Multi-criteria clustering analysis for large-scale public transport performance diagnosis

Imene Soumaya TOUATI[1], Karim BOUAMRANE[2] and Djamila HAMDADOU[1,3]

[1]*LIO Laboratory, University of Oran 1, Algeria*
[2]*LIO Laboratory, University of Oran 1, Algeria*
[3]*LIO Laboratory, University of Oran 1, Algeria*

Abstract: Public transport is a key factor for the global economy; therefore, it has always been a directive of governments to report on its performance to authorities and public. The purpose of the present study is providing a large-scale performance diagnosis dashboard for bus public transport systems to deal with multi-criteria context. The proposed dashboard can assist transportation authorities in undertaking a comprehensive performance evaluation both at route and system level. The methodology of this study is an integration of (i) ordered multi-criteria clustering method based on the K-means algorithm and the FLOWSORT outranking method, (ii) weighted average and (iii) PROMETHEE parameters-based single-criteria analysis. Inspired by an interesting route level evaluation methodology from recent research, a template is generated to illustrate the proposed approach. Outcomes are promising for investing in other multi-criteria clustering methods to deal with large-scale performance evaluation at both route and system levels. The proposed approach can fit any evaluation model based on performance criteria. It allows a detailed presentation of the diagnosis in spite of the large-scale context, which eases the optimization process.

Keywords: Bus public transport, Large-scale performance diagnosis, route-level analysis, GIMSI method, Multicriteria ordred clustering, FlowSort, PROMETHEE I, K-means

1. INTRODUCTION

Public Transport represents a crucial element of socio-economic and environmental development. It provides travel opportunities for average income population, notably for commuting, business, education, and leisure purposes. It continues to attract middle-class population thanks to its accessibility and affordable cost. On the economic side, it holds a high percentage of the international activity, when we take into consideration, public expenses, and operator investments (vehicle fleet, fuel consumption, etc.) or job opportunities with the continuous growth of the workforce; its impact remains very important. A strong correlation exists between public transport and town infrastructure planning. It is also a favourable choice for the environment since it represents an alternative to individual and private transport, especially in large urban agglomerations, and it represents a good investment due to its health and environmental benefits. Given these facts, governments strongly recommend developing public transport within logic of sustainability to achieve such objectives. Unfortunately, this is not the purpose of the majority of current transportation systems; therefore, it is of utmost importance to carry out a performance optimisation which is mainly based on a diagnostic phase. This procedure allows the identification and the quantification of system dysfunctions and under-scores critical sections to be able to suggest possibilities for improvements.

Several studies have addressed this issue and proposed various useful diagnostic approaches. Most of them use multi-criteria decision analysis (MCDA) as part of the diagnostic framework in order to handle the multi-criteria evaluation context. Such approaches consider evaluation either at route level nor systems level. Route-level evaluation provides a ranking within the same system to detect dysfunctions on each route, while at system level, an overall evaluation is performed to rank all systems according to their overall performance quality. The purpose of the present study is to draw benefit from both approaches by performing a large-scale analysis at the route level within a multi-criteria context. In other words, we suggest an approach able to carry out simultaneously a performance diagnosis on all routes from different systems. In this manner, it's possible to provide a diagnosis at both route and system level.

There is no doubt about the growing trend in the application of MCDA methods for performance analysis, notably within the transportation field. However, these methods are less effective when it comes to handle large-

scale evaluation, besides, they require a deep background of the domain so as to carry out a precise analysis. Due to the above reasons, we propose to employ an order multi-criteria clustering method, to deal with these limitations. After performing a survey of the current literature, we have stated that, this proposition represents the first attempt in this regard. For more powerful and efficient performance evaluation, we integrate the ordered multi-criteria clustering with a single-criteria analysis method to extract more information from the resulting distribution.

The present study's steps are as follows : (i) First, the interval multi-criteria clustering approach [1] is used at route level to output an ordered set of clusters according to performance characteristics. The clusters gather routes, from several systems, presenting similar characteristics. This process enables to perform a diagnosis on each route and to detect their deficiencies. (ii) The systems are then ranked according to their overall performance through the weighted mean of their route assignments, whereby a higher weight represents the affectation to the best ranked cluster. In this manner, the performance evaluation is carried out at the system level. (ii) Finally, the single-criterion analysis [2] is carried out on each cluster to yield preference and similarity profiles. The analysis of preference profile provides a detailed diagnosis inter-cluster while similarity profile can help the decision maker in deciding the appropriate action to follow during a possible optimization. On the one hand, our approach can provide a detailed diagnosis with a streamlined interpretation to facilitate the optimization process by providing a scheme of ordered clusters of similar routes according to performance features instead of providing a simple ranking of those routes. On the other hand, the solution allows the decision-maker to better analyse the evaluation model, which can help in checking-up its reliability and propose some improvements.

We carried out some experimentation on an inspired template from the performance evaluation dashboard model proposed by [3]. The model is based on expert evaluation; it takes into consideration different quality aspects (supply, usage, congestion, environment, etc.) and adopts a route level diagnosis according to a set of performance indicators in relation to public policy objectives.

The remainder of this paper is organized in the following manner. In section 2, we introduce the related works within the field of the public transport sector, notably the performance context so to explain our motivations to use multi-criteria clustering methods for large-scale evaluation diagnosis. Such methods are tabled with a brief literature review within section 3. Section 4 is organized in a manner to describe our methodology. We expound in the first part the strategy of the experimental evaluation model [3] used to illustrate our approach. In the second part we present in detail the interval multi-criteria clustering method [1] used for the diagnosis and then, we introduce the single-criteria analyse approach [2] employed to improve the diagnosis

process. We discuss results from experimentations and some analyses in section 5. Finally, we present conclusions and directions for future researches.

2. RELATED WORKS IN PUBLIC BUS TRANSPORT

Many studies have focused on improving freight and public transport sector; since it strongly affects the daily life. Various axes have been considered, notably in the decision support context, customer satisfaction, economic performance of transport networks, passenger information systems, planning and management of exploitation, etc.

In terms of safety, authors by [4] and [5], suggested a web-based interactive decision support system allowing carriers of dangerous materials to check the least risky routes from their outgoing points. Another proposal concerning decision support for diagnosis is found in [6]. In this paper, authors proposed a mathematical modelling to study the bus behaviour of each line from the transport network and detect the exogenous factors which affect the functioning of the system. Many researchers were interested by the passenger information systems. Let's cite the approach in [7], where authors suggested for customers a information system. The challenge of this study was to respect the transparency level of each transport company involved by the developed system. In the field of transport regulation, the authors presented in [8] a multiple criteria decision support system based on the Elimination and Choice Translating Reality (ELECTRE I) method. The regulation system enables incident detection, diagnosis and line-level evaluation.

Let's focus, in the following, on the performance context for public transport systems. Several methods have been applied; some are based on statistical techniques other optimisation methods. Few studies opted for clustering techniques while MCDA approaches are widely used. The following section is limited to some recent studies.

A. Statistical Techniques

Statistical techniques have been widely and effectively used for bus transport performance assessment, especially for the user perception-oriented assessment. In fact, several studies adopted factor analysis to identify relations from a large amount of collecting data. This technique was used in [9], [10] and [11] to identify satisfaction indicators from user point of view. By performing structural equation modelling (SEM), the method in [9] highlights the impact of the demographic characteristics on the user perception of the quality. This technique was also one adopted approach in [12] for bus transport quality assessment. The authors performed the diagnosis, according to various talents and measurable indicators such as safety and economy. Recent researches have opted for the multivariate analysis of variance (MANOVA) technique to carry out public transport performance diagnosis [13] and [14]. In the same context, the well-known economic approach, single-index model (SIM), was also performed [15] and [16].

B. Optimisation Techniques

Optimization techniques are one of the most adopted solutions within the public transport domain, such as planning; however, their use remains limited in terms of performance. We can mention [17] where Bayesian networks helped in quantifying the influence of service aspects of passenger satisfaction. It also used neural networks to assess public transport quality [18].

C. Clustering Techniques

Few studies have opted for the unsupervised clustering techniques for public transport evaluation. Intending to help decision-maker better supervise transportation systems; a decision tree classification model was developed to fetch solutions from an accident dataset [19]. A method based mainly on hierarchical cluster analysis (HCA) attempted to identify countries with similar trends in the field of transportation [20]. Another application of the clustering technique is illustrated in [21]. The authors propose k-mean approach to assess bus routes under time, volume and quality attributes.

D. MCDA Methods

MCDA approaches have several purposes in the field of bus transportation, they are mainly used to deal with performance diagnosis in a multic-riteria context and to identify the operational deficiencies or areas of improvement. Most such studies are based on MCDA methods to carry out an overall performance diagnosis according to subjective and objective dimensions within the same system or involving different systems. Such methods may be classified in two categories.

The first category performs the analysis at the route level according to selected criteria, either from the user's satisfaction survey or from the expert's point of view. Some recent papers focused on this area by integrating survey study, statistical analysis, fuzzy trapezoidal numbers and Technique for order preference by similarity to Ideal Solution (TOPSIS), the authors present in [22] a strong methodology for a general evaluation outline to measure multi-period railway lines performances. The proposed approach holds for all sorts of complex decision problems to deal with vague, unknown and subjective or uncertain data. Similar contribution is presented in [23] based on statistical analysis, fuzzy analytic hierarchy process (AHP), trapezoidal fuzzy sets and Choquet integral. In the same context, the methodology of [24] is based on a two-stage multi-criteria decision-making approach for measuring and improving the service quality of public bus transportation.The AHP method is used to perform a pairwise comparison between service quality attributes in the first stage. In the second stage, the TOPSIS procedure is adopted to rank the service quality scores of bus transit routes. In [25] the authors propose a three-stage approach based on the fuzzy SERVPERF method and envelopment (DEA) to measure and benchmark the quality of urban bus service at route level. They propose to use DEA as an MCDM tool for a full

ranking of bus routes given an overall measure of perceived quality. A recent contribution to electric transportation was proposed by [26] to select conventional bus routes for future electrification. The authors of [27] proposed a multi-criteria approach based on both TOPSIS and multi-objective optimization based on ratio analysis methods (MOORA) for ranking a group of electric buses according to a set of six criteria.

