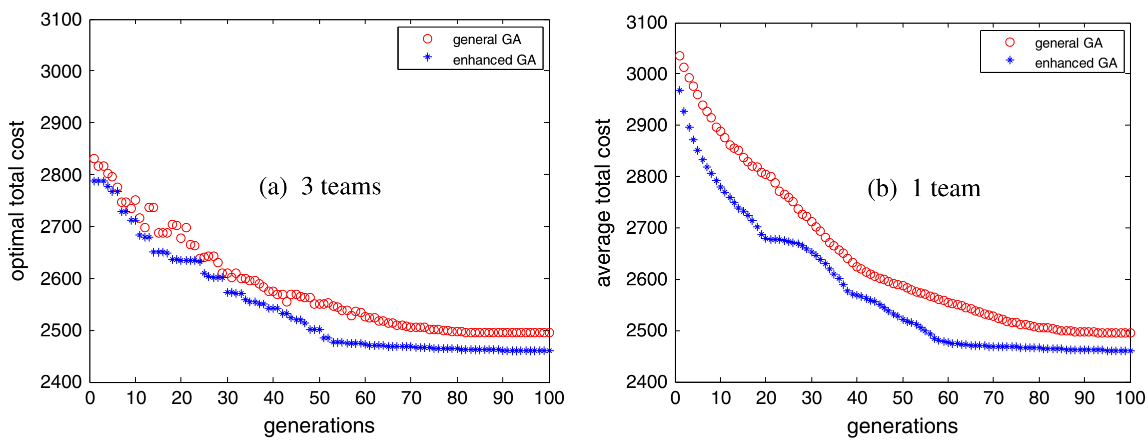

Figure 7. Experimental result of the comparison between the general GA and the enhanced GA

As shown in Figure 7, there are several key observations: (i) in the beginning, both the optimal and the average total cost obtained by the enhanced GA are lower (better) than the general GA, which shows that the use of orthogonal experimental design works better than randomly generated initial population. (ii) In the entire process, the optimal cost of the enhanced GA is lower than that of general GA, which presents that the enhanced GA can achieve better solutions. The superiority might be due to the use of elitist strategy, an advanced variation operator and also the simple diversity maintenance strategy.

5. Conclusions

This article describes the development of a model of maintenance costs for the track condition management of a regional railway network in a finite planning horizon. It then goes on to develop an enhanced GA approach to deduce the optimal scheduling for the maintenance work of one or more available teams. In this approach, the uncertainties in the deterioration process of the segments are considered, and the objective function takes into account four factors: the cost of operating unsafe segments, the cost of losing useful life when maintenance is carried out before it reaches the worst tolerable condition, the maintenance cost and the travel cost of the maintenance teams. Because this kind problem is known to be an NP-hard problem, the capability of enhanced GA-based approach is investigated to search for the best solution. The case study results show that the enhanced GA is superior to the general GA. It can help planners to make decisions for the next planning horizon after the measurement of the current conditions of all segments.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant no. 70971132). This study was carried out while Tao Zhang was visiting in the University of Nottingham. John Andrews is the Royal Academy of Engineering and Network Rail Professor of Infrastructure Asset Management. He is also Director of The Lloyd's Register Educational Trust Centre for Risk and Reliability Engineering at the University of Nottingham. He gratefully acknowledges the support of these organisations. (The Lloyd's Register Educational Trust (The LRET) is an independent charity working to achieve advances in transportation, science, engineering and technology education, training and research worldwide for the benefit of all.)

