

CAHIER DU LAMSADE

Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision
(Université Paris-Dauphine)
Unité de Recherche Associée au CNRS n° 825

P.1824

AN "INTELLIGENT" DSS FOR THE MULTICRITERIA EVALUATION OF RAILWAY TIMETABLES ¹

CAHIER N° 120
octobre 1993

J.C. POMEROL ²
B. ROY
C. ROSENTHAL-SABROUX
A. SAAD ³

¹ Communication présentée à IFORS 93, XIII World Conference on Operations Research, Theme : OR : Expanding Horizons, Lisbonne, Portugal, 12-16 juillet 1993 (contribution nationale).

² Université Paris 6, LAFORIA-IBP, B.P. 169, 75252 Paris Cedex 05.

³ STRATEMS, 20 rue de Mogador, 75009 Paris

CONTENTS

	<u>Pages</u>
Résumé	i
Abstract	i
1. Evaluating timetables	1
2. The design of the system	2
2.1 Various tasks	2
2.2 Architecture of the system	2
2.3 The role of expert systems in multicriteria analysis	4
3. The benchmark of incidents	5
4. The expert system module	7
4.1 Knowledge structuring and automatic decision-making	7
4.2 Analysis of the perturbed period	9
5. The criteria	13
6. Results	16
7. Conclusion	18
Acknowledgement	19
References	20

UN SIAD "INTELLIGENT" POUR L'ÉVALUATION MULTICRITÈRE DES GRILLES HORAIRES FERROVIAIRES

Résumé

Etant donné un réseau ferré, un matériel roulant et une grille horaire (RRT), l'évaluation de la perte de qualité du service offert aux voyageurs en cas d'incidents est une tâche difficile. Nous avons conçu un système d'évaluation de la robustesse (SRE) pour évaluer les effets de ces incidents. Un RRT étant considéré, ce système utilise un système expert pour simuler les décisions des régulateurs en cas d'incident. Une fois les incidents résorbés et le trafic redevenu conforme à la grille horaire, les conséquences de la perturbation sont analysées selon cinq critères. Ce système est maintenant opérationnel. Il se présente comme un SIAD multicritère intégré qui constitue un outil très utile, aussi bien pour apprécier l'intérêt d'investissements améliorant le réseau, de modifications du matériel roulant que pour guider le choix d'une grille horaire.

AN "INTELLIGENT" DSS FOR THE MULTICRITERIA EVALUATION OF RAILWAY TIMETABLES

Abstract

Given a Railway network, a Rolling stock and a Timetable (RRT), it is a very difficult task to ex ante evaluate the loss of quality of the service offered to travelers in case of incidents. We have designed a System for Robustness Evaluation (SRE) to measure the effects of incidents. This system starts from the RRT then, using an expert system the decisions of the dispatchers in case of incident are simulated. After the incidents are resolved, the resulting perturbed timetable is analysed along five criteria. This system is now completed. It appears as an integrated multicriteria decision support system which constitutes a very useful tool, both to appreciate the interest of investments intended to improve the network, of amendmends of rolling stocks, and to guide the choice of a timetable.

1. EVALUATING TIMETABLES

On French Railways, the timetables are prepared by "schedulers" using a DSS called CHAO. This system is described in Bertereix and Beurrier (1990), see also Lévine and Pomerol (1989a, p. 140). A timetable is related to a given railway network and rolling stock (infrastructure, rolling stock and cars) and for a given type of day (e.g. working days or Sundays) during a given period (November to March for instance). For each day of a given type, during the considered period, as long as no incident occurs, trains arrive and leave each station exactly as prescribed by the timetable. Typically, in such a timetable, the trains follow each other very closely, specially at rush hours and any trouble on one train has consequences on the following ones.

Two timetables related to the same railway network and involving the same number of trains differ according to how much they overdevelop or underdevelop the delays due to an original incident.

Many possible outcomes, affecting the company and/or the users, have to be taken into consideration to appreciate the impacts of incidents which inevitably occur during operations. We will measure the behavior of timetables in the case of incidents by the robustness of the timetables. Such a measurement necessarily refers to all the outcomes that seem relevant for comparing the reactions of different timetables to identical incidents. This concept of robustness does not characterize the timetable alone, but the 3-uple timetable, railway network and rolling stock that we will denote RRT in what follows.

Comparing the robustness of different 3-uples RRT is particularly important for evaluating the advantages of different infrastructure investments. For example, the realization of a new siding, or laying a new track parallel to an existing one are firstly justified by enabling timetables on the modified network which are more incident resistant than was possible before the network modification. In other words, with the new RRT, it is hoped that the incidents entail less serious perturbations than with the older one. In this case, we will say that the new 3-uple timetable, railway network and rolling stock is *robuster* than the old one (possibly only one of the terms T or R is modified).

The purpose of our system is (i) to characterize this complex notion of robustness by means of an appropriate family of criteria, and (ii) to evaluate, for a given railway network, any feasible timetable so as to be able to compare the robustness of various timetables related to a given railway network, possibly with variants. Even if two 3-uples RRT are equivalent for the users when no incident occurs, they may react very differently when some type of troubles happens ; the DSS we propose is aimed at evaluating this difference.

2. THE DESIGN OF THE SYSTEM

2.1 Various tasks

While the problem is relatively simple to state : assess the robustness of a timetable with respect to a given 3-uple RRT, it seems, at least to our knowledge, that no decision support system has ever been designed to tackle such a problem.

To assess the robustness, we have therefore been obliged to define :

- a) what are the usual incidents to be considered and how they can be characterized,
- b) how to propagate realistically the perturbations due to the incidents,
- c) what family of criteria would be appropriate for characterizing the notion of robustness.

Accordingly, our task was divided into three parts. A first subtask was concerned with the incidents. We arrived at a typology and a benchmark (see Section 3). The second subtask addresses the problem of the propagation of the incidents. The supervision of the trains is presently performed by persons, called dispatchers. These people obviously react when a real incident occurs ; they make decisions such as : delay or stop a train, have one train overtake another (if the infrastructure allows it), switch a part of the traffic toward another railway network. As our only way of tackling the problem is through the dispatcher's expertise, and because of the large number of possible situations, we have chosen not to conceive a mathematical model, but to design an expert system able to mimic, as far as possible, the dispatcher's reasonings. This subsystem is described in Section 4.

For the last subtask, the building of the criteria (Section 5), we thought that, on the one hand, it would be arbitrary to privilege one single criteria. In fact, we identified : number of perturbed trains, number of people involved in the delays, duration of the overall troubles, sum of the delays (weighted or not with the number of people involved). On the other hand, the aggregation of such a disparate set of criteria raises so many questions that it was decided, preliminarily, to study each of them separately ; this does not rule out any further aggregative decision process (for a further justification of this way of tackling the problem, see Bouyssou (1989) and Roy (1990)).

2.2 Architecture of the system

The reader will recognize in the architecture of our system the different tasks previously described. The system, called SRE (for System for Robustness Evaluation), starts from a given railway network, a timetable and an incident. The RR and the T are two inputs. The incident is introduced via incident editor which is able to define and introduce into the system either one incident or a predefined sequence of incidents.