The second category covers the performance evaluation of several transportation systems to detect efficient and non-efficient transportation systems. We can mention the research study by [28] that provides an effective fuzzy multi-criteria analysis approach for the performance evaluation of urban public transport systems in a multi-criteria context. The analysis enables ranking of all the systems according to subjective criteria. Under the same background, a hybrid fuzzy methodology is proposed to analyse the performance of the public transportation system through customer satisfaction survey results [29]. The methodology consists of SERVQUAL, Delphi method, fuzzy method (AHP) and fuzzy method (TOPSIS). SERVQUAL and Delphi methods are employed to determine the set of criteria based on customer satisfaction and expert evaluation. The MCDM, AHP and TOPSIS methods are used to assign weights to the criteria and rank the different transportation systems. An efficiency-oriented multi-criteria analysis is proposed by [30] in order to rank a set of 18 transport companies. The analysis is based on the coefficient of variation approach and TOPSIS method according to seven performance indicators. Considering electric bus systems, a hybrid AHP-TOPSIS method was introduced by [31] to identify the best electric bus system. The authors suggest applying other MCDA methods, such as the Analytical Network Process (ANP), the PROMETHEE outranking method, etc.

An Interesting approach for bus public transport evaluation on both the route and the system level is presented in [32]. The system-level assessment is carried out based upon a simple qualitative interpretation and quantitative analysis. At the route level, TOPSIS is employed to rank the routes according to a set of subjective and objective criteria. To provide more detail, the routes are then clustered in four categories per criterion using the k-mean method. This procedure provides suggestions for possible route improvements. The approach allows the analysis of various systems through the use of TOPSIS at the system level.

The area of our study is more related to multi-criteria performance evaluation; therefore, we were interested by MCDA approaches used in such context. The proposed MCDA approaches for performance evaluation of public transport can be classified into two categories: route level evaluation methods and systems level evaluation methods. The route level evaluation methods provide a more detailed diagnosis compared to the second category, since the analysis criteria is carried out on the system's internal components. Despite this, the second category can provide

a large-scale performance evaluation since it includes different transportation systems. We attempted in the present study to take advantage from both solutions by performing a large-scale analysis at the route level within a multi-criteria context.

As previously stated, MCDA methods are not as effective when it comes to large-scale evaluation. We therefore propose the use of a multi-criteria clustering approach that perfectly fits such context, when the data set typically exceeds 100 alternatives to be considered for analysis. We present in the following section motivations for the multi-criteria clustering method selected for this study.

3. Multi-Criteria Clustering Techniques

In recent decades, multi-criteria clustering methods has attracted considerable attention and has been widely studied for its powerful and flexible ability to facilitate the analysis of large data sets within multi-criteria context. These approaches handle some clustering problems where preferential information is required during this process. Multi-criteria clustering methods integrate MCDA techniques into the clustering process, providing categories of similar attributes under different criteria. These clustering methods can be classified in three main categories according to their resulting clustering structure: nominal, relational and ordered clustering.

A. Nominal Clustering (No Relations Between The Clusters)

De Smet and Guzman's method [33], is the first contribution in the multi-criteria clustering field. The authors extended the classical K-means algorithm to multi-criteria context by considering a new distance definition based on alternative's profiles to detect clusters. Each profile is characterized with the preference, indifference and incomparability multi-criteria relations. It is worth noting that only outranking methods such as ELECTRE and PROMETHEE can define the incomparability relation.

B. Relational Clustering (Antisymmetric Relation on The Clusters)

Counter to the method presented by [33] where preferential information is considered at the local level, an improved version of [33] is developed to consider the preferential information at the cluster level [34]. The proposed approach is an extension of the K-means multi-criteria clustering method that integrates the multi-criteria nature of the input data. it provides an outranking binary matrix to represent the relational clustering which identify the same multi-criteria relations between each pair of clusters. Within the same context an other version is presented to deal with valued preference relations [35]. The authors propose an extension of the K-means algorithm. Starting from a random partition, the centroids of the clusters are computed, and then outranking relations between them are identified. Finally, alternatives are assigned to appropriate clusters through the K-means algorithm.

C. Ordered Clustering (Order Relation on The Clusters)

By [36] The authors presented a two-step method to deal with the ordered multi-criteria clustering problem. At the first step, clustering method is performed according to the indifference relation provided by some preference information. At the second step, the clusters derived from the clustering processes are ranked by comparing either the cluster centres nor the aggregate information of their alternatives. Heuristic approach is developed to cluster and sort under a partial order, a set of alternatives considering multi-criteria clusters [37]. As a first step, the k-means algorithm is carried out to identify clusters on the basis of the squared error criterion distance. After checking with the decision maker whether the obtained clusters are consistent, they are finally ranked on the basis of the preference relation using the ordinal ranking method Elimination and Choice Translating Reality (ELECTRE). An interesting approach is provided to address the multi-criteria clustering problem. It considers the indifference relation to build the clustering distribution which is improved by a meta-heuristic technique [38]. The strength of this method lies on its capacity to produce three clustering schemes, nominal, relational with partial order or complete order. A new distance measure based on both preference information (preference, indifference and incomparability relations) and the SOKAL and MICHENER similarity index is proposed by [39]. These measures are used to generate four clustering partitions via the k-means algorithm. An aggregation process of these partitions is then performed to produce the final optimal cluster. Following a similar strategy authors suggest in [40] the use of agreement–disagreement similarity index as distance measure for the clustering process and clustering ensemble technique to provide the final optimal partition. An other contribution in multi-criteria clustering by a meta-heuristic algorithm to cluster alternatives defined in terms of multiple incommensurable attributes on different types of scales [41]. The algorithm uses a dual bipolar-valued similarity and dissimilarity relation and performs the clustering process to fetch a set of clustering cores, and then construct a final partition further by adding the left-out alternatives that match the initial bipolar-valued similarity relation.

An extension of the k-mean algorithm with the outranking method PROMETHEE II method to yield totally ordered clusters is provided by [42]. This approach uses the same steps as the classical k-means algorithm, all the same, the alternatives are sorted into only one cluster by means of the FLOWSORT method. An analogous approach is proposed with a slight difference in the non-use of the FLOWSORT method. The approach considers only the relative net flow of PROMETHEE as measure distance while assigning alternatives to the ordered clusters [43]. An improved version of [42] was developed by considering two types of cluster, individual and interval [1]. The individual clusters are ranked according to a complete order, so as to keep the multi-criteria ordered clustering context. The interval clusters are further identified as complimentary

clusters. Using such a clustering structure, outcomes are more flexible and the model can better fit the real-world datasets. Individual clusters represent strong assignments, while interval clusters represent regions where the assignment is less strong. Following the same concept of the classical clustering method k-means extension, a new method integrating evidence theory tools and PROMETHHE preference information is presented to handle the problem of multi-criteria ordered clustering when the decision maker has a doubt on the appropriate cluster for some alternatives [44]. The most distinctive feature of this approach is the ability to generate precise and disjunctive partitions, whereby an alternative can be assigned to more than one cluster when there is no reason to select a single appropriate cluster. With the same rational logic, a more recent approach based on the classical fuzzy c-means clustering extension is developed. The clustering process employs a new objective function built on the PROMETHEE net outranking flow and the traditional validity measure for clustering [45]. A complementary contribution of [42] and [1] is developed to handle the context of hierarchical multi-criteria clustering.

The 3rd category methods are most appropriate for the present study since they generate structures with ordered clusters according to performance quality. Such information is very useful notably for ranking the different systems. It is possible to notice that ordered multi-criteria clustering methods can be divided in two categories according to their process. Some methods are based on a two-step process, the first one performs a classical clustering, and the second one allows to refine the final distribution [36], [37], [38], [39], [40] and [41]. The second category includes recent methods, those are based on an extension of the K-means algorithm by using outranking methods, notably the PROMETHEE method [42], [43], [44], [45] and [1]. As mentioned above, our proposal is the first attempt in this context. Therefore, we opted for the second category methods since their process is quite simple, furthermore, they are strongly based on the PROMTHEE outranking method enabling us to integrate the single-criteria analysis into our process since this method is also based on PROMTHEE method.

As a first trail, we decided to adopt the method proposed by [1]. This method relies on the extension of the K-mean algorithm based on the PROMETHEE I and FLOWSORT sorting procedure to assign the alternatives to either individual cluster or interval cluster. The individual clusters are ranked according to a complete order so that it is still possible to identify a cluster that is better than another one. The interval clusters are defined as complementary clusters, they are located between each two individual clusters. The benefit of this method is the use of the preference information throughout the affectation process of alternatives. Besides, the interval clusters allow to build a new individual cluster if the number of alternatives belonging to the interval cluster is important, which helps to define the appropriate number to consider for the k-means algorithm. The method provides high-quality and robust clustering distributions.

