JOURNAL OF THE TRANSPORTATION RESEARCH FORUM

Volume XXXI Number 1 1990

TRANSPORTATION RESEARCH FORUM

Railroads - Tracks and rails - Maintenance - Scheduling

Expert systems (Computer science)

Application of Expert Systems and Network Optimization Techniques In Rail Relay Scheduling

by Dharma Acharya, Ph.D., Joseph M. Sussman, Ph.D.***
*Carl D. Martland****

ABSTRACT

The North American railroad industry spends hundreds of millions of dollars annually on installation of rails. Currently, rail relay scheduling, determining when to install which segment with what type of rail, is done manually, and is based largely on human judgments. Decisions made by different rail experts are likely to be inconsistent and result in an inefficient utilization of resources. This paper describes a computer-aided scheduling system that uses expert systems and network optimization techniques to work toward an efficient solution of the rail relay scheduling problem. The scheduling system efficiently utilizes the automatically collected rail profile data obtained from the electronic rail profile measurement technique to produce a five-year rail relay schedule. This system has been successfully run for about 20% of Burlington Northern Railroad's mainline track.

Two models, the Categorization Model (CM) and the Model for Optimal Rail Relay Scheduling (MORRS) have been developed to help BN to better relay plans. CM determines which segment must be relaid within the next five years and identifies segments that are likely to need to be relaid within that time. This involves human judgments based on rules of thumb and heuristics. Therefore, CM was developed using a knowledge-based expert system environment.

MORRS was developed to solve the problem of determining an optimal multi-year rail relay schedule given results from CM. MORRS generates a rail relay schedule that minimizes the total life-cycle rail-related costs. MORRS takes into account the fact that each segment should be relaid before it reaches the condition limit, and also considers the possibilities of benefiting by generating secondhand rail and of achieving economies of scale in start-up and transportation cost by relaying a group of neighboring segments in one year (although some of these segments may not need to be relaid yet).

The results showed that expert systems and network optimization techniques can be used effectively in tandem to approach complex problems such as the rail relay scheduling problem described in this paper. Expert systems provide the flexibility to incorporate policies and rules of thumb for limiting the size of the problem. Network optimization techniques provide the computational power to examine a vast number of options in order to identify the "best" solution. The key is to structure the problem so that each approach complements the other and produces a result which would be very difficult to achieve using either approach by itself. The potential benefits obtained are financially substantial and operationally important.

INTRODUCTION

Railroads must relay rail when it is no longer economical and/or safe to operate over the existing rail. They can relay with different types and sizes of rails, depending upon the curvature of the track and the annual traffic being carried by the track. The North American railroad industry spends hundreds of millions of dollars annually on this task. For example, in 1987 and 1989, Burlington Northern, Canadian National, CSX, Norfolk Southern, and Southern Pacific each planned to spend more than 50 million dollars for rail relaying (Wujcik 1987 and Morgan 1989). The problem of determining where to relay rail, using what type of rail, is called the rail relay scheduling problem.

Currently, the annual rail relay scheduling of a typical railroad is based on human judgments requiring hundreds of man hours of effort from various organizational units within the system. Because of the complexity of the rail relay scheduling problem, it is impossible for a human being to determine a single "best" rail relay schedule manually. In addition, decisions made by different rail experts are likely to be inconsistent and result in an inefficient utilization of resources. Even a small percentage of improvement over current rail relay scheduling practice could amount to an annual savings of millions of dollars. This paper describes a computer-aided scheduling model that uses expert systems and network optimization techniques to work toward an efficient solution of the rail relay scheduling problem.

Background

Rail is an important and expensive component of the track structure. It is in direct contact with train wheels. The purposes of having it in a track are to: 1) provide a safe and smooth-running surface for trains, 2) support the weight of the trains, and 3) distribute forces into the track structure (Hay 1982). Under repeated traffic loadings, the condition of rail deteriorates over time. In order to maintain the overall performance of the track, rail needs to be replaced (either by a new rail or by a rail in better condition) before it exceeds the condition limit(s). The condition limit(s) are usually determined by safety and economic factors for a given operating environment. If a rail is not replaced before it exceeds the condition limit(s), the chances of having service delays and train derailments due to rail failure increase rapidly. As a result, the railroad would have to bear high additional costs associated with these events.