The expert system "dispatcher" then makes a decision according to the type of incident and various types of contextual information (e.g. railroad possibilities, time of

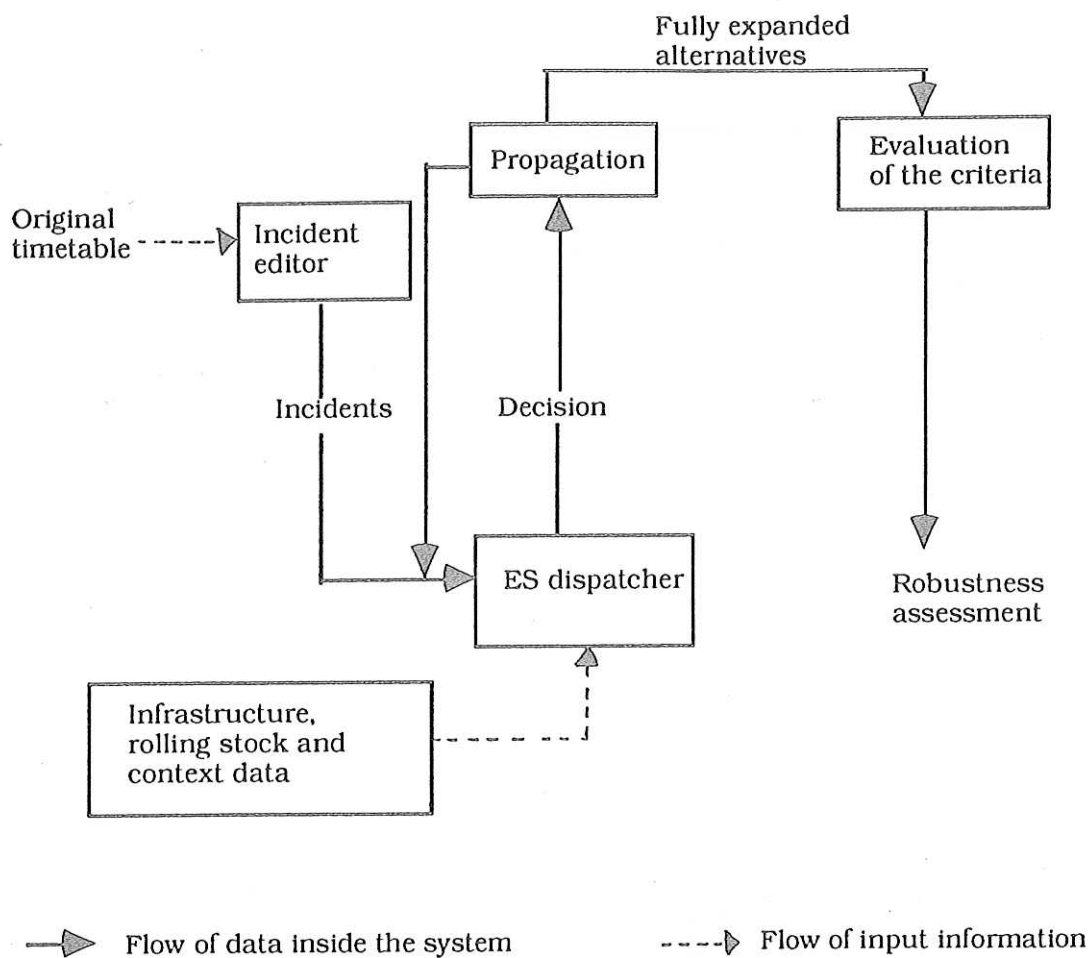


FIGURE 1 : Architecture of the system

operation, number of people in the train). The decision entails some new perturbations in the schedule which in turn produces new incidents on other trains and further decisions made by the ES. Thus, we enter a loop (Figure 1). When all the consequences of the incident under study are completed, i.e. when the trains recover their original schedules in the timetable under consideration, the system runs out of the loop. The point of exit from the loop defines the end of the perturbed period generated by the original incident. During this perturbed time, the schedule of some trains (possibly all the trains) does not conform to the theoretical timetable. These train schedules during the perturbed period are, for the incident under study, the output of the knowledge-based system, depending on the decision it has made. This output is the input for the multicriteria evaluation (Section 5). Thus, each criterion is intended to provide an evaluation of the perturbations to train schedules during the perturbed period. This set of train schedules post incident, denoted perturbed timetable, related to a given 3-uple RRT, may therefore be regarded, from a multicriteria point of view, as a part of the alternative to be evaluated. Actually, in order to get a realistic assessment of the robustness, it is necessary to consider a set of various representative incidents that are likely to appear during the daily operations in a given network. Thus, the multicriteria analysis has to cope with many perturbed timetables. According to Figure 2, an alternative $a = \text{RRT}$ generates, for each incident i or each sequence of incidents, via the expert system, a perturbed timetable defined as an expanded alternative a_i .

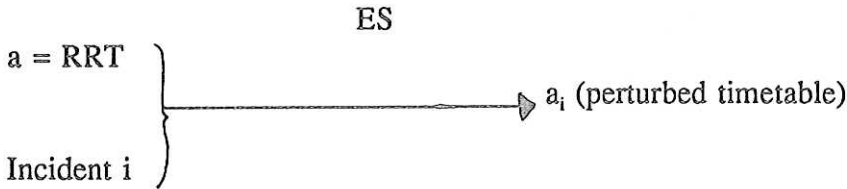


FIGURE 2 : From the given timetable to the perturbed timetable

For a finite set of incidents, called the benchmark, we therefore get a vector $\hat{a} = (a_1, a_2, \dots, a_i)$ which is the *fully expanded alternative* (FEA). It is this fully expanded alternative which will be evaluated according to each criterion.

2.3 The role of expert systems in multicriteria analysis

It is now a common idea to aggregate criteria by rules (see Pomerol (1992) for a review) and various systems have already been produced in this spirit (e.g. Bohanec et al. (1983, 1987), Efstathiou et al. (1986), O'Leary (1986), Rajkovic et al. (1988)). On the contrary, in our system, we intended to maintain the role of the multicriteria analysis by its own methods (see Bana e Costa (1990), ch. 1 and ch. 3). So the system does not trespass on the multicriteria analysis ; it remains completely separate. Nor is the ES used to value the alternatives (or to assess their performance) according to the criteria, as is the case in Lévine et al. (1990). In our system, the evaluation of the fully expanded alternatives according to the criteria is carried out by an adhoc module.

We have seen that the ES produces the new schedules resulting from the incidents. These schedules are the components of the fully expanded alternatives as defined above. Thus, the ES may be viewed in this case as an aid to building a decisive part of the alternatives, from a basic set of data : original timetable (T), infrastructure and type of train (RRT). The ES introduces the decision component into the alternatives. It simulates the reaction of the dispatchers in order to calculate the cascade of consequences of the incidents.

As such, our system provides many fully expanded alternatives according to different original alternatives, in particular railroad possibilities or different initial timetables. Fixing the timetable and changing the railroad network produce different fully expanded alternatives that may be evaluated. The DSS may therefore be used to choose between different infrastructure investments.

Thus, the role of the ES in SRE is to build automatically the expanded alternatives. The only paper we are aware of, using a similar approach, is that of Du Bois et al. (1989) in which a frame based expert system is used to deduce several possible diagnoses relatively to a patient. These diagnoses are then introduced into the PROMETHEE multicriteria software in order to make the decision.

3. THE BENCHMARK OF INCIDENTS

The French railways provided us with a list of 93 types of incidents and their corresponding yearly probability of occurrence. This was the starting point of an analysis, the aim of which was the elaboration of a benchmark of incidents. By benchmark, we mean a set of representative incidents, each of them being precisely described with a weight assigned according to its frequency. The benchmark must be drawn up in such a way that the evaluation of robustness (with respect to each criterion, see section 5) can be done by referring to the only incidents of the set supposed to be representative of what can happen according to their weights. For this purpose, we have been led to define three types of incidents : the isolated incident, the zone incident and the train incident. Each one is briefly described below together with the way it contributes to the benchmark.

a) The isolated incident (type a) is an incident at a precise location affecting a unique train (for example a suicide on the railway or an emergency signal on the train). This type of incident is characterized by :

- a delay for the train ;
- a kilometric point on the network.

