Materials Research Proceedings **27** (2023) 285-291 https://doi.org/10.21741/9781644902455-37

An optimized asset management petri net model for railway sections

Ali Saleh[1,a *], Darren Prescott[2,b], Rasa Remenyte[2,c] and Manuel Chiachio[1,d]

[1]Dept. Structural Mechanics and Hydraulic Engineering, Dr. Severo Ochoa Street, 18071 Granada, Spain

[2]Resilience Engineering Research Group, The University of Nottingham Science Road, Nottingham, University Park, NG7 2RD, United Kingdom

[a] alisaleh@ugr.es, [b] darren.prescott@nottingham.ac.uk, [c] r.remenyte-prescott@nottingham.ac.uk, [d] mchiachio@ugr.es

Keywords: Intelligent Decision Making, Petri Net, Reinforcement Learning, Railway

Abstract. Railways are important ways of transportation that are used massively. This makes it important to create an optimized asset management model that helps in reducing its Operation and Maintenance (O&M) costs while maintaining the quality of service and safety. Reinforcement learning (RL) is an adequate model for optimizing decisions based on unrelated factors as it connects the decision to a final goal without understanding the problem details. Also, it allows for automatic policy updates without any user intervention. On the other hand, the Petri net (PN) model, which is a bipartite graph of transitions and places, are adequate to be combined with Reinforcement learning since RL actions can be directly described by the PN transitions. In addition, PNs are suitable for maintenance modeling since it can model heterogeneous information, parallel operations, and synchronization, and provide a graphical interpretation. In this study, the Petri net method is used with Reinforcement Learning to create a tool for modeling and optimizing decisions within the maintenance of railway sections while taking into account several factors.

Introduction

Railways are a climate-smart and efficient way to move people and freight deployed in most countries worldwide. They are easy for long-distance travel, play an important role in national integration, and can carry huge loads for short and long distances. Only the UK railway industry employs around 710,000 people and contributes £42.9 billion to the economy [1]. However, as reported by *Network Rail*, the railways' Operation and Maintenance (O&M) costs are expensive, with £7.5 billion for 2020/21 [2]. This gives vital importance to find an optimal strategy for performing maintenance with reduced costs while maintaining good service.

The degradation of track geometry is subjected to much uncertainty and can be related to many factors, such as weather, traffic loads, and speed. Maintenance actions should be performed before the condition reaches a point where the asset become unfit for purpose, which may, at the best, result in system downtime, and, at worst, in potential safety risks. An example of catastrophic failure is the Potters Bar train derailment, which resulted in 7 fatalities, 76 injuries, and a £3,150,000 fine for *Network Rail* [3]. Besides, as the track geometry worsens, the probability of having rail faults increases [3]. These faults can result in breaks, safety issues, and speed restrictions if the right measures are not performed. Grinding or welding are two main actions that can be done to correct the rail faults, and sometimes replacing the rail can be considered depending on the severity of the case [3]. The track geometry can be restored by ballast maintenance techniques, which are tamping and stoneblowing, or by ballast renewal if the ballast state is highly fouled. The tamping operation causes ballast breakdown, which may result in a higher degradation rate and faster track settlement, and this result in the necessity to reduce the time interval between required tamps until it becomes no longer economical [4]. At this stage, stoneblowing, which is a more expensive and slower alternative to tamping that causes less ballast breakup can be considered [5].

Several aspects should be considered when modeling the O&M of the rail including the factors that may affect the degradation rate including the maintenance history, the factors that may cause rail faults including the condition and the age of the track, the effectiveness of each maintenance action, the consequences of different conditions of the rail, and the costs of maintenance actions. Petri net (PN) are powerful for modeling O&M as they are able to account for resource availability, concurrency, synchronization, and heterogeneous information [6]. For this, it is chosen to create an asset management model for the case of the railway while taking into account different complexities. Then, Monte Carlo Reinforcement Learning (MCRL) is used to teach an *RL agent* through interacting with the PN model so that it reaches an optimal maintenance policy that reduces the O&M costs while avoiding bad consequences.