4. Methodology

This study is part of the dashboard creation process for public transport large-scale performance evolution. For the design of the dashboard general model, we opted for the GIMISI method. This approach helps locate our contribution by using an ordered Multi-criteria clustering technique [1] with a single-criterion analysis method applied to the distribution obtained [2] to enhance the diagnosis. GIMSI is a dashboard-based steering and decision support system design method. It consists of 10 steps [46] Figure 1.
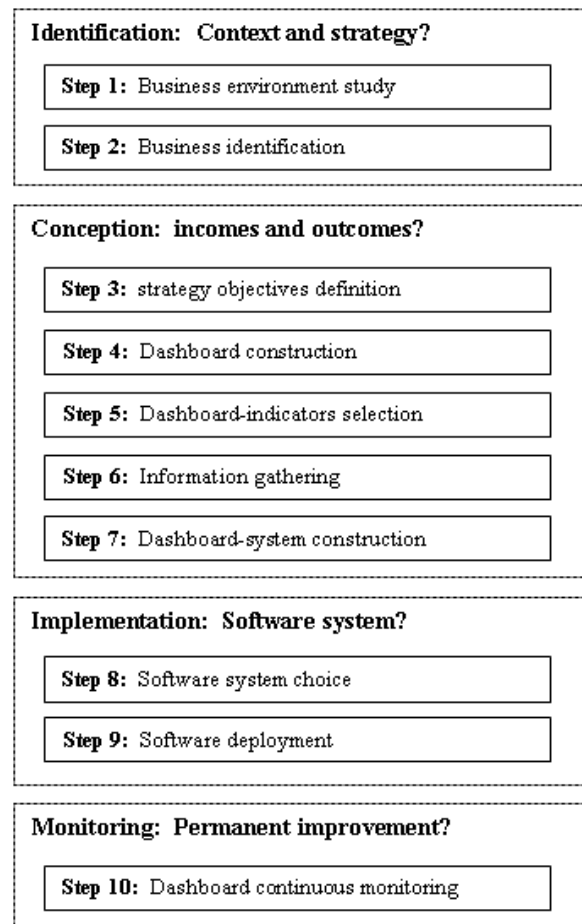

Figure 1. The steps of method GIMSI

The adopted experimental dashboard model covers the first two phases, identification and conception. The model strategy consists of enhancing the public transport through improving commercial efficiency. It considers route level evaluation according to a set of performance indicators, more details below. To achieve the main objective of our study, we generalized the model for a large-scale evaluation by considering several transportation systems. Our proposition involves the implementation phase, where we propose using multi-criteria clustering techniques with a

single-criterion analysis method to improve the dashboard diagnosis.

A. The Proposed Dashboard Experimental Model

The proposed methodology can perfectly match with the dashboard model presented in [3]. The model derives from an emerging study within public transport performance evaluation domain; the study consists of developing models able to carry out an overall analysis according to the user's perception, community satisfaction and provider benefit. This model adopts as well a recent technique, the route level analysis, which allows carrying out a large-scale performance evaluation since it handles individually each route. Besides, it grants a better interpretation of the transport system performance.

The first suggestion proposed in [3] consist of producing precise indicators able to quantify the public transport objectives according to public policies and append these indicators to those proposed by [47] to create an overall diagnosis model based on a multi-criteria performance evaluation. Three objectives were considered, the social, the environmental and the congestion reduction one. The social objective is represented by the number of travels per inhabitant during one year. To estimate the congestion reduction objective the author suggests computing the private vehicle kilometre avoided by using the public transport. The CO_2 emission ratio between that produced by public transport and that assumed produced by a private vehicle is estimated to quantify the environmental objective. The second suggestion is to develop a line-scale benchmarking to better understand the average performance of the network. In fact, the supply efficiency varies greatly from one area to another since the public policy objectives have to be respected even with less demand.

The resulting model performs an evaluation on each route, according to six categories of performance indicators Table I. The evaluation looks particularly at the commercial efficiency; it allows detecting each route's failure for which an optimization is required. By performing a multi-criteria method on the model, the set of routes can be ranked from best to worst.

We should be aware that the diagnosis model carries out a first comparison between routes by sorting them according to a typology of three main categories, mass, connection and diffusion so as to evaluate each category individually Table II.

The performance diagnosis model [3] consists of 28 performance criteria of the 6 categories, which makes difficult to represent the experiments and the clustering analysis. For this purpose, we rather select the most representative criteria of each category Table III. The proposed model aims particularly to explain the commercial efficiency factors, the main objective of any transport authority. It provides an overview of routes performance. The purpose of this study is to use the above model in order to apply our

approach and affirm its reliability in the context of large-scale performance evaluation.

B. Interval Multi-Criteria Clustering

As mentioned above, we suggest operating interval multi-criteria clustering methods on a large set of transport systems with their corresponding routes evaluated by performance criteria. The process produces an ordered set of individual and interval clusters gathering routes with similar characteristics [1]. The method represents an extension of the popular K-mean clustering algorithm by integrating the PROMETHEE I parameters and the FLOWSORT sorting during the affectation process. We should notice that PROMETHEE I threshold values and the choice of the cluster number impact the clustering quality strongly. The particularity of the interval clustering method lies in identifying individual and interval clusters. The distribution of the individual clusters is ordered from best to worst. Between theses clusters interval ones are located. The advantage of such a method is a better interpretation of the data and easy management of cluster number to consider in the classification.

The proposed technique is based on (i) The iterative k-means algorithm [48]; allowing assigning objects to clusters according to the feature similarity and on (ii) The assignment procedure of the FLOWSORT method [49] (iii) and PROMETHHE I net flows [50]; it bases the process on the relative position of an alternative according to the reference limiting or central profiles in terms of net flows, notice that we only consider the central profiles in this study.

Let us consider n alternatives $A = \{a_i, \forall i \in 1 \ldots n\}$ tested by q criteria $F = \{g_j, \forall j \in 1 \ldots q\}$ and k clusters defined by the decision-maker. The interval clustering algorithm defines firstly k individual clusters $\{C_h, \forall h \in 1 \ldots k\}$ and $\left(\sum_{h=1}^{h=k-1} (k-h) \right)$ interval clusters $\{C_{h,l \in 1 \ldots k}, h < l\}$; then performs the same main steps of the k-means algorithm to get the ordered distribution:

- Step 1: Centroids Initialisation

- Step 2: Assignment of the alternatives

- Step 3: Central profile updating

- Step 4: Iteration between the 2nd and the 3rd steps until one of the followed converge conditions is met (the distribution remains unchanged during 10 cycles or carried out the maximum of iterations, (often is set to 100))

1) Centroids Initialisation

In this step, the algorithm establishes the set of the central profiles   to represent the individual clusters $R = \{r_h, \forall h \in 1 \ldots k\}$; each profile $\{C_{h,l \in 1 \ldots k}, h < l\}$ is defined according to the alternative's evaluations on the set of criteria $\{g_j(r_h), \forall j \in 1 \ldots q\}$.

TABLE I. Multi-criteria performance diagnosis model proposed in [3]

Category	Category description	Criterion
Supply	Service quality provided	Number of services per day
		Line length (km)
		Commercial speed (km/h)
		Vehicle kilometre/day
		Road sinuosity coefficient
		Population around stop
		Connection index
Patronage	Use rate	Travels/day
		Passenger˙kilometre/day
Commercial Efficiency	exploitation cost and use rate	Travels/vehicle kilometre
		Distance/travel (km)
Social	travels/year	Employed persons
		Unemployed
		Schoolchildren/students
		Pensioners
		Other persons
		Travel/resident per year
		Travels by non-residents
		Connection travels
Congestion	Vehicle kilometre˙avoided (peak hours)	Home-to-Work Travels
		Other trips
		Total trips
		% of Home-to-Work trips
Environment	CO_2 emission (bus/private vehicle)	Bus emissions per day (CO_2)
		Private vehicle kilometre˙avoided/day
		Private vehicle emissions/day (CO_2)
		Emission ratio (CO_2)
		Private vehicle emissions/km (CO_2)

TABLE II. Typology and function of lines based on the main categories

Type	Serving area	Function	
		Decreasing travellers flow	Provide access to mass lines
Mass	Hyper down-town	X	
Connection	Down-town	X	X
Diffusion	Outskirts		X

TABLE III. The experimental model proposed for the present study, including 10 performance criteria

Category	Criterion	Criterion description
Supply	Number of services/day	Total courses insured
	Line length(km)	Distance between first and last stop
	Commercial speed(km/h)	Average speed of buses(with stops)
	Road sinuosity	Average of line distance/crow flies
	Population around stop	Population located around stops.
Patronage	Passenger˙kilometre/day	Sum of kilometres travelled/day
Commercial Efficiency	Travels/vehicle kilometre	Sum of trips/total vehicle kilometres
Social	Travel/resident/year	Total of trips of residents/year
Congestion	Vehicle kilometre˙avoided(peak hours) (Home-to-Work Travels)	Congestion reducing rate
Environment	Emission ratio(CO_2) (bus/private vehicle)	Bus emissions/ individual cars emission

The clusters are ordered from the best to the worst; the central profiles should respect the dominance condition Equation 1; they are randomly selected from the different alternatives evaluations but have to be sorted for each criterion.

$$\forall r_h, r_l \in R : if(h < l) then \forall g_j \in F, g_j(r_h) \geq g_j(r_l) \qquad (1)$$

2) Alternatives Assignment

Based on the FLOWSORT procedure [49]; the algorithm assigns the alternatives in clusters according to their net flows. The algorithm defines for each alternative $R_i = \{a_i\} \cup R$ to compute the preference degrees $pi(x,y)$ between its alternatives $\pi(x,y)$, $\forall x, y \in R_i$ and determines the set of the positive $\emptyset_{R_i}^+$ and negative $\emptyset_{R_i}^-$ flows Equation 2 and Equation 3.

$$\forall x \in R_i, \ \emptyset_{R_i}^+(x) = \frac{1}{|R_i| - 1} \sum_{(y \neq x) \in R_i} \pi(x,y) \qquad (2)$$

and

$$\forall x \in R_i, \ \emptyset_{R_i}^-(x) = \frac{1}{|R_i| - 1} \sum_{(y \neq x) \in R_i} \pi(y,x) \qquad (3)$$

Then it looks for both the closest profiles according to the positive flow Equation 4 and the negative flow Equation 5 of the alternative a_i.

$$r_h = arg \min r_j \ \left(\left|\emptyset_{R_i}^+(a_i) - \emptyset_{R_i}^+\left(r_j\right)\right|\right) \qquad (4)$$

and

$$r_l = arg \min r_j \ \left(\left|\emptyset_{R_i}^-(a_i) - \emptyset_{R_i}^-\left(r_j\right)\right|\right) \qquad (5)$$

Where C_h^+ represents the assignment to the cluster C_h according to the positive flow, and C_l^- represents the assignment to the cluster C_l according to the negative flow. The alternative a_i is then assigned to the individual cluster C_h if $(h = l)$; or to the interval cluster $C_{h,l}$ with $(h < l)$.