This type of incident is the most frequent and is subdivided, for the purpose of the benchmark, into six sub-types characterized by the location (station or railway) and the

duration (less than 5 minutes, between 5 minutes and 15 minutes, more than 15 minutes).

b) The zone incident (type b) is a repeated incident on the railway corresponding to a railway failure (for example a signal breakdown) which requires all the trains to circulate more slowly at the breakdown point. This type of incident is characterized by :

- a kilometric point ;
- a loss of time due to the slowdown ;
- a duration of the failure.

All the trains which pass by the kilometric point characterizing the incident lose the indicated time as long as the amendment has not been carried out (specified by the duration of the failure). This type of incident is the least frequent but is very important. It has again been decomposed into six sub-types characterized by two levels of duration :

- the loss of time at the kilometric point of incident (less than 10 minutes, between 10 minutes and 30 minutes, more than 30 minutes) ;
- the duration of the failure (less or more than one hour).

c) The train incident (type c) is provoked by a car or motor unit failure (for example a faulty closing of the doors) causing a constant delay of this train at each stop during its whole journey. This third type is characterized only by the delay induced at each occurrence of the incident. Up to now, we have considered that this type of incident only occurs in stations. We have discretized the possible delay into three subclasses (less than 2 minutes, between 2 and 15 minutes, more than 15 minutes).

The above typology and the set of incidents derived from it for building the benchmark, result from many discussions and classical interview process ; they have been finally accepted by the dispatchers and the representatives of the French Railway authority.

These three types of incidents have very different characteristics entailing different decisions. We have also, for each type and sub-type of incidents, a statistical analysis established by the French Railways regarding the delay induced from the different types of incidents and their frequencies. On such a basis, it has been possible to assign a weight to each sub-type of incident.

In order to choose the appropriate kilometric points for each sub-type of incidents (more than one for each sub-type), the experience of the dispatchers has been required (they know where the sensitive points are). Finally, to complete the definition of the benchmark, information like the time of each incident (rush hour or not), the weather (snow, rain, etc.) has been taken into consideration.

An incident editor has been designed in order to give the possibility, to the user, of injecting, into the system the incident he wishes. It allows the introduction of a sequence of incidents with their weights such as the ones constituting the benchmark.

4. THE EXPERT SYSTEM MODULE

4.1 Knowledge structuring and automatic decision-making

When an incident occurs, one or many trains are delayed. Assume that train A is delayed ; then it is easy to check that, at time t_0 and kilometer k_0 , a conflict will appear with train B behind it, which also passes at point k_0 at time t_0 because it is on time. The main task of the dispatcher is therefore to make a decision in order to avoid this predicted conflict. This decision forms the ultimate conclusion of the reasoning process in this type of problem (see also Komaya and Fukuda, 1988 and 1989 who arrive at the same idea in their model). Therefore, the knowledge base is structured around the possible decisions. They are not very numerous, the more frequent being : stop train A in a station, where a siding exists, before or after the conflict point ; if a parallel railroad exists, make train B overtake train A ; otherwise, let train B follow train A.

The knowledge that is necessary to make such a decision and solve the conflicts includes knowledge about the infrastructure, the importance and type of train (express, commuter), the time (rush hour or not) and some other concepts.

These concepts are organized in a semantic tree (see figure 3) of schemas according to the methodology accompanying the shell ARGUMENT (see Lévine and Pomerol, 1990, for another example). The knowledge has been collected by a classical interview process with the dispatchers. Once the tree is consistent with the gathered knowledge, the values related to each concept and the rules making the deduction are written, always in accordance with the interviewed dispatchers. In ARGUMENT, the rules are closely tied to the schema which facilitates a kind of consistency maintenance (see Lévine and Pomerol, 1989b, for more details about the schema representation).

The distance of a concept from the root in the tree gives a good indication of its importance. The reader can observe on Figure 3 that, among these important ideas, are the comparison between the relative importance of the two trains involved in the conflict, the delay, the type of the incident and the reuse of the train after its arrival at the terminus. The reasoning process, paying no attention to the feasibility, orders the decisions from the most desirable to that which seems the worst. Then, according to the infrastructure and the existing possibilities in the stations, the system eliminates the unfeasible decisions and produces an ordered set of feasible decisions. It is the best of these decisions which is implemented first.

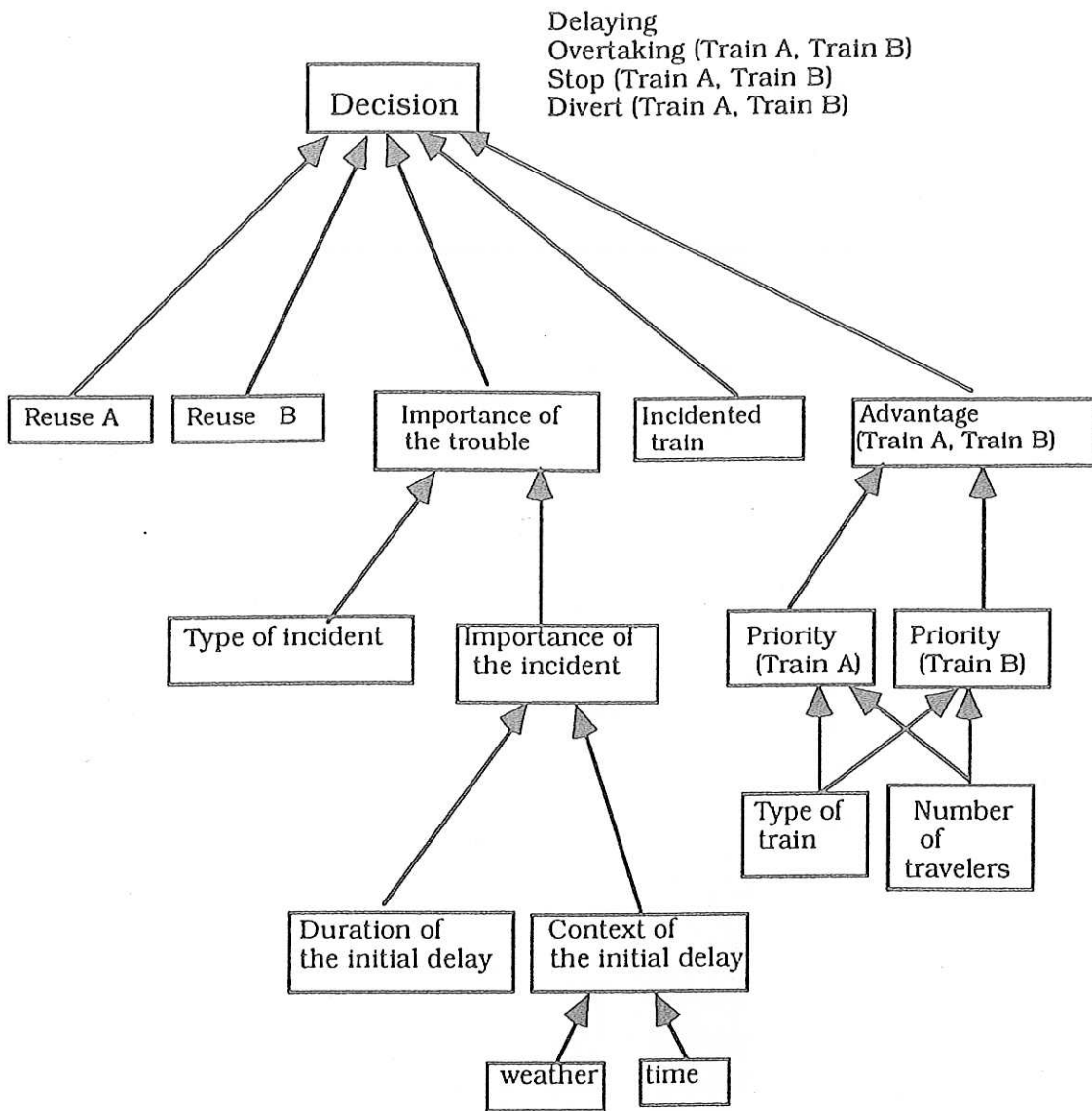


FIGURE 3 : Schema tree of the relevant concepts in the E.S.