Monte Carlo Reinforcement Learning with Petri Net Model

The term reinforcement learning (RL) is applied to machine learning methods that reward or punish desired or undesired behaviors respectively. The method teaches a learning element, called *the agent*, by trial and error. The agent's actions change the state of the environment and result in rewards that are used to evaluate how good that action was. The evaluation of the actions is known as the *state-action value function* and is used to find the optimal policy [7]. This study adopts the Monte Carlo Reinforcement learning (MCRL) method, which works by generating episodes following an initial random policy referred to as $S_0, A_0, R_1, S_1, A_1, R_2, \dots, S_{T-1}, A_{T-1}, R_T$, where S_t, A_t, and R_t are the state, action, and reward at time t respectively, and T is the terminating state of the episode. The summation of the future rewards accumulated from time step *t*, and discounted by discount rate γ, is called the discount expected return, G_t and can be calculated as follows:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots = \sum_{k=t+1}^{T} \gamma^{k-t-1} R_k \quad (1)$$

Thus, a *value function* $Q(S_t, A_t)$ is updated based on the expected return at time *t* as follows:

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha[G_t - Q(S_t, A_t)] \quad (2)$$

where $\alpha \in [0,1]$ is a learning rate parameter, with α =1 meaning that the effect of the latest update will be dominant. The value functions are evaluated to find the optimal policy that increases the long-term rewards. To do this, the actions with higher Q-values are favored to update the initial policy; this is also called *greedy policy*. A problem in the greedy policy is that it does not allow the agent to try actions with lower Q-values, which may be better than how they look if they are updated. This is known as the exploration-exploitation dilemma and can be seen in almost all RL methods. An alternative that solves this issue is the ε-greedy strategy, where the action A_t is given by:

$$A_t = \begin{cases} argmax_a Q(S_t, a) & withprobability(1-\varepsilon) \\ A \in_R A(s) & withprobability\varepsilon \end{cases} \quad (3)$$

with $\varepsilon = [0,1]$ being the exploration rate parameter and $A(s)$ is the set of actions available at state s. To ensure that all actions will be visited and updated despite their low Q-values, this method keeps a probability equal to ε.

Materials Research Proceedings **27** (2023) 285-291 https://doi.org/10.21741/9781644902455-37

Besides, a PN is defined as a directed bipartite composed of connected transitions and places. Each place contain a number of tokens that define the marking of that place and the marking of all places defines the state of the PN. From a mathematical point of view, a PN is defined as a tuple $N = \langle P, T, F, W, M_o \rangle$, where $P, T, F, W,$ and M_o are the sets of places, transitions, arcs, arcs' weights, and initial marking respectively [8]. If the number of tokens in the pre-set places of a transition is greater than or equal to the weights of its pre-set arcs, the transition can fire. This is called the *firing rule* and it controls the dynamics of the PN. When a transition fires it consumes tokens from pre-set places equal to the pre-set arcs' weights and produces tokens in the post-set places equal to the post-set arcs' weights. This causes a change in the markings, which can be described using the state equation defined as:

$$M_{k+1} = M_k + A^T u_k \tag{4}$$

where u_k is the firing vector, which is a binary vector describing the firing states of the transitions, and A^T is the incidence matrix, which represents the difference between weights of input and output arcs connecting places and transitions. Additional definitions are used to model the complexity of practical applications. For this study, timed transitions, inhibitor arcs, and reset arcs are used. Timed transitions takes a delay time before it fires after satisfying the firing rule, and this delay can be a fixed value or sampled from a stochastic distribution. The inhibitor arc, which is an arc with a circular ending, prevents the firing of the transition if the marking of the inhibiting place is more than or equal to the weight of the inhibiting arc. The reset arc, which is an arc with a filled circular ending, assigns the post-set place a marking equal to the weight of the reset arc. To combine RL with PN, an *action groups*, *g*, is defined as a group of conflicting transitions that are fired by RL agent after enabling they are enabled based on the RL policy.

