



**HAL**  
open science

## Clustering approach in maintenance of capillary railway network

Danijela Đorić, Abdessamad Ait El Cadi, Said Hanafi, Nenad Mladenovic,  
Abdelhakim Artiba

► **To cite this version:**

Danijela Đorić, Abdessamad Ait El Cadi, Said Hanafi, Nenad Mladenovic, Abdelhakim Artiba. Clustering approach in maintenance of capillary railway network. *Electronic Notes in Discrete Mathematics*, 2017, 58, pp.239-246. 10.1016/j.endm.2017.03.031 . hal-03402015

**HAL Id: hal-03402015**

**<https://uphf.hal.science/hal-03402015v1>**

Submitted on 18 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Clustering approach in maintenance of capillary railway network

Danijela Đorić<sup>a,b</sup> <sup>1</sup> Abdessamad Ait El Cadi<sup>a,b</sup> Saïd Hanafi<sup>a</sup>  
Nenad Mladenović<sup>a</sup> Abdelhakim Artiba<sup>a</sup>

<sup>a</sup> *LAMIH-UMR CNRS 8201, Université de Valenciennes et du Hainaut-Cambrésis, Le Mont Houy, 59313 Valenciennes Cedex 9, France*

<sup>b</sup> *Institut de Recherche Technologique Railenium, F-59300 Famars, France*

## Abstract

Maintenance optimization of railway infrastructure includes several kinds of aspects, such as safety, economic, operational, organization and regulatory issues. Among them the regulatory issues, that are fixed, increase the maintenance costs significantly. This is especially true in so-called capillary networks (local regional railway networks), where only the freight transport exists. Hence, the question is how to minimize maintenance costs with respect to regulatory issues? To solve this problem, we propose a clustering approach. The idea is to cluster tracks, considering elements of railway infrastructure as attributes. Once railway tracks are clustered in groups with similar attributes, then the maintenance can be organized more efficiently. In this paper, Variable Neighborhood Search metaheuristic is developed to solve minimum sum of squares clustering problem. Based on the results of clustering and available real and simulated data we report 22% savings in maintenance schedule for clusters.

**Keywords :** Maintenance, railway infrastructure, optimization, minimum sum of squares, variable neighborhood search.

---

<sup>1</sup> Email:

danijela.djoric@gmail.com

# 1 Introduction

In the past, railway maintenance procedures have been traditionally planned based on the knowledge and experience of each company, accumulated over many decades of operation, with the major goal of providing a high level of safety to the infrastructures, without much concern over the economic issues [4]. But, according to the standards, maintenance is a “combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function” [5]. Also, there are many aspects of maintenance that can be the subject of optimization, including decisions regarding maintenance intervals, balance of corrective and preventive maintenance, grouping of maintenance activities, and the timing of maintenance and renewal [14]. The academic literature presents a range of methods for optimization and decision support for maintenance of critical infrastructure, but so far, these have rarely been adopted by the studied industries, and also did not treat the regulatory issues.

In this paper we take into account regulatory issues as one of really important aspect of railway maintenance optimization. The regulatory issues are fixed, and they are defined by national safety railway law, then by national safety rules, by standards and by internal rulebooks of the railways companies. This chain increases the maintenance costs significantly. This is especially true in so-called capillary networks (local regional railway networks), where only the freight transport exists. Now, the question is how to minimize maintenance costs with respect to regulatory issues? To solve this problem, we propose a clustering approach. The idea is to cluster tracks, considering elements of railway infrastructure as attributes. Once railway tracks are clustered in groups with similar attributes, then the maintenance can be organized more efficiently. Several variants of Variable Neighborhood Search (VNS) metaheuristic are developed to solve minimum sum of squares clustering problem such as: VNS- $j$ -means, VNS- $j - h - k$ -means, Variable Neighborhood Descent VND- $k - h$ -means and VND-Nested  $j - h - k$ -means. Based on the results of clustering and available real and simulated data we report 22% savings in maintenance schedule for clusters.

The remainder of this paper is organized as follows. Section 2 describes the considered clustering problem with the attributes for classification. VNS variants for classification are explained in Section 3. Furthermore, in Section 4 present our VNS for the preventive maintenance scheduling problem. Finally, Section 5 is devoted to our conclusions.

## 2 Clustering problem

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields. Clustering is a powerful tool for automated analysis of data.