The condition of a rail in a railroad track deteriorates as a function of the type and size of rail, track geometry, and traffic. Traditionally, the condition of a rail in a track in a given year has been characterized by cumulative rail head height and gage face wear and fatigue rate (i.e., defect rate). Rails were replaced when they exceeded wear or fatigue limits.

To reduce rail-related costs, a common practice within the railroad industry is to move rail from one position on the network to another (sometimes called "rail cascading" or "rail relaying"). New rail is installed on a heavily used segment of the railroad and the old rail (thus removed), is installed on a less heavily used portion of the network. The old rail may be removed from the first position before or when it reaches the condition limit(s). In this way, the old rail could last in the second position for a significant number of years and make rail relaying economically attractive. Since rail replacement and rail cascading refer to installation of new rail or of rail that is in a better condition, for simplicity and consistent with common usage within the industry, both will be called rail relaying henceforth.

In addition to using secondhand rail internally on light traffic lines, there may exist an outside market for secondhand or old rail. Under certain circumstances, it may be economically beneficial for a railroad to sell the secondhand rail to other railroads.

The possibilities of increasing the overall life of rail should also be considered while developing a rail relay schedule. For example, the overall life of rail at curves could be increased by transposition and grinding. When the gage face wear of the high rail and the head height wear of the low rail are within certain ranges, the rails on two sides of a curved track can be transposed or swapped. In this way, the high rail and the low rail will have more gage face and head height, respectively, to wear out. As a result, the overall life of the rail at curves is increased.

A rail segment may be relaid early due to high costs in maintaining the segment. For example, the condition of a segment may still be within the condition limits but due to high maintenance, service interruptions, and potential for derailment costs, it may be economical to relay the segment early in order to avoid such costs. Therefore, when considering rail relays, the possibility of relaying a segment early due to high costs associated with maintaining the segment should also be considered.

A typical class I railroad could easily have over 50,000 rail segments. A segment is a portion of track with homogeneous characteristic in terms of rail type and weight, wear, cumulative and annual traffic, and track geometry. It takes a good amount of time, typically one year, to plan for rail relaying.

In addition to the high labor and equipment costs associated with relaying rail, the transportation costs and the start-up costs could also be substantial. The transportation costs refer to costs associated with transportation of the steel (maintenance) gang and equipment to and from the site. The start-up costs refer to the costs associated with starting up at the beginning of every non-contiguous relay job. Therefore, under certain circumstances, it may be economical for the railroad to relay a group of segments in one year (although some of these may not need to be relaid at that point) to achieve economies of scale in transportation costs and start-up costs. Depending on the distance that needs to be travelled to reach the line under consideration, the savings in costs by relaying a group of segments in one year could be substantial. Therefore, it is important to consider these cost factors (i.e., the economies of scale factors) as well while developing a rail relay schedule.

Several models have been developed to understand the costs and impacts of rail deterioration (Wells et al. 1983, TRACS User's Manual 1989, Tew et al. 1986, Uzarski et al. 1988, Zarembski 1986). However, none of the models took into account: 1) previously described economic factors, and 2) the automatically collected rail condition information in developing their rail relay schedules. Reviews of existing rail deterioration/scheduling models are included in Acharya (1990) and Martland et al. (1990). In practice, analytic models have played a small role in rail relay scheduling decisions (Martland et al. 1990). This is primarily because of lack of input data. Hence, railroads have relied heavily on judgement of local officials in developing their rail relay schedules.

Some railroads are using the data obtained from the electronic rail profile measurement technique to overcome the problem of having inconsistencies in rail scheduling within different regions of a railroad (RT&S 1988). The electronic rail profile measurement technique automatically measures the shape of the two rail heads (i.e., the left and right rails) at a very short interval (every 15 to 30 feet) along the length of a track. Information collected by the electronic rail profile measurement technique is useful in making rail relay decisions, but rail experts often find that the information is too detailed. In the absence of a computer model, rail managers spend a good deal of time in manually identifying rail relay segments. With the availability of a computer-based rail relay scheduling system, rail managers will no longer need to analyze the rail profile data manually to prepare their annual rail relay schedule. If data regarding historical information on traffic, rail type, defects and others are available in a railroads' permanent databases, these data together with the rail profile data can be used efficiently in developing their rail relay schedules.