Case study

A 220-yard rail section, known as *poskey*, with a track speed equal to 10 MPH and small concrete sleepers was modeled through PN, and RL was used to find the optimal maintenance schedule. The degradation rate of the section for each condition is sampled from a Weibull distribution whose parameters are related to the track speed, sleeper type, and maintenance history [3]. For a rail with a speed between 5-60 MPH and small concrete sleepers, the Weibull's shape and scale parameters, (β, η), are: (5.64e-1,2.13e-4), (1.3,1.7e-4), (9.73e-1,1.80e-4), (1.77,1.47e-4), (4.3,7.97e-5), (1.82,1.56e-4), (1.34,1.66e-4), and (1.34,1.66) after renewal, 1st tamp, 2nd or 3rd tamps, 4th or 5th tamps, 6th tamp, 7th tamp, 1st stoneblowing, and 2nd stoneblowing respectively [3]. The degradation rate is given in m/Equivalent Million Gross Tonnage (m/EMGT), so to calculate the increase in SD, the following formula is used:

$$SD_2 = SD_1 + DR(U_2 - U_1) \tag{5}$$

where SD, DR, and U represent the standard deviation, the degradation rate, and the usage of the rail respectively, and subscripts 2 and 1 represent the next and the current states. The rate of having rail faults for one *poskey* was related to the deterioration level of the rail [3]. Stacked fault rates are available for the 12 groups of rail fault types: squat, tache ovale, bolt hole, weld, other, rolling contact fatigue (RCF), wheel burn, lipping, side wear, headwear, corrugation, unknown (all). These rates can be calculated using the following formula:

$$FR[poskey/EMGT] = A \cdot \acute{S}D^1 + B \cdot \acute{S}D^2 + C \cdot \acute{S}D^1 + D \cdot 5.3 \tag{6}$$

Materials Research Proceedings **27** (2023) 285-291 https://doi.org/10.21741/9781644902455-37

where $\acute{SD}$ is the average standard deviation, and parameters A, B, C, and D are given in Table 1 for each faults group [3]. Based on the stacked rates, the probability of having a particular fault in a certain period can be modeled as follows:

- Calculate the usage $(U_2 - U_1)$[EMGT] and $\acute{SD}$ [mm] over the modeled period.
- Generate R, a random number between [0,1], $\Rightarrow r = R/\left(L \cdot (U_2 - U_1)\right)$.
- If $r < FR_{12}$: (there is a probability of having one of the faults)
  - For i=1,…11:
    - If $r < FR_i$: there is a probability of having fault i, exit loop.
- Else: there is no probability of having any of the faults

Table 1 Parameters of stacked rail fault rate against track vertical geometry polynomial fits [3].

	1-squat	2-tache ovale	3-bolt hole	4-weld	5-other	6-RCF
A	7.64E-05	7.85E-05	6.90E-05	9.38E-05	1.18E-04	1.13E-04
B	-6.45E-04	-6.55E-04	-5.45E-04	-7.96E-04	-6.94E-04	-5.46E-04
C	3.44E-03	3.73E-03	3.56E-03	4.85E-03	5.22E-03	5.07E-03
D	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	7-wheel burn	8-lipping	9-side wear	10-headwear	11-corrugation	12-unknown
A	1.12E-04	7.34E-05	1.66E-04	1.61E-04	1.60E-04	1.70E-04
B	-3.87E-04	1.31E-04	-3.48E-04	-1.98E-04	-1.92E-04	-2.31E-04
C	4.93E-03	3.87E-03	4.61E-03	4.28E-03	4.27E-03	4.33E-03
D	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

Now, several maintenance actions can be done for each of the faults, which are given as stacked probabilities in Table 2. Other than the faults, the deterioration of the rail can have a bad consequences including speed restriction and accidents. Three maintenance actions can be done to fix deteriorated rails: tamping, stoneblowing, and renewal. Each of the actions has a maintenance effectiveness that is described as the reduction of the poskey's SD. This study adopts the Network Rail (NR) maintenance effectiveness model given as [9]:

$$SD_2 = A_G + (SD_1 \cdot B_G) \tag{7}$$

where A_G and B_G for a track speed less than 20 MPH is 0.365 and 0.754 respectively for tamping and 0.88 and 0.577 respectively for stoneblowing. It is assumed that the renewal return the SD of the poskey to 0.