Among many criteria used for cluster analysis, the minimum sum-of-squares is one of the most popular since it expresses both homogeneity and separation. Minimum sum-of-squares clustering (MSSC) consists in partitioning a given set of  $n$  entities  $\{P^1, P^2, \dots, P^n\}$ , where  $P^i \in \mathbb{R}^s$  with  $s$  is the number of attributes, into  $k$  clusters in order to minimize the sum of squared distances from the entities to the centroid of their cluster. Let  $X^j \in \mathbb{R}^s \forall j = 1, \dots, k$  the continuous variable representing the center of cluster  $j$ ;  $w_{ij} \in [0, 1]$  variable corresponding to the weight of the Euclidean distance between the center  $X^j$  and the entity  $P^i$ . A mathematical programming formulation of MSSC is as follows:

$$\min_{w, X} \sum_{i=1}^n \sum_{j=1}^k w_{ij} \|P^i - X^j\|^2 \quad (1)$$

subject to :

$$\sum_{j=1}^k w_{ij} = 1, \forall i = 1, \dots, n \quad (2)$$

$$w_{ij} \in [0, 1], \forall i = 1, \dots, n; \forall j = 1, \dots, k \quad (3)$$

$$X^j \in \mathbb{R}^s, \forall j = 1, \dots, k \quad (4)$$

Based on the regulation parameters and the rulebooks, we selected a larger spectrum of attributes: length of track, train speed, number of the trains, traffic load, traffic load of dangerous goods, type of traffic, axle load, integrated functional safety, age and quality, alternative route, time constrains, type of track, type of sleepers, and type of switches. In total we consider 14 attributes for generated data, as shown in the Table 1, and five attributes for the real data. For real data attributes are limited due to the available informations and they are: Traffic load, train speed, number of trains, number of trains with dangerous goods and traffic load dangerous goods.

The generated data in Table 1 were obtained and computed based on real references, for example: train speed is taken from the network statement of Dunkerque's port [2], as volume of the traffic load, and traffic load of dangerous goods [1]. The age of the switches is maximum 30 years (page 371, [11]).

Average rail age: 4.4 – 39 years [13]

Table 1  
The 14 parameters used for the generated data

Truck	Length	Speed	Number of trains	Tonns	Danger goods tonns	Axle load	Integrated functional safety	Ages	Alternative route	Time constraints	Elect or not	Seepers	Rails	Switches
1	0,2500	0,3333	0,1984	0,1984	0,1210	0,3182	0,0000	0,9250	0,0000	0,5000	1,0000	0,3333	0,5556	0,6000
2	0,1875	0,3333	0,1082	0,1082	0,1256	0,5000	0,0000	0,2000	0,0000	1,0000	1,0000	0,6667	0,2222	0,2000
3	0,1250	0,8333	0,8297	0,8297	0,7961	0,7273	1,0000	0,8250	0,0000	0,8333	0,0000	0,3333	0,1111	0,8000
4	0,1875	0,6667	0,7455	0,7455	0,7589	0,5000	0,0000	0,5250	0,0000	1,0000	1,0000	0,6667	0,8889	0,2000
5	0,1875	0,1667	0,0401	0,0401	0,0128	0,3182	0,0000	0,1750	0,0000	0,5000	0,0000	0,6667	0,5556	0,2000
...														
71	0,6250	1,0000	0,2926	0,2926	0,2297	0,0455	1,0000	0,0250	1,0000	0,3333	0,0000	1,0000	0,7778	0,2000
72	0,8125	0,3333	0,2325	0,2325	0,0270	0,9091	0,0000	0,0750	0,0000	0,8333	1,0000	1,0000	0,0000	1,0000
73	0,1250	0,8333	0,3026	0,3026	0,0088	0,5455	1,0000	0,6750	1,0000	0,6667	1,0000	0,0000	0,2222	0,6000
74	1,0000	1,0000	0,0721	0,0721	0,0713	0,0909	1,0000	0,5000	0,0000	0,3333	1,0000	0,6667	0,5556	1,0000
75	0,1250	0,5000	0,0922	0,0922	0,0295	0,3636	0,0000	0,3750	1,0000	1,0000	0,0000	0,3333	0,5556	0,8000

It is known that MSSC problem is an NP-hard problem [3], hence we develop VNS for this problem in the next section.

### 3 VNS for classification

VNS is a metaheuristic, or a framework for building heuristics, first time proposed by Mladenović and Hansen (1997) [12]. It is based on the systematic exploration of different neighborhood structures within a local search routine. Generally speaking, there is a change of neighborhood each time that the local search algorithm stops (i.e., when the local optimum is reached). This avoids the algorithm to be trapped, given that a local optimum may not remain optimal for another neighborhood structure [9,10]. VNS is a flexible framework for building heuristics to solve approximately combinatorial and global optimization problems. It exploits systematically the possibility of changing the definition of neighborhood structures within the search for a globally optimal (or near-optimal) solution. The VNS algorithm has been successfully applied to optimization problems in different fields, such as logistics, transportation, clustering, location, scheduling etc. The success of VNS is mainly because it exhibits the desirable properties of meta-heuristics [9,10], such as simplicity, robustness, user-friendliness, and generality.