APPROACH

Determining a rail relay schedule is not difficult, but determining a close to optimal rail relay schedule that minimizes the total rail-related life-cycle costs occurring during the planning horizon with a reasonable set of constraints is a complex problem. When determining such an optimal rail relay schedule, we need to consider the following factors:

1) A typical railroad will have several rail relay gangs and relay equipment which can relay rail on different subdivisions of the railroad. A rail relay gang (sometimes called a "steel gang") consists of a number of crews each assigned for carrying out a certain task that is involved in rail relaying. And, a subdivision is a linear portion of track and normally connects two cities or junctions and has relatively constant traffic characteristics. In order to consider the possibilities of achieving economies of scale in rail relaying, we need to know which of the steel gangs and relay equipment should relay which subdivisions in each of the years during the planning horizon. This problem of determining the optimal rail relay-routes for several steel gangs and relay equipment for a given rail relay schedule itself is very complicated. A relay-route is a complete path (i.e., beginning from and ending to the corresponding maintenance plant) taken by the steel gang when relaying rail. A relay-route may normally include more than one subdivisions in its path. Moreover, the decision to relay certain segments (i.e., the rail relay schedule) will depend on the magnitude of the transportation cost, which will depend on the particular relay-route. Therefore, the determination of the relay-routes and the decision as to which segment should be relaid in which year need to be considered simultaneously. This problem is more complicated than a multiple vehicle routing problem.

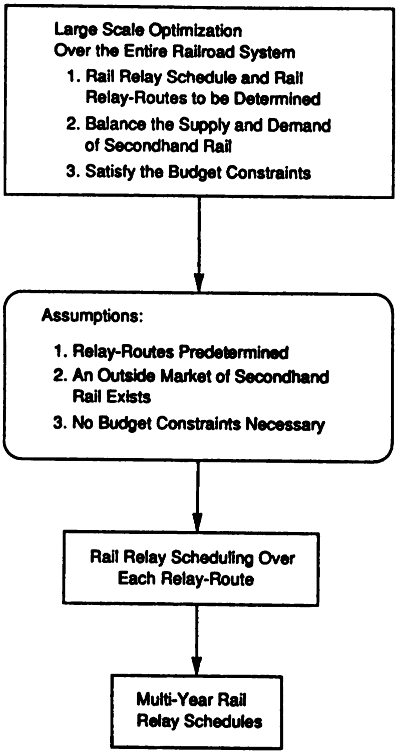
2) The problem could be further complicated by the fact that there may be constraints on balancing the supply and demand of secondhand rail and on the budget availability at the system, regional, or divisional level of the railroad.

If we were to formulate the above described multi-year rail relay scheduling problem for a typical railroad with over 50,000 rail segments as one large scale optimization problem, it would be impossible to solve. The total number of options for each possible combination of segment relay year and relay-route that we would need to evaluate would be extremely large (Acharya et al. 1989 and Acharya 1990). Thus, it would be impossible to identify a single best rail relay schedule from such a large number of possible options. Therefore, we need to make the problem smaller and solvable by making certain assumptions regardingsome of the factors described above and yet find a solution close to the global optimum.

Figure 1 shows our overall approach for reducing the size of the rail relay scheduling problem. The three assumptions shown in the figure will allow us to consider the rail relay scheduling problem over a small number of subdivisions at once and to find a solution close to the global optimum. The first assumption of the relay-routes being predetermined is reasonable and realistic because a rail manager would know the rail relay-routes that are commonly used over the years when relaying rail in his divisions or regions. This determination may have been based on many factors such as operational feasibilities and work rules. Of course, there is a possibility that the current or the existing relay-route is not the best of all the feasible relay-routes. However, we believe

FIGURE 1

Overall Approach



that considering all the constraints, the existing relay-route will be very close to the best possible relay-route.