4.2 Analysis of the perturbed period

When a decision is made to resolve a conflict, it is implemented by the scheduling system CHAO. Generally, this implementation (e.g. stop train for a while in a station) produces new conflicts. We enter into the loop. The decisions are made in order to minimize the perturbations, specially those related to the important trains, and to retrieve the original schedule as quickly as possible. The moment where all the conflicts are solved, i.e. when the original schedules come back into operation, defines the end of the perturbed period. So the system produces the perturbed timetable in which one can read the new schedule of each train.

With these more or less new delayed schedules, all the conflicts have disappeared. This process is always possible because, at the end of the day, the trains are sufficiently spaced that, even with large delay, no new conflict appears.

Let us illustrate these ideas. Figure 4 shows the representation used by the schedulers and the system CHAO to display a timetable. Time is on the horizontal axis while the kilometers and the stations appear on the vertical one. On Figure 4, a train L3 is stopped at 5.30 p.m. (17h30) for fifteen minutes at Station SUO0. This generates conflicts (indicated by an arrow) when it restarts at 5.45 p.m. In the windows of Figure 5, one can see the decisions made in each loop of the algorithm ("passage", in French). The most important decisions in this case are to transfer the trains V3, S1 and L5 onto another line between PUBV and SCBP and slightly to delay L5. This produces a new conflict visible in Figure 5 which will be solved by slightly delaying 05. So all the conflicts are resolved and we get the perturbed timetable.

But it does not suffice to resolve all the conflicts ; there remains one problem : even if we have chosen the best decision for each conflict, it may happen that the overall solution is bad. For example, taking a good myopic decision about a delayed commuter train may result, a little while and several conflicts later, in delaying an express by half an hour, which is unacceptable.

To avoid this phenomenon, we have implemented two ideas. Firstly, we do not resolve the conflict following the time order but following a priority order. We have defined a simple evaluation function which, according to the kilometer of conflict, the hour, the type and importance of the train, computes a priority level. Thus, at a given step, the system processes what seems to be the more crucial conflict before the others.

The second idea is to estimate the overall quality of a partial solution before the resolution of all the conflicts and to backtrack on already-made decisions if the partial solution is becoming too bad. For example, the resolution of a conflict relative to a commuter train entails the decision of letting the train continue ; it produces after a while a major conflict with an express which results in stopping the express a quarter of an hour in a country station. In this case, the system estimates that the uncompleted solution becomes unacceptable and it backtracks to the closer preceding conflict where it has left a non-implemented decision with the highest priority (among the remaining

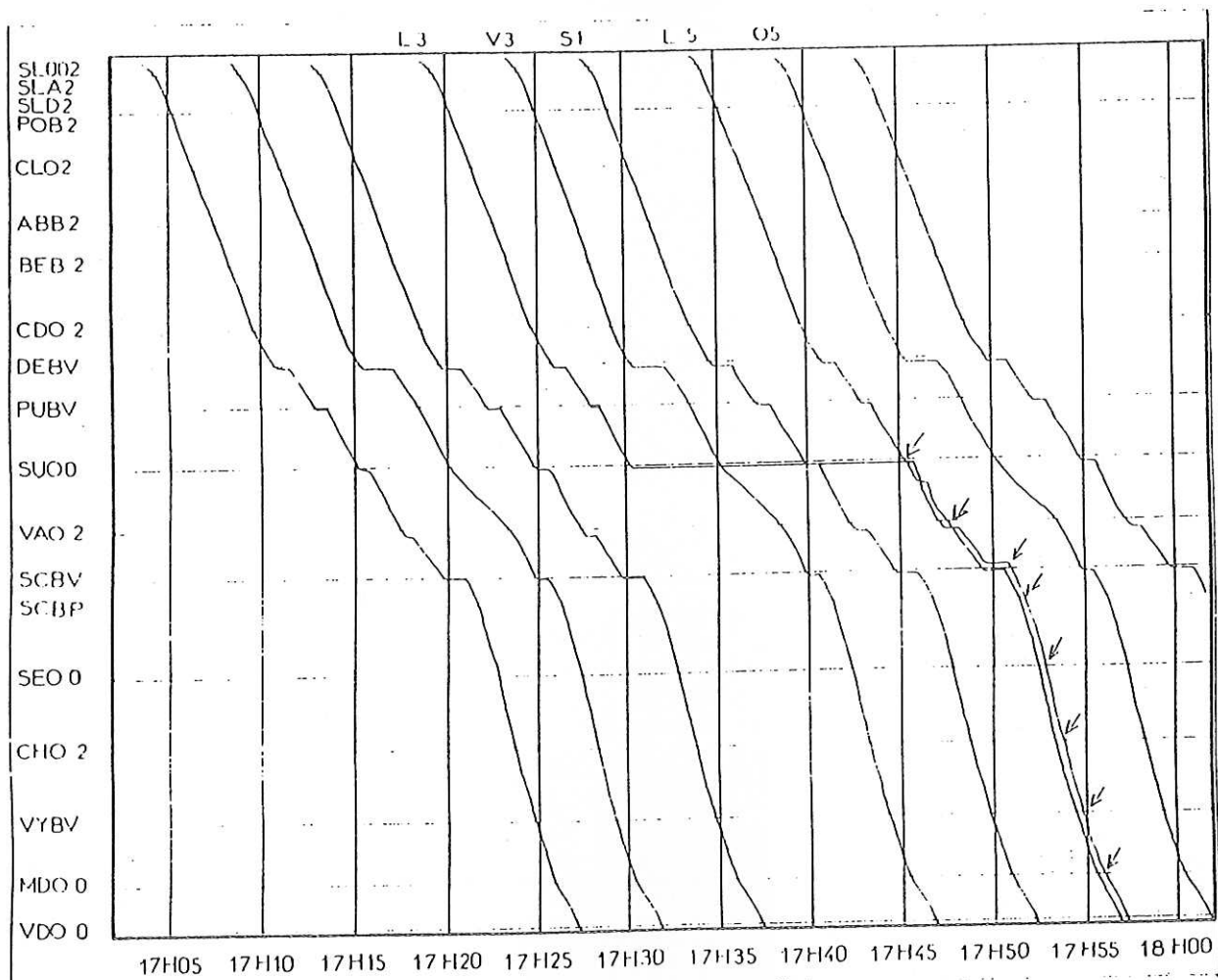
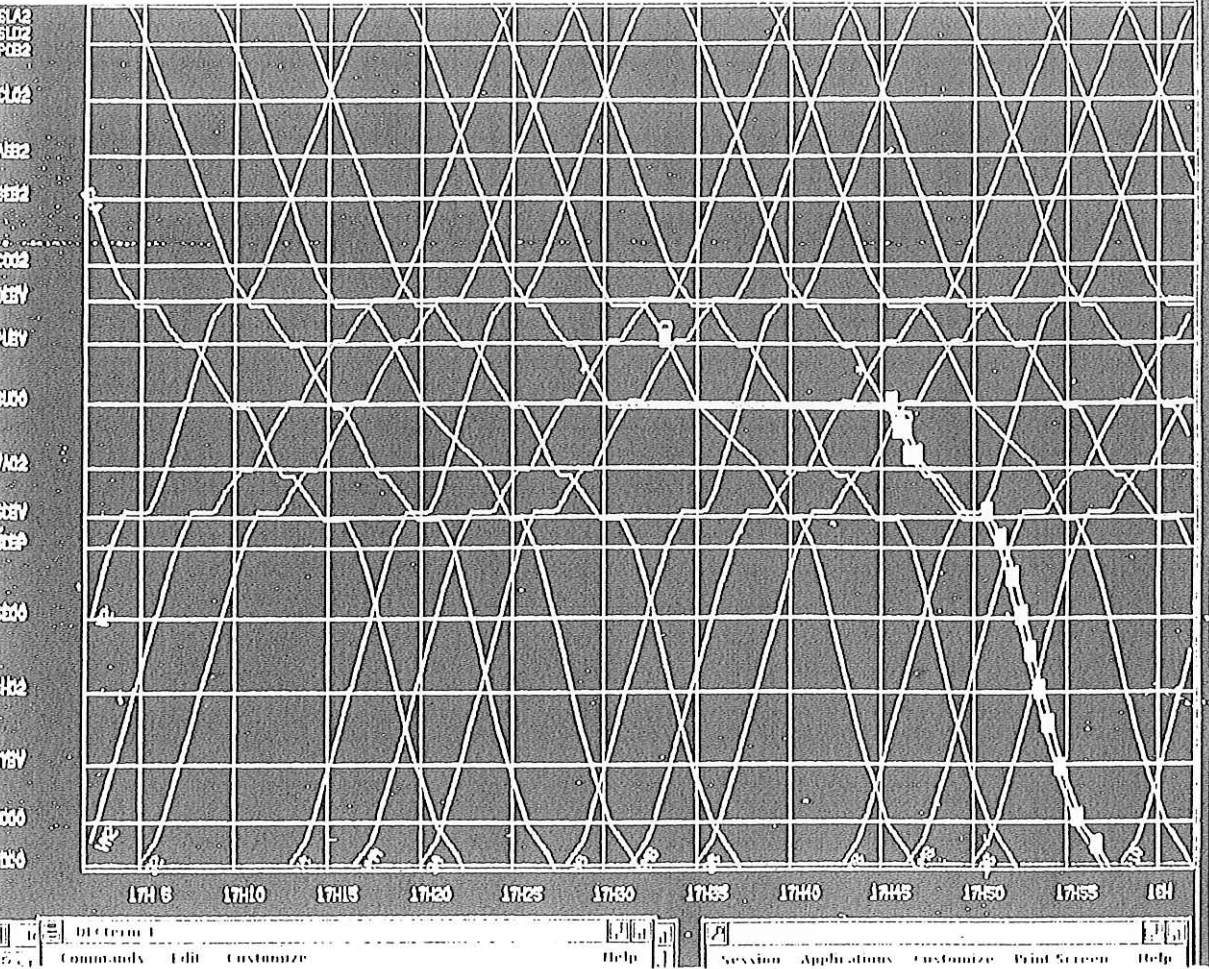