Among many heuristics for MSSC [6], the  $k$ -means local search algorithm is the most popular. Given an initial partition, an entity  $P^j$  that belongs to the cluster  $C_l$  in the current solution is assigned to some other cluster  $C_i$ ,  $i \neq l$ . New centroids and objective function values are then updated by using simple formula; a neighborhood of the current solution is defined by all possible such exchanges (i.e., for all  $i$  and  $j$ ). A move is made if the best solution in the neighborhood is better than the current one. Otherwise, the procedure stops. Another popular heuristics is the so-called h-means algorithm. It is the same

as  $k$ -means, the only difference is the step at which the objective function is updated: in  $k$ -mean, it is done after a movement of each entity to its close cluster. But in  $h$ -means, it is done after that all entities are moved to their closest clusters respectively. A new descent local search heuristics, called  $j$ -means, is proposed in [8], where the cluster centroid  $X^i$  is relocated at some entity which does not already coincide with a centroid. Since this move may be a large one and corresponds to  $n_i$  relocations (or  $n_i$   $k$ -means moves), we refer to it as jump, hence the name  $j$ -means. Obviously, heuristic solution obtained with the jump neighborhood could be improved by the  $k$ -means and/or  $h$ -means heuristics. Using them after  $j$ -means gives a VND heuristics called  $j + h + k$ -means.

We developed, and test on our data, eight methods:  $k$ -means,  $k$ -means+,  $h$ -means,  $j$ -means, VNS:  $j$ -means, VNS:  $j - h - k$ -means, VND:  $k - h$ -means and VND: Nested  $j - h - k$ -means. All the methods work within the same cpu-time  $t_{max}$  defined by user. They are all multi-start heuristics (methods from 1 to 6) and the code at its output reports, for each method, the number of local searches within a given CPU time. VND based methods (number 7 and 8) are not multistart type. They work until CPU time elapsed. Method 7 uses  $j$ -means as a local search.

Clustering was done with  $n = 75$  elements (see Table 1) and tested for a number of clusters between five and ten. All heuristics are coded in FORTRAN 77 and run on a HP notebook core i7 45104, 8Go RAM, 256 Go SSD.

The best results were obtained, in fact, with General VNS that uses VND with three neighborhoods in mix-nested fashion (VND: Nested  $j - h - k$ -means), as a local search. These best results were found for 10 clusters. Table 2 present the best solution of ten clusters. The numbers of elements (tracks) in clusters are, respectively, 8, 10, 8, 9, 5, 9, 10, 8, 6, and 2.

Table 2  
The solution for the generated data in Table 1

Number of clusters	Number of elements (tracks)	f= 672202182.2722218									Total number of elements inside the cluster
		13	19	22	37	43	54	70	73		
1	13	19	22	37	43	54	70	73			8
2	3	6	8	15	28	31	32	42	51	58	10
3	10	20	21	26	40	48	64	72			8
4	4	7	11	16	24	34	39	60	61		9
5	9	12	33	45	68						5
6	17	18	29	41	51	59	62	63	75		9
7	1	2	5	23	46	49	50	52	55	74	10
8	14	25	27	36	44	56	66	67			8
9	35	38	47	53	65	71				6	
10	30	69									2

## 4 VNS for periodic maintenance

In the last part of this paper, we will apply our clustering approach to the scheduling preventive maintenance problem. The problem is to determine the cyclical schedule for the preventive maintenance activities under a given cost-structure assuming a fixed cycle length. The solution should minimize the total incurred costs [7].

In a such problem, it is assumed that there is a number of elements  $m$  (machines or railway infrastructures components), that need a preventive maintenance and a time interval  $T$ . During each period  $t$  of the time interval  $T$ , at most  $k$  machines can be serviced. When machine  $i$  is serviced, a given non-negative servicing cost of  $b_i$  is incurred, regardless of the period. In each period  $t$ , a machine  $i$  not yet serviced is in operation; this incurs an operation cost of  $a_i \times g_i(t)$ , where  $a_i$  is a given positive number, and  $g_i(t)$  is the number of periods elapsed since the last servicing of the machine  $i$ . The problem is now to determine a maintenance schedule such that total servicing costs and operating costs are minimized.