A PN model shown in Figure 1 was created to model the degradation, inspection, and maintenance of the poskey. Transition t_1 represents the periodic inspection that is performed every half a year. It is assumed that the *super-Red* condition can be revealed visually since it causes noise, vibration, and great fluctuation in the poskey, and this is modeled by transition t_7. Place p_1 represents that a change occurred in the poskey condition, and action is required. This enables the transitions in the action group g_1, which are t_2, t_3, t_4, and t_5 and they represent the *no-action*, *renewal*, *tamping*, and *stoneblowing* decisions respectively. Accordingly, RL agent chooses an action using the followed policy and based on the RL-state. Choosing any of the available repair types marks p_2, which represents that logistic preparation starts, then after this period ends, t_6 is fired to represent the maintenance of the poskey. An additional node type called *function* is defined in this PN model. If a function is connected from a transition, it runs when the transition fires; otherwise, it runs every time the state of the PN changes. Function f_1 updates the condition and the SD of the poskey and includes the faults rate and degradation rate models. Function f_2 calculates the time needed of the poskey to change to *super-Red* condition and assigns a delay for transition t_7 accordingly. Function f_3 checks the available actions at the current condition (the tamping action can't occur after the stoneblowing). Function f_4 updates the maintenance history of the poskey according to the chosen action. Function f_5 calculates the logistic time based on the condition and assigns the delay of t_6 accordingly, and finally, function f_6 updates the SD of the poskey based on the maintenance effectiveness model and updates the degradation rate based on the degradation model.

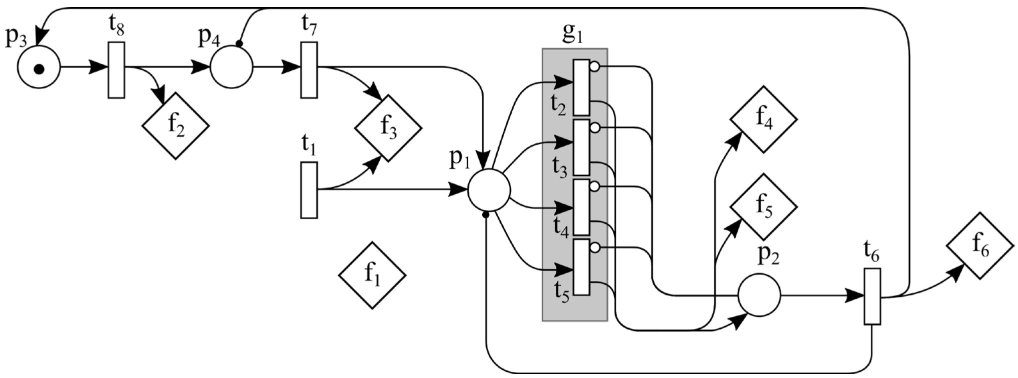

Figure 1 PN model for the degradation and maintenance of rail poskey.