FIGURE 4 : Timetable after the incident

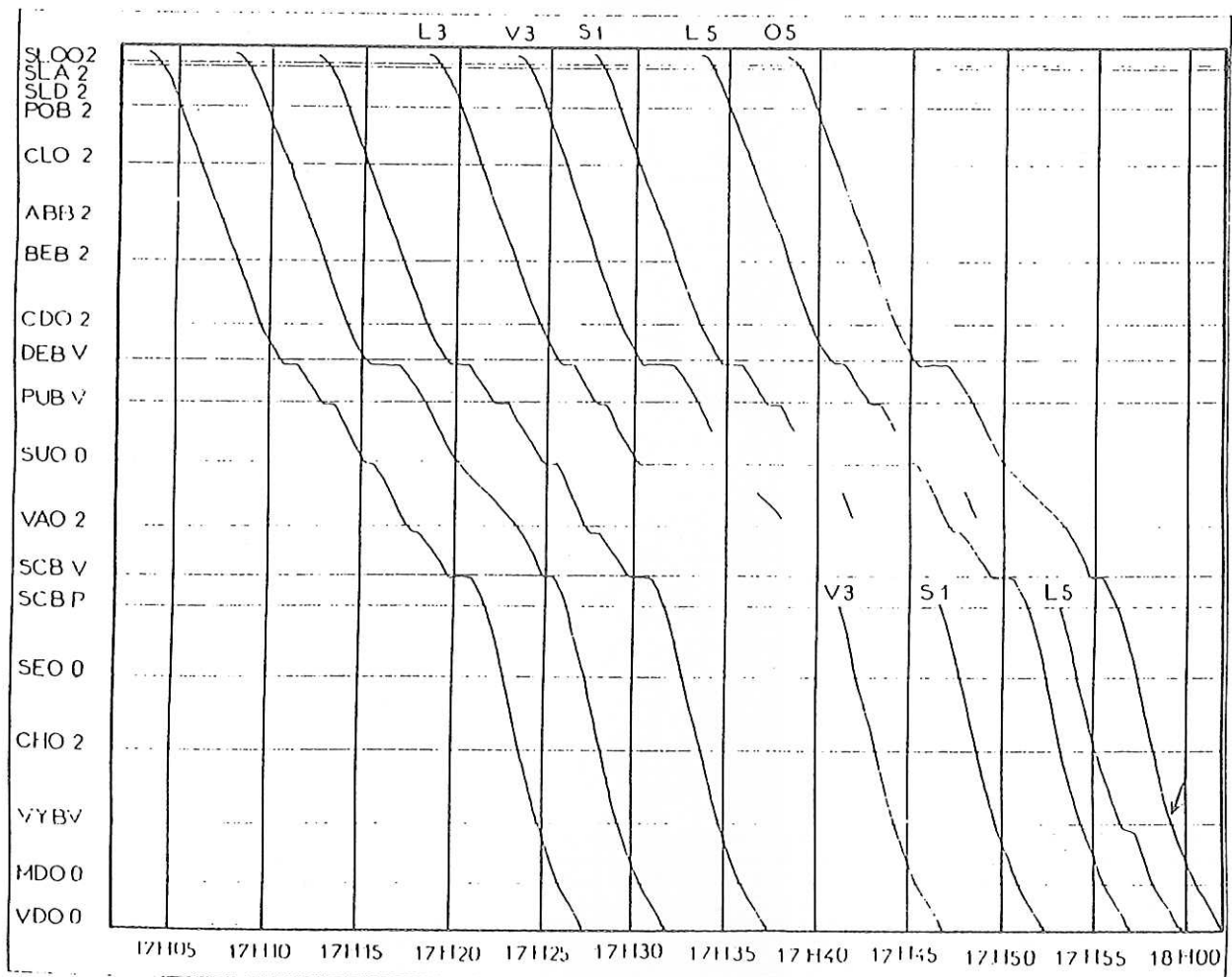
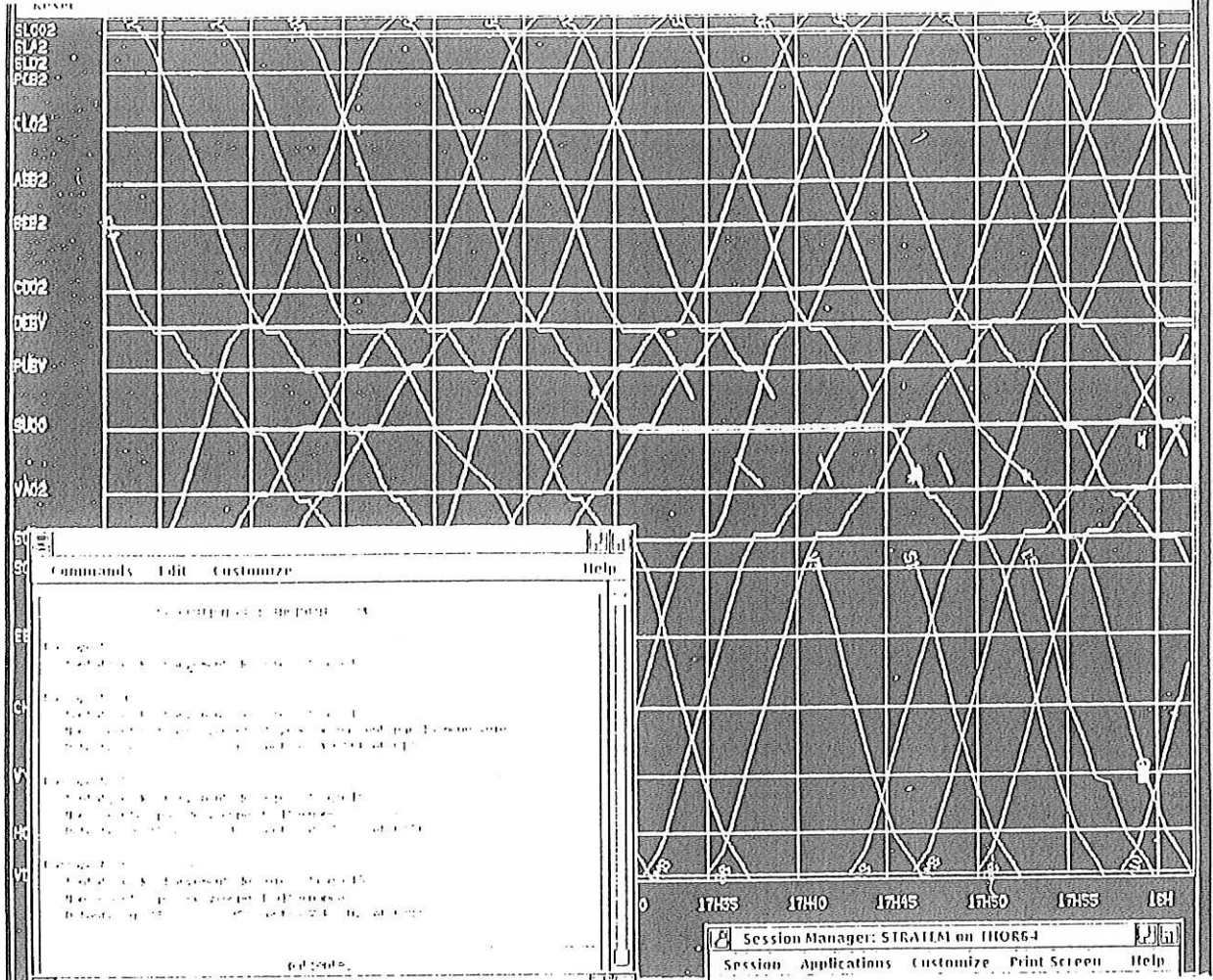