Based on the results of clustering and available data we determine optimal maintenance schedule for clusters (which present the groups of tracks with similar attributes) over  $n = 24$  time periods horizon. At each time period at most one cluster can be serviced. For any cycle, all the clusters must be serviced at least once. In period  $k \times n + t$  ( $k \in \mathbb{Z}, t \in T$ ), the same cluster that was serviced in period  $t$  will be serviced again. To each cluster is assigned a maintenance cost with different scenarios. The maintenance cost is calculated as  $a_i \times j_i(t)$ , where  $a_i$  is a given positive integer, and  $j_i(t)$  is the number of periods elapsed since the last servicing of the cluster  $i$ . The objective is to find a schedule that minimizes the total cost.

No /element	1 month	2 month	3 month	4 month	5 month	6 month	7 month	8 month
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								

Fig. 1. Maintenance schedule before clustering

For the real data, we have information about the maintenance costs. There are 14 elements (tracks) and a period of eight months. First schedule of the maintenance is presented in Figure 1 and it is done for all elements separately.

The maintenance cost, of this schedule, was 18 082 (Unit of costs). After applying cluster approach to the 14 elements, we get four clusters. For these four clusters, we get a new schedule for the maintenance activities, as in Figure 2. Now, the maintenance cost is 14 710 (Unit of costs). We get a saving of 22% of the total cost.

No /element	1 month	2 month	3 month	4 month	5 month	6 month	7 month	8 month
Cluster 1								
Cluster 2								
Cluster 3								
Cluster 4								

Fig. 2. Maintenance schedule after clustering

## 5 Conclusion

This paper presents a clustering problem for the maintenance of a capillary railway network. The idea is to cluster tracks, considering elements of railway infrastructure as attributes. Once railway tracks are clustered in groups with similar attributes, then the maintenance can be organized more efficiently. We develop Variable Neighborhood Search metaheuristic to solve it. Finally, based on the results of clustering and available real and simulated data, we compare two situation of the preventive maintenance scheduling problem, first without clustering and the second with clustering. We report 22% savings in maintenance scheduling cost for clusters.

We could notice that the clustering is beneficial, we save on costs, it is possible for the case of the capillary railway network and we develop a VNS metaheuristic to solve it. By this approach, we present a new direction of the maintenance philosophy in railway system. This approach takes not only into account the rail infrastructure characteristics but also the maintenance parameters, maintenance strategies, the organizational issues and regulatory ones. The research reported in this paper was concentrated on a specific organization working in railway maintenance.

## References

- [1] *Official annual report of port Dunkerque, part 2, page 6*, Technical report (2014).
- [2] *Network statement of Dunkerque's port, page 79*, Technical report (2015).
- [3] Aloise, D., A. Deshpande, P. Hansen and P. Popat, *Np-hardness of Euclidean sum-of-squares clustering*, Technical Report G-2008-33 (2008).



- [4] Carretero, J., J. M. Pérez, F. García-Carballeira, A. Calderón, J. Fernández, J. D. García, A. Lozano, L. Cardona, N. Cotaina and P. Prete, *Applying RCM in large scale systems: a case study with railway networks*, Reliability Engineering & System Safety **82** (2003), pp. 257–273.
- [5] EN, *13306 Maintenance, Maintenance terminology* (2010).
- [6] Gordon, A., “Classification: Methods for the exploratory analysis of multivariate data,” Chapman & Hall, New York, 1981.
- [7] Grigoriev, A., J. van de Klundert and F. C. Spieksma, *Modeling and solving the periodic maintenance problem*, European Journal of Operational Research **172** (2006), pp. 783–797.
- [8] Hansen, P. and N. Mladenović, *J-means: a new local search heuristic for minimum sum of squares clustering*, Pattern Recognition **34** (2001), pp. 405–413.
- [9] Hansen, P., N. Mladenović and J. A. M. Pérez, *Variable neighbourhood search: methods and applications*, 4OR **6** (2008), pp. 319–360.
- [10] Hansen, P., N. Mladenović and J. A. M. Pérez, *Variable neighbourhood search: methods and applications*, Annals of Operations Research **175** (2010), pp. 367–407.
- [11] Lichtberger, B., “Track compendium: track system, substructure, maintenance, economics,” DVV Media Group, 2011.
- [12] Mladenović, N. and P. Hansen, *Variable neighborhood search*, Computers & Operations research **24** (1997), pp. 1097–1100.
- [13] Odolinski, K. and A. S. Smith, *Assessing the cost impact of competitive tendering in rail infrastructure maintenance services: evidence from the Swedish reforms (1999 to 2011)*, Journal of Transport Economics and Policy **50** (2016), pp. 93–112.
- [14] Optirail, *Characteristics of different approaches to and frameworks for maintenance optimization methodologies* (2013).