The RL rewards were assigned in terms of the maintenance costs and the consequences of rail bad condition. For a rail with a track speed between 5 and 20 MPH, the condition is considered super-red, poor, good, very good, and excellent for SD more than 9900, 8300, 7400, 5200, and 0 [μm] respectively [9]. The RL environment is defined to include the deterioration level of the vertical geometry, the age, and the maintenance history of the rail. The deterioration level is assigned by discretizing the SD between values 5200 and 9900 to 10 levels, while the aging in considered by discretizing the age into intervals of 2 years each. A learning process with 400'000 episodes was considered to find the optimal maintenance strategy that reduces the Operation and Maintenance (O&M) costs while elongating the life of the rail. An episode terminates when a renewal action is chosen or when the life of the rail reaches 200 years. The RL rewards for a state-action pair were calculated as the summation of the O&M costs coming after that state divided by the time from taking the action until the end of the episode. If the rewards are not normalized by the time, the agent will choose the action with the minimal cost without considering the effect on the life of the rail, and this leads to increasing the costs per time. The O&M costs include the vertical geometry maintenance costs which are assigned as 1000, 2000, and 20'000 units for the tamping, stoneblowing, and renewal respectively, the faults maintenance costs which are assigned as 120, 100, and 80 units for the rerail, weld, and grind actions respectively, and the consequences of being in a super-red condition which is assigned as 1 unit per minute of service.

Table 2 Stacked probabilities for each maintenance action according to fault type [3].

	squat	tache ovale	bolt hole	weld	other	RCF
Rerail	0.328	0.722	0.9	0.519	0.641	0.508
Weld	0.954	0.963	0.919	0.904	0.918	0.786
Grind or other	1	1	1	1	1	1
	wheel burn	Lipping	side wear	headwear	corrugation	unknown
Rerail	0.267	0.029	0.044	0.213	0.706	0.464
Weld	0.874	0.134	0.338	0.752	0.765	0.63
Grind or other	1	1	1	1	1	1

Results

Figure 2 shows the increase of the total rewards of each episode as a function of episode number. It can be noted that the curve has reached stability by the end of the learning process, which results in a decrease in the annual O&M costs. The learning process resulted in finding the optimal action for 3878 different states. The effect of the final policy can be seen in Figure 3 which shows the SD and maintenance history for 20 random poskeys as a function of time. The figure shows that once a tamping action is taken, the frequency of maintenance actions increases. This is because each tamping operation causes ballast breakup of as much as 20 EMGT of traffic [10], resulting in fouling, faster settlement of subgrade, and faster degradation rate [4, 5]. For some of the samples, tamping was not considered at all, and this may be due to the advantages of stoneblowing over tamping including the maintenance effectiveness and degradation rate. The resulting policy takes into account various factors including the relation between the SD and the probability of having faults, the costs of maintenance actions, the consequences of being in a bad condition, the effect of each action on the life of the poskey, the maintenance history, and the age of the poksey. For this, the decisions at the same vertical geometry conditions are not the same for all poskeys. This shows the importance of considering the age of the rail and the maintenance history in the RL environment.