FIGURE 5 : Timetable after partial rescheduling

Type of incident	Weight of incident	Number of perturbed trains	Duration of the perturbation	Delayed travelers	Average delay	Cancelled trains
det vm1	8.33	5	31m08s	3525	02m05s	0
s1 rep	9.23	6	31m15s	2650	03m07s	0
s1 rep2	6.92	3	18m22s	1175	02m43s	0
s1 sta su	5.00	6	26m06s	1700	04m17s	0
v01 det	8.33	4	20m16s	1625	02m30s	0
v01 sta def	10.00	8	37m34s	5925	06m17s	0
vm1 rep	13.85	5	34m46s	1875	03m59s	0
vm1 sta su	8.33	3	23m58s	850	08m22s	0
vm7 sta def	10.00	6	29m01s	4500	04m16s	0
zone 1	7.06	14	01h20m	10725	01m45s	0
zone 2	7.06	5	35m35s	3000	01m45s	0
zone 3	5.88	8	50m35s	5925	02m17s	0

- CRIT. 1 : Number of perturbed trains
- CRIT. 2 : Duration of the perturbation
- CRIT. 3 : Number of delayed travelers
- CRIT. 4 : Average delay per traveler
- CRIT. 5 : Number of cancelled trains

RESULTS	CRIT.1	CRIT.2	CRIT.3	CRIT.4	CRIT.5
	7.022868	34m23s	3530	04m48s	0

FIGURE 6 : Results of the multicriteria analysis

neglected decisions). In our case, the neglected decision might have been to stop the commuter train 10 minutes in a station where it should have been overtaken by the express.

The system stops when it has completed a "satisficing" (in Simon's sense) decision. As such, the system mimics quite well the reasoning of the dispatchers, who always try to avoid subsequent "hard" conflicts with express trains while they attempt to solve the light conflicts. The system does not anticipate the "hard" conflicts but it recognizes them when they occur and it backtracks looking for a new solution. As such, the problem is very similar to a planning process with interacting goals (Nilsson, 1980). The adopted solution is, on the one hand, hierarchical by solving the high level conflict before the others ; on the other hand, it follows a backward search graph strategy relying on an evaluation of the path. Thus, the solution may be regarded as a variation of ABSTRIPS (Nilsson, 1980).

5. THE CRITERIA

The objective of this third subtask is to build a family of criteria taking into account all the outcomes pertinent for comparing two given alternatives (RRT with respect to the complex notion of robustness introduced in Section 1). For this purpose, two different viewpoints have been considered : that of the operators in charge of the circulation of trains and that of the users. We present below the final result of an analysis carried out in collaboration with different representatives of the French Railway authority, within the framework of the methodology described in Roy (1985, chapters 8, 9, 10).

For evaluating the alternative robustness, we finally retained six criteria denoted g_1 , g_2 , g_3 , g_4 , g_5 , g_0 . Each one respectively deals with :

- g_1 : the total number of trains included in the timetable concerned by a delay (from the original incident to the return to the theoretical schedule).
- g_2 : the total duration of the perturbation (the perturbed period is defined in 2.2).
- g_3 : the total number of travellers concerned by the perturbation.
- g_4 : the average delay of the travelling time.
- g_5 : the total number of cancelled trains.
- g_0 : the maximum delay allowed to any train without any perturbation being caused.

Let us draw attention to the fact that criterion g_0 , contrary to the others, does not depend on what happened during the different perturbed periods considered in connection with the various incidents of the benchmark. Consequently, the ES does not play any role for evaluating an alternative according to this criterion. In what follows, we will focus our attention on the other five criteria.

To define the criterion g_j ($j = 1, \dots, 5$), we built a computational rule which associates, to each fully expanded alternative, a numerical performance $g_j(\hat{a})$ reflecting the general meaning indicated above. For this, two types of item must be specified :

- Considering any single incident i (or a sequence), how can we define the contribution $g_j^i(a_i)$ of the incident i ¹ to the overall performance $g_j(\hat{a})$?
- Considering the set of all the incidents of the benchmark, how can we aggregate the components $g_j^i(a_i)$ for defining $g_j(\hat{a})$?

We will now provide answers to these two questions for each of the five criteria.