3) Central Profiles Updating

At the end of the previous step, the central profiles of the individual clusters C_h, $\forall h \in 1 \dots k$ should be updated. However, some alternatives may be assigned to interval clusters, which lead to different updating cases.

    a) Non-empty individual cluster

If the individual cluster is non-empty $|C_h| \neq 0$, its central profile on each criterion $g_j(r_h)$, $j \in F$ is equal to the average value of the evaluations $g_j(a_i)$ of the assigned alternatives to this cluster Equation 6.

$$|C_h| \neq 0 \Rightarrow \ then \ g_j(r_h) = \frac{1}{|C_h|} \sum_{a_i \in C_h} g_j(a_i) \qquad (6)$$

    b) Empty individual cluster

To update the empty interval cluster $|C_h| = 0$, the algorithm distinguishes between the two following cases; the first corresponds to extreme clusters while the second represents the reminder of the clusters. For both the best cluster C_1 and the worst C_k; if at least one of their

related interval clusters in non-empty the central profile $g_j(r_{1,k})$ is equal to the average value of the evaluations of alternatives in interval clusters. In the opposite case the central profiles, $g_j(r_1)$ and $g_j(r_h)$, are respectively a random value, between the best alternative evaluation and the second profile, and the penultimate one and a null value. For the non-extreme cluster, each central profile $g_j(r_h)$ $h \in 2, k-1$ is a random value from the interval comprised in the upper cluster profile and the lower cluster one.

We should notice that, the central profiles are sorted after each updating, so it respects the dominant condition.

4) Iteration and Convergence

The assignment and updating steps are iterated until one of the convergence conditions (the distribution remains unchanged during 10 cycles or carried out the maximum of iterations, (often is set to 100)) is satisfied.

5) Clustering Quality

To fix the number of the optimal clusters, for the best ordered clustering distribution, a new quality index D is defined considering both intra-homogeneity and inter-heterogeneity. The intra-homogeneity index Δ_h considers, for each individual cluster, the set of the preferences degrees $\pi(a_i, a_j)$ between each pairwise of its alternatives n_h Equation 7.

$$\Delta_h = \frac{1}{|n_h|} \sum_{a_i, a_j \in C_h} \pi(a_i, a_j) \qquad (7)$$

Intra-heterogeneity index δ_h represents the preference degree comparison between the central profile of the individual cluster C_h and the one of C_{h+1} Equation 8.

$$\delta_h(C_h, C_{h+1}) = \ \pi(r_h, r_{h+1}) - \pi(r_{h+1}, r_h) \qquad (8)$$

Finally, the quality index D is obtained by considering both indexes of all individual clusters Equation 9. This measure has to be maximised since a high quality of distribution is characterised by high intra-homogeneity and a low inter-heterogeneity.

$$D = \frac{\sum_{1,k-1} \delta_h(C_h, C_{h+1})}{\sum_{1,k} \Delta_h} \qquad (9)$$

C. Single-Criterion Analysis

Once the Clustering performed, an ordered distribution from best to worst cluster is produced. Each cluster aggregates similar lines with convergent evaluations according to the proposed criteria by the evaluation model. Subsequently, we suggest carrying out a single-criterion analysis on each cluster [2], which enables to examine in-depth the characteristics of each group and provide useful information for the diagnosis. The single-criterion analysis proposed in [2] aims to interpret the resulting distribution from an ordered clustering technique based on the PROMETHEE method. It considers the preference degree concept when performing

the analysis. The algorithm identifies for each cluster three profiles:

- The preference profile detects for each criterion, whether it makes up strength or weakness for the concerned cluster.

- The similarity profile identifies the subset of criteria where the affected alternatives are similar and the one where they are less same.

- The inconsistency profile enables to measure of the quality of the clustering distribution and identifies the inconsistency origin for each cluster.

For this study, we limit the analysis to preference and similarity profiles since we check the quality of the clustering distribution during the previous phase.

1) The Preference Profile

For a distribution comprising k clusters C_h , $\forall h \in 1 \ldots k$ the preference profile identifies k single-criterion net flow vectors Equation 10.

$$PRP(C_h) = \left[\emptyset_1(C_h), \ldots, \emptyset_q(C_h) \right] , \; \forall h \in 1 \ldots k \quad (10)$$

Where $\emptyset_j(C_h)$, $j \in 1, \ldots, q$ is the net flow of the cluster C_h for the criterion j . To obtain the $PRP(C_h)$ vector, the algorithm computes for each criterion $j \in 1, \ldots, q$ the set of uni-criterion preference index between each pairwise of alternatives $P_j(a_i, a_{i'})$, $a_i \in C_h$, $a_{i'} \in C/C_h$.

According to this set, the algorithm computes the positive single-criterion flow $\emptyset_j^+(C_h)$ Equation 11 and the negative single-criterion flow $\emptyset_j^-(C_h)$ Equation 12.

$$\emptyset_j^+(C_h) = \frac{\sum_{a_i \in C_h} \sum_{a_{i'} \in C/C_h} P_j(a_i, \; a_{i'})}{n_h(n - \; n_h)} \quad (11)$$

and

$$\emptyset_j^-(C_h) = \frac{\sum_{a_i \in C_h} \sum_{a_{i'} \in C/C_h} P_j(a_{i'}, \; a_i)}{n_h(n - \; n_h)} \quad (12)$$

Where n_h is the number of the assigned alternatives to the cluster C_h.

The single-criterion net flow is then obtained by Equation 4-C1

$$\emptyset_j(C_h) = \; \emptyset_j^+(C_h) - \; \emptyset_j^-(C_h) / \emptyset_j(C_h) \in [-1, 1] \quad (13)$$

$\emptyset_j(C_h) = \; \emptyset_j^+(C_h) - \; \emptyset_j^-(C_h) / \emptyset_j(C_h) \in [-1, 1](13)$

According to the computed vector, a criterion is considered whether strength ($\emptyset_j(C_h) > 0$) or weakness $\left(\emptyset_j(C_h) \le 0 \right)$ for the cluster.

The covariance between each pair of criteria according to the net flow can identify:

- Similarity (criteria strongly positively correlated)

- Independence (covariance tends to zero)

- Inconsistency (criteria strongly negatively correlated)

2) The Similarity Profile

The similarity profile for a cluster C_h considers intra-similarity and inter-similarity. The intra-similarity represents the indifference degree between the alternatives affected to the cluster ($a_i \in C_h$). In contrast, the inter-similarity checks up the preference degree of the assigned alternatives with the remains ones.

The similarity $SA_j(a_i, a_{i'})$ between each pair is obtained by Equation 14

$$SA_j(a_i, a_{i'}) = 1 - P_j(a_i, a_{i'}) - P_j(a_{i'}, a_i) \quad (14)$$

Where $P_j(a_i, a_{i'})$ represents the preference function between the alternatives a_i and a_j. Notice that there are six types of preference function [50].

The intra-similarity $SA_j(C_h)$ is thus computed by Equation 15:

$$SA_j(C_h) = \frac{\sum_{a_i \in C_h} \sum_{a_{i'(i \neq i')} \in C_h} SA_j(a_i, \; a_{i'})}{n_h(n_h - \; 1)} \quad (15)$$

Where n_h is the number of the assigned alternatives to the cluster C_h.

While identifying the inter-similarity of each pair, we consider the positive preference $SE_j^+(a_i, \; a_{i'})$ Equation 16, the alternative strength, and the negative ones $SE_j^-(a_i, \; a_{i'})$ Equation 17, the alternative weakness.

$$SE_j^+(a_i, \; a_{i'}) = \frac{1}{n - n_h} \sum_{a_k \in C/C_h} (1 - \left| P_j(a_i, \; a_k) - P_j(a_{i'}, \; a_k)) \right|$$
$$(16)$$

and

$$SE_j^-(a_i, \; a_{i'}) = \frac{1}{n - n_h} \sum_{a_k \in C/C_h} (1 - \left| P_j(a_k, \; a_i) - P_j(a_k, \; a_{i'})) \right|$$
$$(17)$$

The equations above help computing the positive $SE_j^+(C_h)$ and the negative $SE_j^-(C_h)$ similarity of the cluster in Equation 18 and Equation 19.

$$SE_j^+(C_h) = \frac{\sum_{a_i \in C_h} \sum_{a_{i'(i \neq i')} \in C_h} SE_j^+(a_i, \; a_{i'})}{n_h(n_h - \; 1)} \quad (18)$$

and

$$SE_j^-(C_h) = \frac{\sum_{a_i \in C_h} \sum_{a_{i'(i \neq i')} \in C_h} SE_j^-(a_i, \; a_{i'})}{n_h(n_h - \; 1)} \quad (19)$$

Finally, the global similarity of each cluster is obtained by Equation 20:

$$S_j(C_h) = SA_j(C_h) + \; SE_j^+(C_h) + \; SE_j^-(C_h) \quad (20)$$

According to the similarity vector of each cluster $PS(C_h) = [S_1(C_h), \ldots, S_q(C_h)]$, we can detect the subset of criteria

where alternatives of the same cluster are more similar $S_j(C_h) \cong S_{j'}(C_h)$. This information can speculate on the impact of future changes on each criterion. Optimisation in terms of these criteria is favourable and sometimes, it is preliminary.