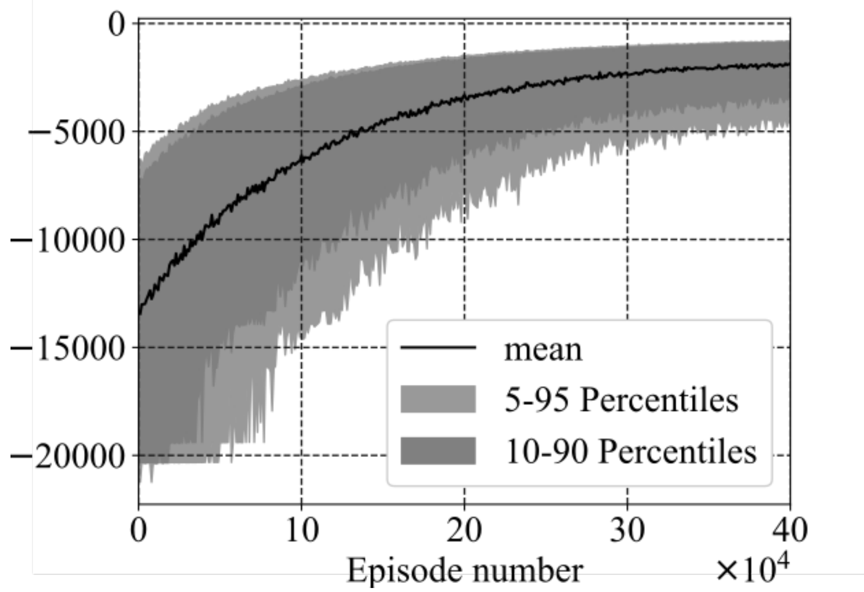

Figure 2 Total rewards as a function of episode number

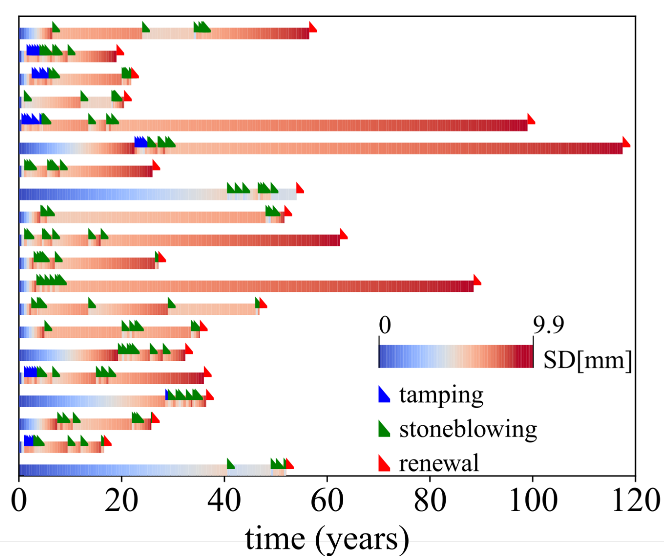

Figure 3 SD and maintenance history of 20 random Poskey samples as a function of time

Materials Research Proceedings **27** (2023) 285-291 https://doi.org/10.21741/9781644902455-37

Conclusion

A PN asset management model was created and optimized through RL for railway sections. Different factors are considered in the created model including the effectiveness of the maintenance, the effect of the rail condition on the probability of having faults, the factors influencing the degradation rate, the consequences of bad rail condition, and costs of the maintenance actions. The RL rewards are all expressed in monetary terms, and RL agent was left to interact with the PN model, which reached at the end an optimized maintenance strategy that can reduce the O&M costs while maintaining the safety and quality of service. Additional work can be done by considering a full track with different sections that has different properties when creating a full asset management model.

References

[1] O. Economics. "The economic contribution of UK rail." https://www.oxfordeconomics.com/resource/the-economic-contribution-of-uk-rail/ (accessed.

[2] Office of Rail and Road 2021, "Annual efficiency and finance assessment of Network Rail 2020-21," 2021.

[3] Steven Clarke, Darren Prescott, and J. Andrews, "Railway track asset management modelling," University of Nottingham, Nottingham eTheses, 2021. [Online]. Available: http://eprints.nottingham.ac.uk/id/eprint/64797

[4] E. T. Selig and J. M. Waters, *Track geotechnology and substructure management*. Thomas Telford, 1994. https://doi.org/10.1680/tgasm.20139

[5] B. Aursudkij, "A laboratory study of railway ballast behaviour under traffic loading and tamping maintenance," University of Nottingham Nottingham, UK, 2007.

[6] M. Chiachío, A. Saleh, S. Naybour, J. Chiachío, and J. Andrews, "Reduction of Petri net maintenance modeling complexity via Approximate Bayesian Computation," *Reliability Engineering & System Safety,* vol. 222, p. 108365, 2022/06/01/ 2022. https://doi.org/10.1016/j.ress.2022.108365

[7] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[8] T. Murata, "Petri nets: Properties, analysis and applications," *Proceedings of the IEEE,* vol. 77, no. 4, pp. 541-580, 1989. https://doi.org/10.1109/5.24143

[9] N. Rail, "Track geometry and ballast condition on NR track - parametric analysis for tier 1 and 2 modelling.," Technical Report 1 2012.

[10] M. Ottomanelli, M. Dell'Orco, and D. Sassanelli, "A Decision Support System Based on Neuro-Fuzzy System for Railroad Maintenance Planning," in *ICEIS (2)*, 2005, pp. 43-49.