Criterion g_1 : $g_1^i(a_i)$ is the number of trains which have been delayed or cancelled after incident i . This number is easy to obtain by comparing the original timetable to the schedule produced by the ES during the perturbed period. Let us denote by w_i the weight (see section 3) assigned to the incident i within the framework of a specified benchmark B ($\sum_{i \in B} w_i = 1$). The average :

$$\sum_{i \in B} w_i g_1^i(a_i)$$

did not seem appropriate for defining the performance $g_1(\hat{a})$. In fact, the same average can be obtained from a set of components g_1^i ($i \in B$), all very close to the average and from another set of components all very different to this average. The first case was preferred to the second one. To take into consideration for defining $g_1(\hat{a})$, only the worst component, i.e. $\max_{i \in B} g_1^i(a_i)$, was not judged more satisfactory than the average because it causes the disappearance of all information on the probability of occurrence of each incident. We looked for an intermediate solution² between those two extremes and we decided to adopt the following :

$$g_1(\hat{a}) = \frac{2}{1 + r} \sum_{k=1}^r w_k' g_1^k(a_k) \quad \text{with } r = |B|$$

and $g_1^k(a) = k^{\text{th}}$ figure of the sequence obtained by ranking the $g_1^i(a)$ from the smallest to the highest. For example, given the incidents $i = 1, \dots, 4$ with weights $w_1 = 0.2$, $w_2 = 0.3$, $w_3 = 0.4$, $w_4 = 0.1$ and values $g_1^1(a) = 10$, $g_1^2(a) = 5$, $g_1^3(a) = 7$ and $g_1^4(a) = 2$, we

¹ that is to say the impact of the component a_i of $\hat{a}$ on the evaluation of the fully expanded alternative $\hat{a}$ (see § 2.2 in fine).

² This, nevertheless, supposes that the difference $[(\max_{i \in B} w_i) - 1/r]$ is sufficiently small.

get $g_1^1(a) = 2$, $g_1^2(a) = 5$, $g_1^3(a) = 7$ and $g_1^4(a) = 10$ with weights $w_1' = 0.1$, $w_2' = 0.3$, $w_3' = 0.4$ and $w_4' = 0.2$.

Criterion g_2 : $g_2^i(a)$ is merely the duration of the perturbed period due to the incident i . For this criterion, it has been judged appropriate to let :

$$g_2(\hat{a}) = \sum_{i \in B} w_i g_2^i(a_i).$$

Criterion g_3 : $g_3^i(a)$ is the total number of s delayed due to the incident i ¹. The computation of this number requires statistical data giving the number of travelers boarding and leaving the train at each station according to the type of train (destination, express or commuter, etc.), to the day and the hour, etc. Here, as for criterion g_2 , the average has been adopted for aggregating the $g_3^i(a)$'s.

Criterion g_4 : $g_4^i(a)$ is the average delay of s for whom the travelling time has been perturbed. The delay of each train arriving at each station can be computed on the basis of the scheduling during the perturbed period. $g_4^i(a)$ is the average of such a delay weighted by the number of s leaving the train at the station considered and divided by the sum of these numbers, i.e. $g_3^i(a)$. $g_4(a)$ is deduced from the $g_4^i(a)$'s by the same formula (and for the same reason) as the one used for criterion g_1 .

Criterion g_5 : $g_5^i(a)$ is the number of trains which have been cancelled² following the incident i . The cancellation of a train can be the inevitable consequence of a failure corresponding to the incident i considered or the result of a decision made by the dispatcher to solve a conflict (see Section 4). Only the trains cancelled for the second reason are taken into account by $g_5^i(a)$. Here, the aggregation is made by the same formula as the one used for criterion g_2 .

Criteria g_1 , g_2 (and g_0) reflect essentially the viewpoint of the operators in charge of the circulation of trains while criteria g_3 , g_4 , g_5 are directly related to the satisfaction of the travelers. It is noteworthy that, at least for commuters, punctuality appears the most desirable quality of the transportation system (Felici and Negri, 1992). So the travelers' criteria must measure the possible lack of punctuality in case of incident. This is the case of criterion g_3 . Observe that propagating simulated incidents seems to be the only way to get an a priori measure of the lack of punctuality. Up to now, all the known measures seem to be a posteriori ones. The easiest way for minimizing g_1 and g_2 consists in cancelling all delayed trains (by doing this, the operators avoid the perturbation snowballing) but this is the worst solution for the travelers, a phenomenon

¹ For simplicity's sake, we do not take into account the phenomenon of the users changing from one delayed train to another.

² The cancellation of trains introduces some complications in the computation of $g_3^i(a)$ and $g_4^i(a)$, not considered here.

explicitly taken into account by g_5 . More details on the concrete significance and the pertinence of each criterion can be found in Sranon (1990).

Discussions and practical considerations have led us to verify that two alternatives, which give to each criterion the same performance, can really be viewed as equally robust. Moreover, such a property disappears as soon as one of the criteria is withdrawn. Such tests are important (see Roy and Bouyssou (1993)) for establishing the consistency of the family of criteria.

6. RESULTS

The system was experimented during the Summer 1992. It is implemented on a VAX computer under the operating system VMS and was written in C language. SRE is connected to ARGUMENT as an Expert System Shell and to CHAO which calculates the schedules according to the rail data. SRE is used to study the new timetables and investment projects. As was planned, it is not intended to be used to replace the dispatchers during real time operations, although its speed would authorize this use. As such, SRE produces very appealing outputs which show the variability of the reactions of different timetables to the same benchmark of incidents and their sensitivity to the railway layout.

As an example, let us give an output of the system. For each type of incident (see Section 3), the analyst creates a certain number of incidents of weight w_i . These weights appear in the third column of Figure 6. This figure displays all the data relative to a benchmark.

Figure 7 shows the result of the same benchmark as on Figure 6 on a railway network with three railways between the two stations of "Courbevoie" (i.e. CDO2 on Figure 4) and "Puteaux" (i.e. PUBV on Figure 4). This third railway is a fictitious one. In reality, the infrastructure has only two lines taken into account in Figure 6. The criteria clearly indicate that, with the extra railway, the situation is better for the travelers. In fact, the number of people concerned with incidents decreases from 3.500 to 2.000, but the average delay slightly increases, which indicates that the remaining perturbed travelers suffer a relatively long delay.

At this stage, the system appeared to be very valuable for the analysis of the choice between complex alternatives (timetable, network and rolling stock) we were confronted with. However, the values obtained for each criterion must not be regarded as absolutely exact. Let us draw attention to the fact that the notion of accuracy is not really pertinent here. The question is not mainly a matter of imprecision of certain data but more essentially a problem of uncertainty and ill-determination (see Roy, 1989).

	Type of incident	Weight of incident	Number of perturbed trains	Duration of the perturbation	Delayed travelers	Average delay	Cancelled trains
det vm1	a	8.33	2	24m02s	1325	04m33s	0
sl rep	c	9.23	6	31m15s	2350	03m29s	0
sl rep2	c	6.92	3	18m22s	1175	02m43s	0
sl sta su	a	5.00	6	26m06s	1700	04m17s	0
v01 det	a	8.33	4	20m16s	1325	02m54s	0
v01 sta def	a	10.00	2	20m23s	300	09m07s	0
vm1 rep	c	13.85	5	34m46s	1875	03m59s	0
vm1 sta su	a	8.33	3	23m58s	850	08m22s	0
vm7 sta def	a	10.00	4	26m55s	1850	03m55s	0
zone 1	b	7.06	8	01h03m	5700	01m27s	0
zone 2	b	7.06	5	35m35s	3000	01m16s	0
zone 3	b	5.88	8	51m25s	5625	02m16s	0

- CRIT. 1 : Number of perturbed trains
- CRIT. 2 : Duration of the perturbation
- CRIT. 3 : Number of delayed travelers
- CRIT. 4 : Average delay per traveler
- CRIT. 5 : Number of cancelled trains

	CRIT.1	CRIT.2	CRIT.3	CRIT.4	CRIT.5
RESULTS	5.279162	30m42s	2094	05m32s	0

FIGURE 7 : Results of the multicriteria analysis (modified network)

We have of course made some approximations about the number of travelers, and the decisions. It is also possible that, in some cases, the dispatchers would have made a better decision than SRE ; this does not disqualify our system, since it reacts equally for each alternative the comparisons are meaningful : this was our fundamental objective. After various adjustments and after completing the layout and rolling stock data files, the system is now fully operational for the choice of the investments and of the timetables at the central station "Saint-Lazare". According to the users, the system has proved to be valuable in some important cases. For example, it saved one million of francs in Summer 1993 by showing that a foreseen modification (adding a derivation in Achères station) of the network was not really valuable to increase the robustness of the timetables for travellers in case of incident. In another case, on the overcrowded line of Saint-Cloud and Versailles, the system proves, after many simulations, that the present ordering of the train was not optimal as regards the robustness in case of incident. Accordingly, the succession of the different trains with different destinations was changed, which results in lowering the perturbations. Thus, the sensitivity analysis that can be performed with SRE provides a worthy help to schedulers and investors. Up to now, the decisions of the system, when submitted to real dispatchers, were always validated as a possible interesting solution. It is the reason why the system is going to be used for the training of the dispatchers.