5. RESULTS AND DISCUSSION

In order to check the effectiveness of the proposed solution for large-scale performance evaluation of public bus transport systems. we generated a dataset based on the benchmarking presented in the experimental dashboard [3]. The dataset sample present the evaluation of 90 lines from 9 cites according to 10 performance criteria Table IV.

We suppose that all lines are from connexion type deserving down-town area since the comparison is meaningful only when the lines are of the same type. The Table V presents an extract of the performance criteria evaluation matrix for the whole cities.

- Cr1 (Number of services per day)

- Cr2 (Commercial speed)

- Cr3 (Line length)

- Cr4 (Road sinuosity coefficient)

- Cr5 (Population around stop)

- Cr6 (Passenger kilometre per day)

- Cr7 (Travels per vehicle kilometre)

- Cr8 (Vehicle kilometre avoided in peak hours)

- Cr9 (Travel per resident per year)

- Cr10 (CO2 emission ratio (bus/ private vehicle))

A. Pre-processing Phase

The interval multi-criteria clustering [1] is based on the K-means algorithm and the outranking method PROMETHEE . We must first lay out the parameters for both algorithms and carry out a priori a data normalisation.

1) Data Normalisation

Several studies focus on choosing a proper normalisation technique as pre-processing step for clustering since it improves the final outcome. Although Z-score normalisation is considered best practice for k-means clustering [51] there is no granite that it's appropriate for the multi-criteria clustering method adopted in this study [1] since the distance function is rather based on PROMETHEE method. Beside [52] suggest performing different normalisation techniques and then select the one with the most useful relevant outcome. For these reasons we start by performing both min-max [53] and z-score normalisations and then we selected the one with the best clustering outcome. For the z-score

normalisation we transformed a priori the cost type criteria to profit type criteria.

Notice that we didn't focus on the normalisation problem since our main objective is only to stimulate the use of multi-criteria clustering rather than MCDA methods for large-scale performance evaluation. We will handle the normalisation problem in further studies.

2) PROMETHEE Parameters

In order to determine the criteria weighting we based on logic of providing high quality of service; therefore criteria with the greatest weights are number of services per day (Cr1) commercial speed (Cr2) and population around stop (Cr5). The linear preference function is selected for the PROMETHEE parameters. Preference and indifference thresholds are set by using a mathematical method to define threshold ranges and select appropriates values Table VI. For more information about the PROMETHEE method refer back to [50].

3) K-means Parameters

Interval multi-criteria clustering [1] benefit consists of the ability to choose the optimal number of clusters. A large number of alternatives assigned to interval clusters increases inter-cluster homogeneity and decreases intra-cluster heterogeneity. Thus we performed different scenarios of clustering according to both datasets from min-max and z-score normalisations.

B. Clustering Outcome Analysis

The comparison of datasets promotes min-max normalization. According to all scenarios the number of alternatives assigned to the interval clusters for the dataset from z-score normalization exceeds greatly the one from min-max normalization Figure 2. Besides, the clustering outcomes obtained from the z-score normalization dataset are unstable. For example, the line (L12) was assigned to the first cluster for (k=4) then to the second one for (k=5) and reassigned to the first cluster for (k=6).
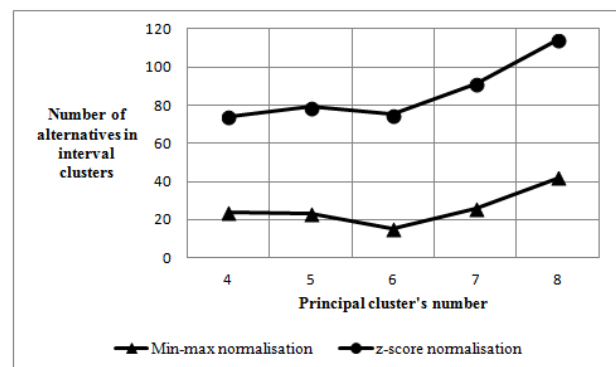

Figure 2. Comparison between min-max and z-score normalization according to the number of alternatives in the interval clusters

TABLE IV. Line distribution per city according to the experimental sample

City number	Connexion lines number per city (down-town area)
City 1	16
City 2	15
City 3	11
City 4	12
City 5	9
City 6	8
City 7	8
City 8	6
City 9	5

TABLE V. The performance criteria evaluations matrix per line and city (an extract)

Cities	Lines	Line number	Criteria				
			Cr1	Cr2	Cr3	Cr4	Cr5
City 1	L11	Line 1	356	13.5	13.7	0.3	3725
	L12	Line 2	320	14	12	0.75	3600
City 2	L21	Line 17	334	19.7	3.7	0.62	3643
	L22	Line 18	158	17.8	8	0.75	3500
City 3	L31	Line 32	190	14.8	5.7	0.35	3278
	L32	Line 33	301	14.3	6.1	0.52	4364
City 4	L41	Line 43	108	15.7	5.4	0.57	2755
	L42	Line 44	210	13.7	9	0.48	3407
City 5	L51	Line 55	160	19.3	10.3	0.52	397
	L52	Line 56	385	15.8	7.4	0.77	3497
City 6	L61	Line 64	133	14.8	18	0.35	680
	L62	Line 65	243	16.8	15.7	0.58	1537
City 7	L71	Line 72	275	15	3.2	0.76	3174
	L72	Line 73	382	16.6	8.2	0.32	4219
City 8	L81	Line 80	368	14.9	17.2	0.53	4268
	L82	Line 81	222	18.8	5.4	0.71	4024
City 9	L91	Line 86	122	19.3	5.9	0.53	2694
	L92	Line 87	356	19.5	9	0.62	2242

TABLE VI. PROMETHEE parameters

Criteria	Preference function	Indifference threshold	Preference threshold	weight	type
Cr1	Linear	0.099	0.493	0.3	profit
Cr2	Linear	0.354	1.662	0.2	profit
Cr3	Linear	0.097	0.429	0.075	profit
Cr4	Linear	0.556	2.037	0.075	cost
Cr5	Linear	0.035	0.176	0.1	profit
Cr6	Linear	0.031	0.155	0.05	profit
Cr7	Linear	0.061	0.362	0.05	profit
Cr8	Linear	0.043	0.217	0.05	profit
Cr9	Linear	0.077	0.37	0.05	profit
Cr10	Linear	0.353	1.294	0.05	cost

Let's focus on the outcomes from min-max normalization dataset. From the first scenario (k=4) to the third one (k=6) the number of alternatives assigned to interval clusters continues to decrease, unlike the following scenarios (K = 7) and (K = 8). As a result, the most relevant outcome is the clustering result with (k=6) clusters applied on the min-max normalisation dataset Table VII. The distribution consists of 75 alternatives located in principal clusters and 25 assigned to interval clusters Table VII.

Besides, the decision-maker can fetch more deep down information about each city performance. For instance, it can observe that 11 lines from city 1 are assigned to principal clusters while only 5 lines are located in interval clusters. L11, L12, L111 and L115 present great performance evaluations. The line L19 is located between the 5th and 6th principal cluster with bad evaluations, some of them are nearby to the performance of the 5th principal cluster and others are one of the worst performance evaluations, Table VII and Table VIII. The clustering outcome can at least determine cities with good overall performance. To enhance this result, we suggest determining the lines affection percent to clusters according to each city, then compute the weighted mean of each city based on the line assignments assuming a higher weight to the first cluster affectation Table IX.

This step allows ranking cities according to their overall performance. We can observe from Figure 3. that City "3" is considered as the best one unlike the city "4".
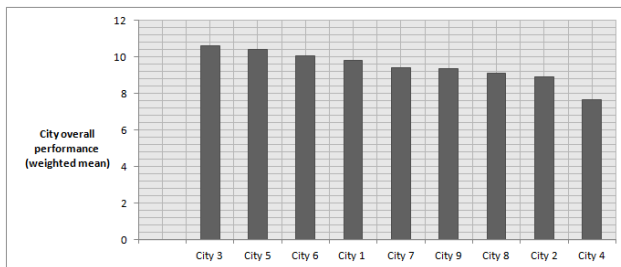

Figure 3. City overall performance comparison

1) Profile Analysis

Once the clustering accomplished, we carry out the single-criterion analysis [2] on each cluster to yield preference and similarity profiles. The analysis of both profiles provides information on how to proceed during eventual optimisation.

a) Preference profile

The preference profile identifies for each cluster the criteria with either good evaluation (profile greater than or equal to zero) or bad evaluation (profile less than zero). This information assists the decision-maker to select the set of criteria where similar lines from the same cluster require significant improvements.

We can observe from Figure 4 and Figure 6 that 50% of

the criteria for the first three clusters exhibit a favourable performance, while most criteria for the last one present bad evaluations. Lines from the best cluster (cluster 1) have almost the greatest performances, they are characterised by a high rate of line use since the criterion passenger kilometre per day (Cr6) present the highest preference profile value.
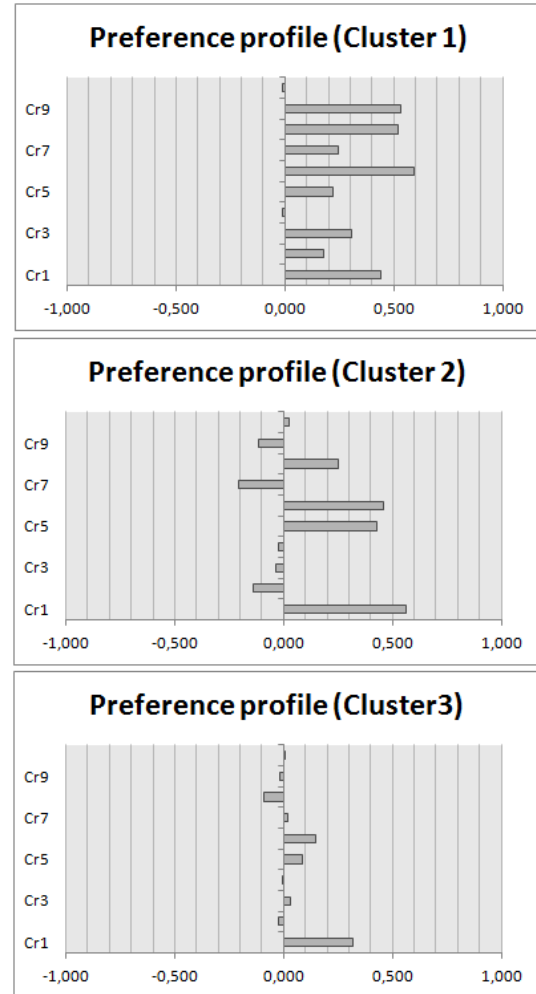

Figure 4. Preference profile outcomes for the best three clusters

2) Covariance Analysis

Based on the criteria covariance analysis according to their preference profile vectors the decision-maker can benefit from an evaluation of the adopted model for the performance diagnosis which provides its better comparison. The covariance matrix detects three main relations between criteria. Similarity when criteria are highly positively correlated; independence for covariance closes to zero and inconsistency for criteria strongly negatively correlated.