The second assumption that there exists an outside market of secondhand rail is also reasonable and realistic because railroads do sell and buy secondhand rail to or from other railroads. If there is a high market price for certain types of secondhand rail for the area under consideration, it may be better for the railroad to relay certain segments early and benefit by selling the generated secondhand rail at a high market price. Conversely, if the market price of certain types of secondhand rail is very low, it may be economically beneficial for the railroad to buy and relay some segments with such secondhand rail. Therefore, there is no need to balance the supply and demand of secondhand rail within a railroad.

The third assumption of "No Budget Constraints Necessary" allows us to relax the budget constraints. Some railroad experts claimed that if the engineering department of the railroad can provide a reliable prediction of rail relay needs for next few years to its finance department, it will have enough time to allocate appropriate resources for each of the justified projects; hence adequate resources can be made available. The rail relay scheduling system described in this paper will help the engineering department of the railroad to make a more reliable prediction of the rail relay needs for the next few years. Therefore, we will not include any budget constraints in our formulation. However, as a by-product of the scheduling model, a prioritization mechanism can be devised later in the event a rail manager is still faced with limited resources.

Once we make the three assumptions described above, we no longer need to consider rail relay scheduling over the entire system of the railroad all at once. We can consider rail relay scheduling over the group of subdivisions that is in the relay-route. This is so because there are no other factors that need to be considered at a higher than relay-route level. Based upon the above-described assumptions, the optimal rail relay schedule obtained by solving the problem as one large scale optimization problem and the set of optimal rail relay schedules obtained by solving a multiple number of small optimization problems by each relay-route will be the same.

When we develop a rail relay schedule by relay-route, there are two major tasks that we need to accomplish: 1) predict or assess the condition of each segment of rail so that we know which segment must be relaid by which year, and 2) based on the predicted condition of segments, develop an optimal rail relay schedule that minimizes the total life-cycle cost incurred by the railroad. To accomplish each of the two tasks, we develop two separate models: 1) the Categorization Model (CM) and 2) the Model for Optimal Rail Relay Scheduling (MORRS), respectively. Each of the two models will be described, respectively, in the next two sections.

Figure 2 shows the different components of the rail relay scheduling system. As shown in the Figure 2, there are two major sources of input data, rail wear and rail statistical data. The rail wear data contains information on rail wear and the rail statistical data contains information such as the rail type and size, cumulative traffic and annual traffic, and defect rate. We use two models to compress the two sets of raw input data, which are too detailed for our rail scheduling purpose. The two compression models produce two independent sets of wear and rail segments of different lengths, which are used along with the other parameters in the CM to create another set of homogeneous segments in terms of wear and rail attributes. Each of the homogeneous segments are then categorized by the CM.

The segment categories along with some other parameters are used by MORRS or REPLACER to develop rail relay schedules. MORRS uses network optimization techniques to produce a five-year rail relay schedule and REPLACER uses a knowledge system environment to produce a one-year rail relay schedule (Mishalani 1989). In this paper, we will focus on MORRS, which is also a rapid way of producing a multi-year rail relay schedule.

THE CATEGORIZATION MODEL

The CM determines which segments must be relaid within the next five years and identifies segments that are likely to need to be relaid during that time period. To do this, we need to assess the condition of each segment of rail for each year of the five-year period. To predict the condition of each segment of rail for each year of the five-year period, we need to use rail condition deterioration models along with some heuristic rules, judgments, and empirical associations obtained from experts in the field. These can be incorporated better and can be used efficiently and flexibly in an expert system environment as opposed to a regular programming environment (Waterman 1986, Harmon and King 1985, Martland et al. 1989 and 1990). Therefore, an expert system will be developed to solve the categorization subproblem.

The condition of each segment in a subdivision will be classified with one of the following four categories for each year of the five-year program period:

MUST if rail condition will exceed critical relay parameters in analysis year.

SHOULD if rail condition will exceed critical relay parameters within two years from the analysis year. The decision to relay early will be based on economic factors.

MAYBE if rail condition will exceed critical relay parameters within three to five years from the analysis year. The decision to relay early will be based on economic factors.