All these results appeared sufficiently valuable for the French Railway that they have now decided, for 1994, to launch a new development of the system to include the total course of a train from its departure from a suburb town to central station and return. Up to now, the robustness was studied separately for the timetables concerned by trains from suburbs to Paris and then another study for timetables from Paris to suburbs, only the importance of the reusing time at the central station was taken into account but not the whole return course. With this new tool and the associated scheduling system CHAO, to our knowledge, the French Railway will possess the only system in the world able to really support investment decision and timetable design taking into account the travelers requirements in term of punctuality.

7. CONCLUSION

SRE is an analysis tool for assessing, mainly from travelers' point of view, some very complex alternatives including a timetable, a railway network and a set of rolling stocks and railcars. Although many people are now convinced of the utility of multicriteria analysis, this methodology is not yet widespread in many organizations. One reason is that, generally, people in firms and administrations deal with very involved alternatives which are not only difficult to evaluate, but above all impossible to handle. This situation actually seems to be very frequent. So the first goal of the system is to help the decision-maker to construct realistic fully expanded alternatives from his basic ideas. Let us give another example among thousands : the same thing occurs when somebody wants to make a decision affecting his (her) competitors (e.g. introduction of a new product, decreasing a price). The basic alternatives are simple ; what is difficult is to

manage the chain of consequences according to the multiple nested decisions that will follow. Each basic alternative thus generates a fully expanded alternative that no decision-maker can handle and even build without a computer aid. This is what our system is intended to do in the case of timetable robustness. We think that this is a prominent issue for the credibility of decision analysis in general and, in particular, multicriteria analysis.

To our knowledge, this question has rarely been addressed up to now. It is only when the analysts are able to present realistic alternatives, evaluated along non-trivial criteria, that multicriteria analysis will be more widely used by real decision-makers. Unfortunately, it needs a lot of work and several tools (multicriteria methodology, expert systems, game theory, etc.), as SRE shows, to pass from a basic idea to a fully expanded alternative ; in any case it demands much more work than to invent a new aggregation function, for example, but we do believe that this is the price that has to be paid by multicriteria community to tackle real problems.

ACKNOWLEDGEMENT

We thank the French Railways (SNCF, Direction de Saint-Lazare) and especially Messrs. Romestain, Gautié, Gaultier, Louvière, Roux and Savin without whom this study would not have been possible. The authors also gratefully acknowledge the help of P. Lévine who, as Stratems CEO, directed the computer implementation of the system and the work of J.L. Morano, as Stratems engineer. Furthermore, we gratefully acknowledge the work of A. Sranon who completed her DEA report on this subject.

REFERENCES

- Bana e Costa C.A. (Ed.) (1990) : **Readings in multiple criteria decision aid**, Springer-Verlag.
- Bertereix G. and Beurrier A. (1990) : "Conception des horaires assistée par ordinateur (CHAO)", *Revue Générale des Chemins de Fer*, 109e année, mai, 29-36.
- Bohanec M., Bratko I., Rajkovic V. (1983) : "An expert system for decision making", in H.G. Sol (ed.), **Processes and Tools for Decision Support**, North-Holland, 235-248.
- Bohanec M., Rajkovic V. (1987) : "An expert system approach to multi-attribute decision making", **IASTED Conference on Expert Systems**, Geneva.
- Bouyssou D. (1989) : "Modelling inaccurate determination, uncertainty, imprecision using multiple criteria", in A.G. Lockett, G. Isley (eds.), **Improving Decision Making in Organisations**, *Lecture Notes in Economics and Mathematical Systems* 335, Springer-Verlag, 78-87.
- Bouyssou D. (1990) : "Building criteria : A prerequisite for MCDA", in C.A. Bana e Costa (ed.), **Readings in Multiple Criteria Decision Aid**, Springer-Verlag, 58-80.
- Du Bois Ph., Brans J.P., Cantraine F., Mareschal B. (1989) : "MEDICIS : An expert system for computer-aided diagnosis using the PROMETHEE multicriteria method", **European Journal of Operational Research** 39, 284-292.
- Efstathiou J., Rajkovic V., Bohanec M. (1986) : "Expert systems and rules based decision support systems", in G. Mitra (ed.), **Computer Assisted Decision Making**, Elsevier Science Publishers, 165-174.
- Felici A., Negri L. (1992) : "La qualité du service voyageurs des chemins de fer, Perception et souhaits des clients d'affaires et de banlieue", **Recherche Transports Sécurité** 36, 47-52.
- Komaya K., Fukuda T. (1988) : "An expert system for train-traffic control", **Mitshubishi Electric Advance** 43, June, 25-28.
- Komaya K., Fukuda T. (1989) : "Estrac-III : An expert system for train traffic control in disturbed situation", in J.P. Perrin (ed.), **Control, Computers, Communications in Transportation**, AFCET, Paris, 327-333.
- Lévine P., Pomerol J.Ch. (1989a) : **SIAD et Systèmes Experts**, Hermès, Paris.

Lévine P., Pomerol J.Ch. (1989b) : "Knowledge representation by schemata in financial expert systems", *Theory and Decision* 27, 147-161.

Lévine P., Pomerol J.Ch. (1990) : "Railcar distribution at the French railways", *IEEE Expert* 5, n° 5, 61-69.

Lévine P., Pomerol M.J., Saneh R. (1990) : "Rules integrate data in a multicriteria decision systems", *IEEE Transactions on Systems, Man and Cybernetics* 20, n° 3, May/June, 678-686.

Nilsson N.J. (1980) : *Principles of Artificial Intelligence*, Tioga Publishing Company, Palo Alto.

O'Leary D.E. (1986) : "Multiple criteria decision making in accounting expert systems", *Comptes Rendus des 6e Journées sur les Systèmes Experts et leurs Applications*, Avignon, 1017-1035.

Pomerol J.Ch. (1992) : "A comprehensive view of the relationships between MCDM and AI", Communication presented at **EURO XII/TIMS XXXI Joint International Conference**, Helsinki, Finland, June 29-July 1, 1992.

Rajkovic V., Bohanec M. and Batagelj V. (1988) : "Knowledge engineering techniques for utility identification", *Acta Psychologica* 68, 271-286.

Roy B. (1985) : *Méthodologie multicritère d'aide à la décision*, Economica, Paris.

Roy B. (1989) : "Main sources of inaccurate determination, uncertainty and imprecision in decision models", *Mathematical and Computer Modelling*, Vol. 12, n° 10/11, 1245-1254.

Roy B. (1990) : "Decision-aid and decision-making", *European Journal of Operational Research* 45, 324-331.

Roy B., Bouyssou D. (1993) : *Aide multicritère à la décision : Méthodes et cas*, Economica, Paris.

Sranon A. (1990) : *Réalisation d'un système basé sur la connaissance pour l'évaluation de la robustesse des grilles horaires de trains*, Université de Paris-Dauphine, Mémoire de DEA 103.