TABLE VII. The (k=6) clustering outcome (lines subset assignment per city to principal clusters)

Principal cluster	City 1	City 2	City 3	City 4	City 5	City 6	City 7	City 8	City 9
Cluster 1	L11 L12 L112 L115		L34 L311						
Cluster 2		L211			L52 L56	L66			
Cluster 3	L16 L110 L116	L26 L210 L212 L213	L311 L36 L38 L310	L43 L410	L54 L57 L59	L62 L64 L65 L67 L68	L72 L73 L74 L78	L81 L82 L85 L86	L95
Cluster 4	L13 L14 L113	L22 L23 L25 L27 L28 L29 L214	L33 L35 L39	L42 L44 L47 L48 L49 L411 L412	L55	L61 L63	L76	L83	L91 L93

TABLE VIII. The (k=6) clustering outcome (lines subset assignment per city to interval clusters)

Principal cluster	City 1	City 2	City 3	City 4	City 5	City 6	City 7	City 8	City 9
Cluster 1-2					L58				
Cluster 2-3		L21	L37				L75		L94
Cluster 3-4	L17 L114		L32	L45	L53		L71		L91
Cluster 4-5	L18 L111								
Cluster 5-6	L19								

TABLE IX. Clusters weighting for the weighted mean computation

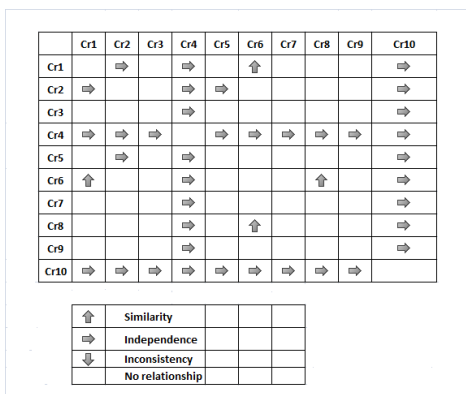| Clusters	Principal	Interval									
Cluster number	1	2	3	4	5	6	1-2	2-3	3-4	4-5	5-6
Cluster weight	6	5	4	3	2	1	5.5	4.5	3.5	2.5	1.5

Figure 5. Interpretation of the correlation matrix between the criteria

According to Figure 5 the number of services per day (Cr1) the passenger.km per day (Cr6) and the vehicle.km avoided in peak hours (Cr8), are similar; they strongly impact each other. In this case, the decision-maker can figure out beforehand that an improvement of one of these criteria enhances directly the others. For example, it's possible to consider the number of services per day (Cr1) to lay out the three criteria.

Let's look at the criteria independence context. The commercial speed (Cr2) is not correlated to both the number of service per day (Cr1) and the population around stops (Cr5); neither of them can have an effect on the commercial speed. However, we cannot figure out if there's either or nor an impact between the speed (Cr2) and the line length (Cr3). The road sinuosity (Cr4) and the CO2 emission

(Cr10) are independent of all others criteria; accordingly, they are required for the diagnosis model since they may target different individual evaluation field.

Fortunately, the profile analysis didn't detect any inconsistency relation between criteria. In the opposite case, optimisation has to balance between such criteria, a process requiring more efforts.
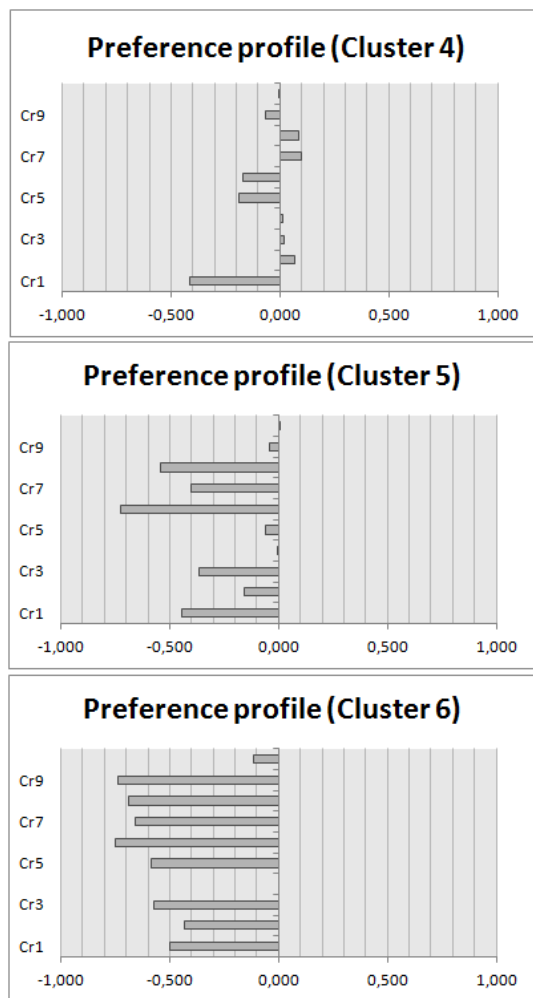

Figure 6. Preference profile outcomes for the three worst clusters

3) Similarity Profile

The similarity profile identifies, for each cluster, the criteria that promote the similarity of the alternatives; in other words, the criteria in which the profile's values of alternatives are analogous. Thereby, the decision-maker can prioritize the identified criteria, by the similarity profile, among those which require improvements according to the preference profile.

For example, alternatives from the 4th cluster are highly similar for almost the half set of criteria line length (Cr3), population around stop (Cr5), travels per vehicle kilometre

(Cr7) and travel per resident per year (Cr9) Figure 7. This case leads to an easy optimization process, unlike the other clusters with only two criteria.
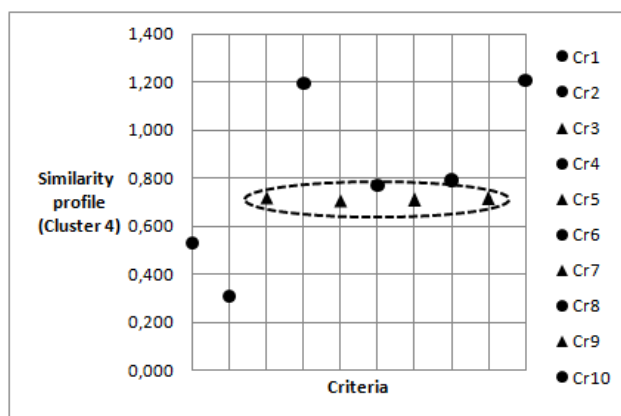

Figure 7. Similarity profile outcome (cluster 4)

Let's focus on the 2nd cluster; its alternatives are highly similar for both the road sinuosity coefficient (Cr4) and the CO_2 emission ratio (Cr10). The analysis suggests first considering improvements for the 4th criteria (Cr4) since it is a weakness for the cluster, but it won't impact the similarity between alternatives in case of future optimization. Notice that for clusters with less than two assigned alternatives, the similarity profile doesn't get consideration Figure **??**.
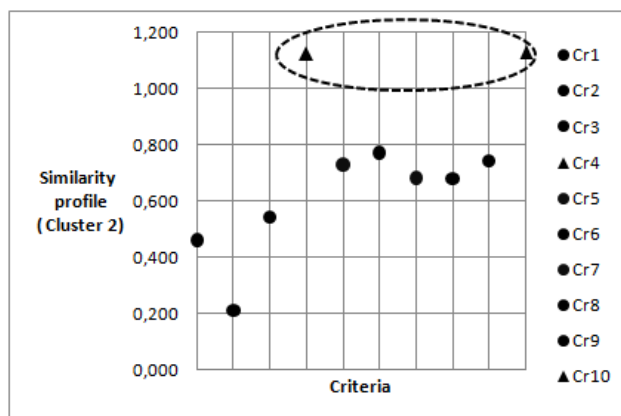

Figure 8. Similarity profile outcome (cluster 2)

6. Conclusions and Future Work

The large-scale performance evaluation is a helpful mean to lead an overall dashboard-based diagnosis of the public transport systems. This process is advantageous for the government and the large transport companies. Performing our integrated methodology on a dashboard model based on a route level evaluation provided interesting results. From a sample composed of 90 lines of 9 cities (bus transport systems), the solution generated an ordered set of 6 categories; each category holds lines with similar features.

The first cluster gather lines with almost best criteria evaluations while lines from the last cluster represent the worst evolutions. In this manner, the characteristics of the appropriate cluster for each line enable the decision-maker in identifying line's overall performance and at the same time detecting its deficiencies relative to each criterion. Besides, the resulting distribution relieves the optimization phase, since the decision-maker can by this way considers the cluster-level instead of the line-level to establish the diagnostic. Our approach provides, through weighted mean, a ranking of the different systems on the basis of lines assignment. This extra information is very useful, it allows to enhance the diagnosis with an analysis at both line and system level.