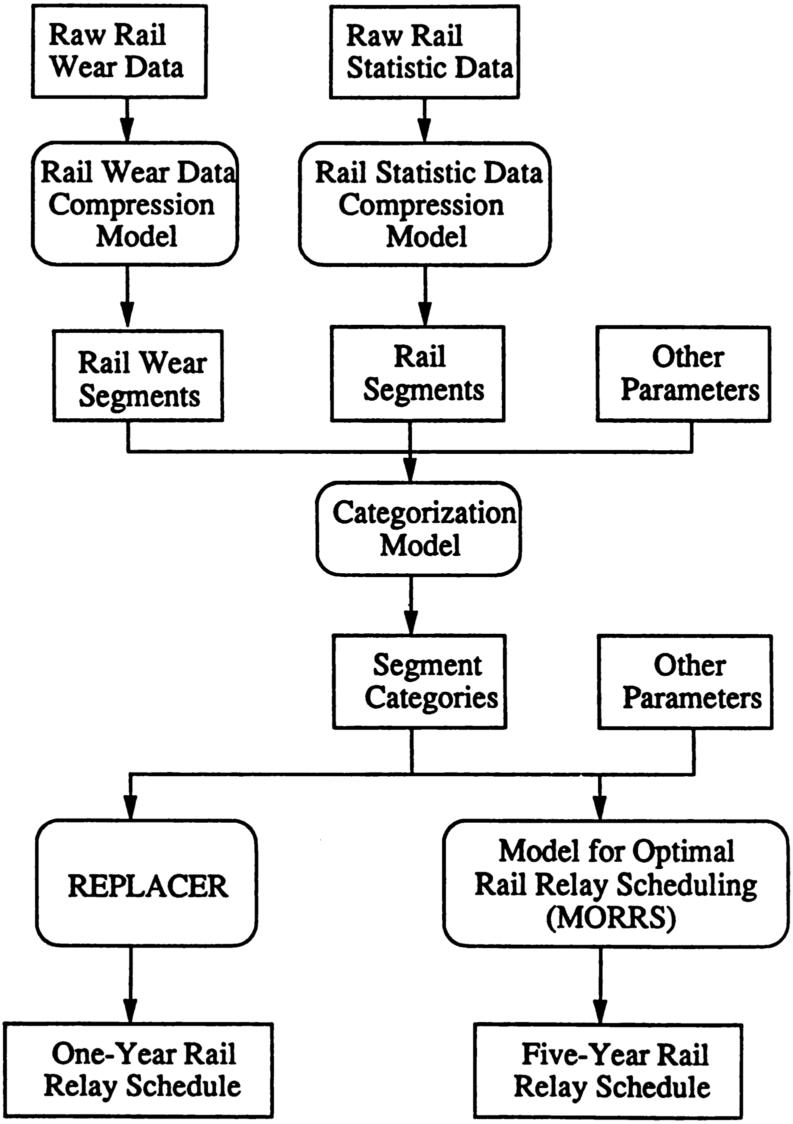
OK if the rail condition will **not** exceed any critical relay parameters in the year under consideration.

The process of assessing the condition of each segment (i.e., categorizing it) is consistent with what experts do in practice. At an early stage, rail managers identify all the segments that they would like to have relaid by examining the condition of the segment. As they start to get a better picture of the system level resources and used rail availability, they begin to defer relaying of less critical segments.

By categorizing the condition of the segments for each of the five years, we will be able to determine which segment must be relaid by which year (i.e., which segment-category becomes MUST in which year). We can also identify segments that are in good condition throughout the five-year period and are not likely to need to be relaid within that time period. We can ignore these segments when developing the five-year rail relay schedule. The definition of the "good-condition" rail may be relative and may depend on a particular circumstance. In this research, we will include only those segments that have a Year 5 category of at least MAYBE. In other words, when developing the five-year rail relay schedule, we will ignore segments that have the category of OK for all of the five years. The cost of ignoring such OK

FIGURE 2

Flow Diagram of the Rail Relay Scheduling System



segments would be either nothing or relatively small compared to other costs (Acharya, 1990).