References

1. Railtrack Pubic Limited Company. Estimation of permanent way material service lives. *Railtrack line specification*. RT/CE/S/014, Issue 1, UK, 1996.
2. Federal Railroad Administration. Development of objective track quality indices. *Research Results*, RR05-01, 2005.
3. Nyström B. Punctuality and Railway Maintenance. *Licentiate thesis* 2005; 4I, ISSN: I402-I757.
4. Swier J. Outsourcing in The Netherlands; experiences and developments. http://www.promain.org/images/brussel/ProRail_ProMain_okt_2003_2.pdf. (available in April, 2011)
5. Peng F, Kang S, Li XP, Ouyang YF, Somani K, Acharya D. A Heuristic Approach to the Railroad Track Maintenance Scheduling Problem. *Computer-Aided Civil and Infrastructure Engineering* 2011; **26**:129–145.
6. Zhao J, Chan AHC, Burrow MPN. A genetic-algorithm-based approach for scheduling the renewal of railway track segments. *Proceedings of the IMechE* 2009; **223**(Part F):533–541.
7. Levitin G, Lisnianski A. Short communication optimal replacement scheduling in multi-state series–parallel systems. *Quality and Reliability Engineering International* 2000; **16**:157–162.
8. Zhang T, Nakamura M. Reliability-based Optimal Maintenance Scheduling by Considering Maintenance Effect to Reduce Cost. *Quality and Reliability Engineering International* 2005; **21**:203–220.
9. Mosheiov G, Sarig A. Scheduling a maintenance activity to minimize total weighted completion-time. *Computers and Mathematics with Applications* 2009; **57**:619–623.
10. Wang JB, Wei CM. Parallel machine scheduling with a deteriorating maintenance activity and total absolute differences penalties. *Applied Mathematics and Computation* 2011; **217**:8093–8099.
11. Cheng TCE, Hsu CJ, Yang DL. Unrelated parallel-machine scheduling with deteriorating maintenance activities. *Computers and Industrial Engineering* 2011; **60**:602–605.
12. Higgins A, Ferreira L, Lake M. Scheduling rail track maintenance to minimise overall delays. *Proceedings of The 14th International Symposium on Transportation and Traffic Theory*, Jerusalem, Israel, 1999.
13. Budai G, Dekker R, Kaymak U. Genetic and memetic algorithms for scheduling railway maintenance activities. *Econometric Institute Report* 2009; 30.
14. Budai G, Huisman D, Dekker R. Scheduling Preventive Railway Maintenance Activities. *Journal of the Operational Research Society* 2006; **57**:1035–1044.
15. Ng M, Lin DY, Waller ST. Optimal long-term infrastructure maintenance planning accounting for traffic dynamics. *Computer-Aided Civil and Infrastructure Engineering* 2009; **24**:459–469.
16. Lapaa CMF, Pereira CMNA, de Barros MP. A model for preventive maintenance planning by genetic algorithms based in cost and reliability. *Reliability Engineering and System Safety* 2006; **91**:233–240.
17. Montgomery DC. Design and analysis of experiments. John Wiley & Sons Inc.,: 111 River Street, Hoboken, NJ, USA, 2008.
18. Zeng SY, Kang LS, Ding LX. An orthogonal multi-objective evolutionary algorithm for multi-objective optimization problems with constraints. *IEEE Transactions on Evolutionary Computation* 2004; **12**(1):77–98.
19. Jiang S, Cai Z. Faster convergence and higher hypervolume for multi-objective evolutionary algorithms by orthogonal and uniform design. *Advances in Computation and Intelligence* 2010; **6382**:312–328.
20. Leung YW, Wang Y. An orthogonal genetic algorithm with quantization for global numerical optimization. *IEEE Transactions on Evolutionary Computation* 2001; **5**(1):41–53.
21. Larranaga P. Genetic algorithms for the travelling salesman problem: A review of representations and operators. *Artificial Intelligence Review* 1999; **13**(2):129–170.
22. Deb K, Kumar A. Real-coded Genetic Algorithms with Simulated Binary Crossover: Studies on Multimodel and Multiobjective Problems. *Complex Systems* 1995. **9**(6):431–454.
23. Price KV. An introduction to differential evolution. McGraw-Hill Ltd.: UK, 1999.
24. Storn R, Price K. Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces. *INTERNATIONAL COMPUTER SCIENCE INSTITUTE-PUBLICATIONS-TR*, 1995.
25. Das S, Suganthan PN. Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation* 2011; **15**(1):4–31.
26. Deb K, Pratap A, Agarwal S, *et al*. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 2002; **6**(2):182–197.

Authors' biographies

Tao Zhang is an associate professor in the College of Information Systems and Management of National University of Defense Technology (NUDT). He got his Ph.D degree from NUDT in 2004. His research interests are complex system availability and reliability, maintenance scheduling and spare management.

John Andrews is The Royal Academy of Engineering and Network Rail Professor of Asset Management in the Nottingham Transport Engineering Centre (NTEC) at the University of Nottingham. He is also the Director of The Lloyd's Register Educational Trust Centre for Risk and Reliability Engineering. Prior to this he worked for 20 years at Loughborough University where his final post was Professor of Systems Risk and Reliability. The prime focus of his initial research has been on methods for predicting system availability and reliability. Much of this work has concentrated on the Fault Tree technique and the use of Binary Decision Diagrams (BDDs) as an efficient and accurate solution method. More recently his research has been in the area of maintenance modelling and asset management.

Rui Wang is a Ph.D student in the Department of Automatic Control and Systems Engineering, University of Sheffield. His study is supported by the Chinese Scholarship Council and the National University of Defense Technology. His research interests are multi-objective optimization and Evolutionary Computation.