The single-criterion analysis provides detailed information of the diagnosis, thanks to preference profile, similarity profile and the criteria covariance analysis. The preference profiles outcomes allow the decision-maker to select the set of criteria where similar lines from the same cluster require significant improvements. Covariance analysis pinpoints all the criteria that are highly positively correlated for each cluster. Thus, if the decision maker focuses on one of these criteria, he can anticipate an improvement on the others. Based on the similarity profile, the decision maker can recognize for each cluster, which criteria raise the similarity of the alternatives. These criteria are more privileged for further improvement according to the preference profile because they increase the homogeneity inside clusters.

The experimental findings are an opportunity to prove that multi-criteria clustering can represent a relevant solution for large-scale performance evaluation at both route and system level, which leads to many directions for future researches in this sphere. Since it's a first attempt in this context, a number of research issues are still to be investigated. In order to further reinforce our study, we intend to carry out a comparative study with other methods of multi-criteria clustering. Further effort must be accomplished for the pre-treatment phase since it strongly impacts the quality of the clustering outcome. For example, perform other data normalization techniques. We can also integrate techniques helping while determining the PROMETHEE parameters to relieve the decision-maker diagnosis process. We also suggest investing in the management of qualitative and imprecise evaluations by considering fuzzy multi-criteria clustering; an aspect we did not address in this research.

7. Acknowledgement

References

[1] R. Sarrazin, Y. De Smet, and J. Rosenfeld, "An extension of PROMETHEE to interval clustering," *Omega (United Kingdom)*, vol. 80, pp. 12–21, oct 2018.

[2] M. A. Boujelben, "A unicriterion analysis based on the PROMETHEE principles for multicriteria ordered clustering," *Omega (United Kingdom)*, vol. 69, pp. 126–140, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.omega.2016.08.007

[3] B. Faivre d'Arcier, "Measuring the performance of urban public transport in relation to public policy objectives," *Research in Transportation Economics*, vol. 48, pp. 67–76, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.retrec.2014.09.033

[4] D. Hamdadou, E. Alnafie, K. Bouamrane, I. Behilil, and A. Benbella, "A MultiCriteria Spatial WEB Decision Support System for Hazardous Materials Road Transport," vol. 6, no. 1, pp. 1–20, 2018.

[5] K. Bouamrane, D. Hamdadou, K. Yachba, and K. Guenachi, "Towards a decision support system, application: itinerary road modification for road transport of hazardous materials," *International Journal of Decision Sciences, Risk and Management*, vol. 4, no. 3/4, p. 175, 2012.

[6] K. Bouamrane, B. H. Ali, and N. Belayachi, "Towards a numerical model for the representation of an urban transportation system," *CEUR Workshop Proceedings*, vol. 867, no. March 2016, pp. 93–101, 2012.

[7] Z. Bendaoud and K. Bouamrane, "A traveller information system: Minimisation of the number of graphs' nodes involved when processing route requests," *Transport and Telecommunication*, vol. 15, no. 4, pp. 280–291, 2014.

[8] K. Yachba, Z. Bendaoud, and K. Bouamrane, "Toward a Decision Support System for Regulation in an Urban Transport Network," *International Journal of Strategic Information Technology and Applications*, vol. 9, no. 2, pp. 1–17, 2018.

[9] L. H. P. Yaya, M. F. Fortià, C. S. Canals, and F. Marimon, "Service quality assessment of public transport and the implication role of demographic characteristics," *Public Transport*, vol. 7, no. 3, pp. 409–428, 2015.

[10] C. Morton, B. Caulfield, and J. Anable, "Customer perceptions of quality of service in public transport: Evidence for bus transit in Scotland," *Case Studies on Transport Policy*, vol. 4, no. 3, pp. 199–207, 2016.

[11] H. Mohd Noor, A. R. Md Noor, N. Nasrudin, and J. Foo, "Public Transport Service Attributes: Travelers' perception in Malaysia," *Journal of ASIAN Behavioural Studies*, vol. 3, no. 10, p. 124, 2018. [Online]. Available: http://dx.doi.org/10.21834/jabs.v3i10.311

[12] X. Cheng, Y. Cao, K. Huang, and Y. Wang, "Modeling the Satisfaction of Bus Traffic Transfer Service Quality at a High-Speed Railway Station," *Journal of Advanced Transportation*, vol. 2018, 2018.

[13] F. F. Gonul and R. A. Solano, "Empirical insights on improving bus reliability at a rural transit system," *Asian Journal of Empirical Research*, vol. 7, no. 12, pp. 306–319, 2017.

[14] J. Weng, X. Di, C. Wang, J. Wang, and L. Mao, "A bus service evaluation method from passenger's perspective based on satisfaction surveys: A case study of Beijing, China," *Sustainability (Switzerland)*, vol. 10, no. 8, pp. 1–15, 2018.

[15] I. Olivková, "Model for measuring passenger satisfaction and assessing mass transit quality," *Journal of Public Transportation*, vol. 18, no. 3, pp. 52–70, 2015.

[16] J. de Oña, R. de Oña, L. Eboli, and G. Mazzulla, "Index numbers for monitoring transit service quality," *Transportation Research Part A: Policy and Practice*, vol. 84, pp. 18–30, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.tra.2015.05.018

[17] J. Wu, M. Yang, S. Rasouli, and C. Xu, "Exploring passenger assessments of bus service quality using bayesian networks," *Journal of Public Transportation*, vol. 19, no. 3, pp. 36–54, 2016.

[18] C. Garrido, R. De Oña, and J. De Oña, "Neural networks for analyzing service quality in public transportation," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6830–6838, 2014.

[19] H. Ibrahim and B. H. Far, "Data-oriented intelligent transportation systems," *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration, IEEE IRI 2014*, pp. 322–329, 2014.

[20] O. Nežerenko, O. Koppel, and T. Tuisk, "Cluster approach in organization of transportation in the Baltic Sea Region," *Transport*, vol. 32, no. 2, pp. 167–179, 2017.

[21] R. A. Kadir, Y. Shima, R. Sulaiman, and F. Ali, "Clustering of public transport operation using K-means," in *2018 IEEE 3rd International Conference on Big Data Analysis, ICBDA 2018*, no. Ivi, 2018, pp. 427–432.

[22] N. Aydin, "A fuzzy-based multi-dimensional and multi-period service quality evaluation outline for rail transit systems," *Transport Policy*, vol. 55, no. December 2015, pp. 87–98, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.tranpol.2017.02.001

[23] N. Aydin, E. Celik, and A. T. Gumus, "A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul," *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 61–81, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.tra.2015.03.029

[24] S. Güner, "Measuring the quality of public transportation systems and ranking the bus transit routes using multi-criteria decision making techniques," *Case Studies on Transport Policy*, vol. 6, no. 2, pp. 214–224, 2018. [Online]. Available: https://doi.org/10.1016/j.cstp.2018.05.005

[25] K. Zehmed and F. Jawab, "A combined approach based on fuzzy SERVPERF and DEA for measuring and benchmarking the quality of urban bus transport service at the route level," *Industrial Engineering and Management Systems*, vol. 19, no. 2, pp. 442–459, 2020.

[26] S. Analysis, "Multi-Criteria Analysis in the Decision-Making Process on the Electrification of Public Transport in Cities in Poland : A Case," 2021.

[27] M. Hamurcu and T. Eren, "APPLICATIONS OF THE MOORA AND TOPSIS METHODS FOR DECISION OF ELECTRIC VEHICLES IN PUBLIC TRANSPORTATION TECHNOLOGY," vol. 37, no. 4, pp. 251–263, 2022.

[28] C. H. Yeh, H. Deng, and Y. H. Chang, "Fuzzy multicriteria analysis for performance evaluation of bus companies," *European Journal of Operational Research*, vol. 126, no. 3, pp. 459–473, 2000.

[29] Ö. N. Bilişik, M. Erdoğan, I. Kaya, and H. Baraçli, "A hybrid fuzzy methodology to evaluate customer satisfaction in a public transportation system for Istanbul," *Total Quality Management and Business Excellence*, vol. 24, no. 9-10, pp. 1141–1159, 2013.

[30] R. Vavrek, "Capital City as a Factor of Multi-Criteria Decision Analysis — Application on Transport Companies in the Czech Republic," pp. 1–17, 2020.

[31] M. Hamurcu and T. Eren, "Electric Bus Selection with Multicriteria Decision Analysis for Green Transportation," 2020.

[32] M. N. Hassan, Y. E. Hawas, and K. Ahmed, "A multi-dimensional framework for evaluating the transit service performance," *Transportation Research Part A: Policy and Practice*, vol. 50, pp. 47–61, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.tra.2013.01.041

[33] Y. De Smet and L. M. Guzmán, "Towards multicriteria clustering: An extension of the k-means algorithm," *European Journal of Operational Research*, vol. 158, no. 2, pp. 390–398, 2004.

[34] Y. De Smet and S. Eppe, "Evolutionary Multi-Criterion Optimization," *Evolutionary Multi-Criterion Optimization*, vol. 5467, no. April, pp. 380–392, 2009. [Online]. Available: http://www.springerlink.com/content/x79325217n618v74

[35] S. Eppe, J. Roland, and Y. De Smet, "On the use of valued action profiles for relational multi-criteria clustering," *International Journal of Multicriteria Decision Making*, vol. 4, no. 3, pp. 201–233, 2014.

[36] E. Fernandez, J. Navarro, and S. Bernal, "Handling multicriteria preferences in cluster analysis," *European Journal of Operational Research*, vol. 202, no. 3, pp. 819–827, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.ejor.2009.05.034

[37] C. Rocha, L. C. Dias, and I. Dimas, "Multicriteria Classification with Unknown Categories: A Clustering-Sorting Approach and an Application to Conflict Management," *Journal of Multi-Criteria Decision Analysis*, 2013.