The domain knowledge for the CM was obtained from a panel of field experts from BN. A preliminary set of rules for categorizing segments was developed based on interviews with field experts (Martland et al. 1988, Piech 1988). We felt that we needed more rules for categorization. Therefore, in order to obtain a more complete set of knowledge for categorization, we met again with field experts from the railroad. A series of meetings were held with systems engineering, regional engineering, financing, and research and development staff to discuss how they develop their annual rail relay program. During this stage, we also decided to present the domain knowledge in terms of decision trees and tables rather than in rules (although both of them are equivalent), which made it easier to understand and obtain feedback from field experts. The framework of the decision trees and tables were based on economic analyses, expert judgments, and the policy of the railroad. This set of decision trees and tables were reviewed and recommendations for further improvements were provided by the regional and divisional experts of BN (Acharya 1990). Rail experts believe that this set of decision trees and tables include most of the criteria a rail expert looks at when he or she develops a rail relay schedule.

An expert system (i.e., the CM) with the domain knowledge mentioned above was developed. The domain knowledge represented in the decision trees and tables was converted into a number of rule sets, each designed to accomplish certain tasks. In addition to the knowledge described by the decision trees and tables, rule sets for preprocessing the data and for computing some other parameters were included. The expert system was developed on a Compaq 386 in Goldworks, a LISP-based expert system development tool developed by Gold Hill Computers, Inc. (Goldworks 1987).

THE MODEL FOR OPTIMAL RAIL RELAY SCHEDULING

MORRS was developed to solve the problem of determining an optimal multiyear rail relay schedule given the results from CM. MORRS generates a rail relay schedule that minimizes the total life-cycle cost. MORRS also takes into account the fact that each segment should be relaid before it reaches the condition limit, and also takes into account the possibilities of benefiting by generating secondhand rail and of achieving economies of scale in start-up and transportation cost by relaying a group of neighboring segments in one year (although some of these segments may not need to be relaid yet).

The optimal rail relay scheduling problem can be formulated as a network optimization problem. Figure 3 shows the optimal rail relay scheduling problem in a time-space network diagram. The time-space network diagram of an optimal rail relay scheduling subproblem is obtained by: 1) expanding each segment of the linearized network over time, and 2) adding additional arcs to represent the cost structure associated with rail relaying. As described previously, we take out all the segments that are in good condition throughout the five-year period because they are not likely to need to be relaid for at least another five years.

The first layer of the network in Figure 3 represents the option of not relaying the segments (e.g., arc a). The second layer of the network represents the option of relaying segments in Year 1. The third layer represents the option of relaying segments in Year 2 and so on. Each node in the network is numbered by two indices, the first one representing the beginning or end of a segment (the index with no prime represents the beginning of a segment and the index with a prime represents the end of the segment) and the second index represents the year in which the relay option is being considered. The relay-option arcs pointing from the beginning of a segment to the end of a segment in Year y represent the option of relaying that particular segment in Year y, where y may be any number between 1 and 5 (e.g., arc b in Figure 3 is the relay-option arc for Segment 1 in Year 4). Some of the segments must be relaid by Year x, where x is the year when the segment category becomes MUST (e.g., the category of Segment 2 becomes MUST in Year 4). In that case, that particular segment will not have:

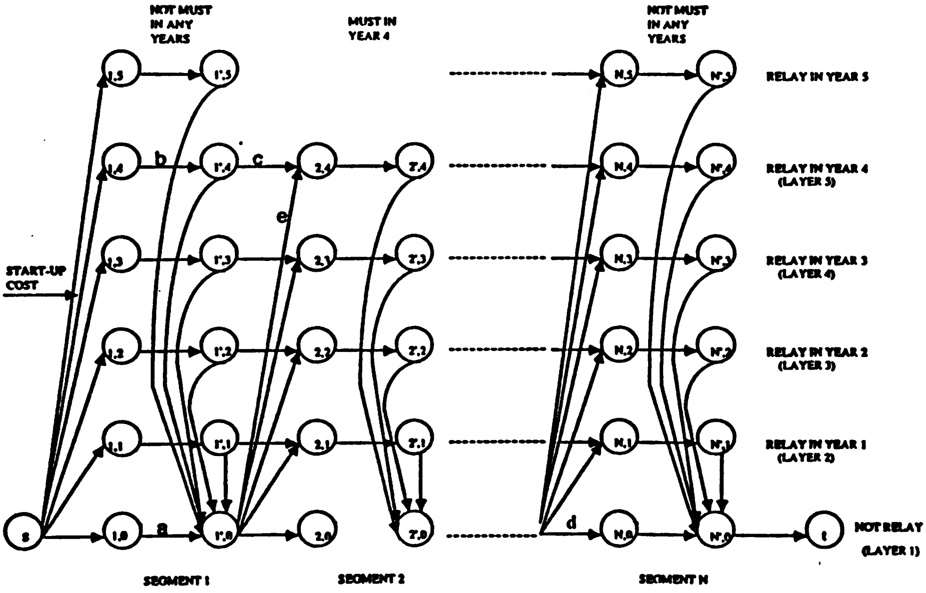
1) a not-relay-option arc (the corresponding arc in the first layer of the segment, e.g., the lack of an arc from Node (2,0) to Node (2',0) of Segment 2), and
2) arcs with relay option from Year x+1 to Year 5, e.g., the lack of an arc from Node (2,5) to Node (2',5) of Segment 2.