[38] P. Meyer and A. L. Olteanu, "Formalizing and solving the problem of clustering in MCDA," *European Journal of Operational Research*, 2013.

[39] B. Rouba and S. Nait Bahloul, "Towards multicriteria analysis: A new clustering approach," *2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings*, pp. 126–131, 2010.

[40] B. Rouba and B. Safia Nait, "A multicriteria clustering approach based on similarity indices and clustering ensemble techniques," *International Journal of Information Technology and Decision Making*, 2014.

[41] R. Bisdorff, P. Meyer, and A. L. Olteanu, "A clustering approach using weighted similarity majority margins," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011.

[42] Y. De Smet, "P2CLUST: An extension of PROMETHEE II for multicriteria ordered clustering," in *IEEE International Conference on Industrial Engineering and Engineering Management*, 2014.

[43] L. Chen, Z. Xu, H. Wang, and S. Liu, "An ordered clustering algorithm based on K-means and the PROMETHEE method," *International Journal of Machine Learning and Cybernetics*, 2018.

[44] M. A. Boujelben and Y. De Smet, "A multicriteria ordered clustering algorithm to determine precise or disjunctive partitions," *Interna-*

tional Journal of Multicriteria Decision Making, vol. 6, no. 2, pp. 157–187, 2016.

[45] C. Bai, R. Zhang, L. Qian, L. Liu, and Y. Wu, "An ordered clustering algorithm based on fuzzy c-means and PROMETHEE," *International Journal of Machine Learning and Cybernetics*, 2019.

[46] Alain Fernandez, "La démarche GIMSI," in *Les nouveaux tableaux de bord des managers*, 5th ed. 1998: La démarche GIMSI, 1998, ch. La démarch, pp. 117–132.

[47] M. Schlossberg, C. Meyer, J. Dill, and L. Ma, "Measuring the Performance of Transit Relative to Livability," p. 66, 2013. [Online]. Available: http://dx.doi.org/10.15760/trec.135%0Ahttp://archives.pdx.edu/ds/psu/13085

[48] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967.

[49] P. Nemery and C. Lamboray, "Flow sort: A flow-based sorting method with limiting or central profiles," *TOP*, 2008.

[50] J. P. Brans and Y. De Smet, "PROMETHEE methods," *International Series in Operations Research and Management Science*, 2016.

[51] I. B. Mohamad and D. Usman, "Standardization and its effects on K-means clustering algorithm," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 6, no. 17, pp. 3299–3303, 2013.

[52] R. James, G., Witten, D., Hastie, T., Tibshirani, *An Introduction to Statistical Learning - with Applications in R — Gareth James — Springer*, 2013.

[53] K. Palczewski and W. Sałabun, "Influence of various normalization methods in PROMETHEE II: An empirical study on the selection of the airport location," *Procedia Computer Science*, vol. 159, pp. 2051–2060, 2019. [Online]. Available: https://doi.org/10.1016/j.procs.2019.09.378

Imene Soumaya TOUATI Imene Soumaya TOUATI is full professor and member of the LIO laboratory at University Oran 1, Algeria. She obtained her engineering and master degree respectively in 2009 and 2012 from computer sciences department at the same institution. Besides her research interests in artificial intelligence, decision support systems and multi-criteria analysis, she is very involved in developing skills in programming languages.

Karim BOUAMRANE Karim BOUAMRANE received the PhD Degree in computer science from the Oran University in 2006. He is full Professor of computer Science at the same university. He is member of the computer science laboratory (LIO). He is the head of the team diagnosis and decision support system . His current research interest's deal with decision support system in maritime and urban transportation system, production system, application of bio-inspired based optimization meta-heuristic, health systems. He participates in several scientific international/national committees conferences in Algeria and others countries in the same domain and collaborates in many scientific projects. He is co-author of more than 70 scientific publications and communication.
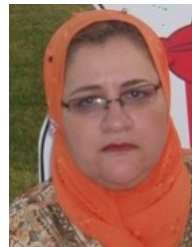
Djamila HAMDADOU Djamila Hamdadou received her Engineering degree in Computer Science and her Master of Science degree from the Computer Science Institute in 1993 and 2000, respectively. She also obtained her doctorate in 2008. She received her PHD in 2012 from the Computer Science Department. She is specialized in Artificial Intelligence, Decision Support Systems, Multi Criteria Analysis, Collaborative and Spatio Temporel Decisional Systems and Business Intelligence . She is a Professor at the University Oran 1 in Algeria where she leads the research team artificial Intelligence Tools at the service of Spatio-Temporal and Medical Decision Support at the laboratory of computer science of Oran (LIO).

APPENDIX - Performance diagnosis process

Methodology of the present study in the context of the large-scale performance evaluation.

Algorithm 1 - The overall process

Inputs - The performance criteria evaluation matrix $M = (m_{i,j}) \; \forall \, i \in 1 \, ..n, \; \forall \, j \in 1 \, ..q$ with n lines from different cities concerned with the diagnosis and q the selected set of criteria for the performance evaluation.

Outputs - A detailed diagnosis by analysing the resulting set of ordered clusters gathering similar line from various cities.

Begin

1) Pre-processing
   - *Data normalisation* : generate the normalized matrix $M_{min-max}$ by the min-max normalisation and the normalized matrix $M_{z-score}$ by the z-score normalisation (transform the cost type criteria to profit type criteria).
   - *PROMETHEE parameters* : set the weight of each criterion according to the strategic objectives and compute the preference and the indifference thresholds by applying mathematical method.
   - *k-means parameters* : Initialise the number of individual clusters ($k = 4$).
2) Clustering process
   Repeat
   - Generate the distribution of individual clusters C_h , $\forall h \in 1 \ldots k$ and the one of interval clusters $C_{h,l \in 1...k}$, $h < l$ by performing the interval clustering algorithm on the matrix $M_{min-max}$.
   - Generate the distribution of individual clusters C_h , $\forall h \in 1 \ldots k$ and the one of interval clusters $C_{h,l \in 1...k}$, $h < l$ by performing the interval clustering algorithm on the matrix $M_{z-score}$.
   - Increment the number of individual clusters

   Until a distribution of individual clusters C_h , $\forall h \in 1 \ldots k$, obtained from one of the two matrix, with $k = argmin_k \left(min \left| C_{h,l} \right| \right)$
   Sort the set of the cities concerned with the diagnosis, according to their weighted mean based on the line assignments, assuming a higher weight for the first cluster.
3) Profile analysis
   - Perform the single-criterion analysis algorithm on each individual cluster from the resulting distribution to yield the preference and similarity profiles.
   - Display the outcomes in different charts in order to facilitate the interpretation of the diagnosis.

End

The under mentioned algorithm determines the main steps of the ordered interval clustering technique used to figure out an optimal distribution

---

Algorithm 2 - Clustering process

Inputs - The set of alternatives evaluations $g_i(a_i)$, $\forall a_i \in A$, $\forall g_i \in F$ where $A = a_i$, $\forall i \in 1 \ldots n$ the set of alternatives and $F = g_j$, $\forall j \in 1 \ldots q$ the set of criteria and initial k value.

Outputs - An ordered distribution of individual clusters C_h , $\forall h \in 1 \ldots k$ and interval clusters $C_{h,l \in 1...k}$, $h < l$.

Begin

1) *Centroids initialisation* : Initialise the set of central profiles $R = r_h$, $\forall h \in 1 \ldots k$ with random values from $g_i(a_i)$, $\forall a_i \in A$, $\forall g_i \in F$.. Each central profile r_h represents an individual cluster C_h.
   Repeat

2) *Alternatives assignment* : Assign the alternatives a_i to the individual C_h or interval C_h, l clusters according to their positive $\emptyset_{R_i}^+(a_i)$ and negative $\emptyset_{R_i}^-(a_i)$ flows computed by the FLOWSORT sorting process. *Central profile updating* : Update the set of central profiles $R = r_h$, $\forall h \in 1 \ldots k$ according to the state of the individual clusters, empty or non-empty. **Until** one of the followed converge conditions is met (the distribution remains unchanged during 10 cycles or carry out the max iterations)
3) *Index quality computing* : Compute the quality index (D of the resulting distribution of individual clusters. This index assists in selecting the best value k.

End

The profile analysis is achieved by the single-criterion analysis algorithm presented bellow :

Algorithm 3 - Profile analysis

Inputs - An ordered distribution of individual clusters C_h , $\forall h \in 1 \ldots k$

Outputs - Profile of each individual cluster

Begin

1) *Preference profile computing*:

- For each individual clusters C_h , $\forall h \in 1 \ldots k$ compute the single-criterion net flow vector $PRP(C_h) = \left[\emptyset_1(C_h), \ldots, \emptyset_q(C_h) \right]$; where $\emptyset_j(C_h)$, $j \in 1, \ldots, q$ is the net flow of the cluster C_h for the criterion j.
- Compute the covariance matrix between criteria according to the net flow vectors $PRP(C_h)$, $\forall h \in 1 \ldots k$

2) *Similarity profile computing*: For each individual clusters C_h , $\forall h \in 1 \ldots k$ Compute the global similarity vector $PS(C_h) = [S_1(C_h), \ldots, S_q(C_h)]$.

3) *Profiles interpretation*: For each individual clusters C_h , $\forall h \in 1 \ldots k$

- Identify from the single-criterion net flow vector $PRP(C_h)$ the criteria with either good evaluation ($\emptyset_j(C_h) > 0$) or bad evaluation ($\emptyset_j(C_h) \leq 0$). This allows selecting the set of criteria where similar lines from the same cluster, require significant improvements.
- Detect from the covariance matrix the relation between each pair of criteria, similarity, dependency or inconsistency. These relations help to figure out the impact of an eventual improvement of a criterion on the others.
- Identify the criteria that promote the similarity of the alternatives according to the global similarity vector $PS(C_h)$. The optimization process is easier for clusters with a big set of similar criteria.

End