Each of the relay-option arcs and not-relay-option arcs have costs associated with them and all the costs included in the network are in terms of discounted costs. The discounted costs associated with each relay-option arc in the network include:

a) Rail relay material, work gang, and equipment costs
b) The expected future maintenance and derailment costs

FIGURE 3

Time-Space Network Diagram of an Optimal Rail Relay Scheduling Problem



c) Credit for generating the secondhand rail, and

d) Credit for the terminal value of the rail in track at the end of Year 5.

The not-relay-option arc in the network (e.g., arc a) will only include the expected future maintenance and derailment costs because the other three types of costs that are included in the relay-option arcs (i.e., items a, c, and d listed above) are not relevant to the not-relay-option arcs. Arc a is a not-relay-option arc; there would be no rail relay cost incurred and there would be no secondhand rail generated.

The credit for the terminal value of rail in track is used for making the solution obtained with short planning horizon similar to that with long or infinite planning horizon (Acharya 1990). Essentially, the effect of such a credit is that when minimizing the costs, we include only a portion of rail relay cost that is used until Year 5.

The credit for the terminal value of rail in track is also not applicable to the not-relay-option arcs in Figure 3. The rail that is already in the track was relaid before this analysis period and the costs associated with the relaying of the segment is a sunk cost, which can not be considered as a cost incurred during our five-year analysis period. Therefore, we cannot give credit for the terminal value of the rail in track at the end of the five-year analysis period for the not-relay-option arc.

In any of the years, the arcs pointing from the end node of a segment's not-relay-option arc to the following segment's beginning node of the relay-option arc in any of the years are the fixed cost arcs (e.g., arc e in Figure 3), which represent the discounted fixed or start-up cost incurred in starting or setting up the equipment for relaying if the previous segment is not relaid in the same year.

The arcs pointing from the end node of a relay-option arc to the beginning node of the relay-option arc of the following segment in any year represent the economies of scale arcs (e.g., arc c in Figure 3). These arcs take into account the condition that there would be no steel gang and equipment start-up costs incurred while relaying the successive segments if the consecutive and physically connected segments are relaid in the same year. The cost associated with such an economies of scale arc is zero, because there would be no work gang and equipment start-up costs incurred if the consecutive and physically connected segment is relaid in the same year. If two consecutive segments in the network are not physically connected (i.e., if there are OK segments between them and we are ignoring them in the time-space network representation while developing a rail relay schedule for the next five years), there then would be no such economies of scale arcs between the relay-option arcs of the two segments. However, there still will be a zero cost arc between the not-relay-option arcs of the two consecutive but physically separate segments.

As described above, all the arcs shown in the network represent a particular activity and each arc has costs associated with that activity. Therefore, in order to find a combination of relay schedule for each segment that minimizes the total cost, we have to find a shortest cost path between the beginning point, s, to the end point, t, of the network. We have developed a shortest path algorithm that finds the shortest cost path between the two nodes s and t efficiently (Acharya 1990). The algorithm is based on the recursive relation of a dynamic programming problem. This algorithm is faster than any of the regular shortest path algorithms because it exploits the special structure of the network.

A network is a topological network if there are no directed cycles in the network. In other words, it is a topological network if all the nodes in the network can be numbered in such a way that every arc in the network always points from a low numbered node to a high numbered node. The network representation of the optimal rail relay scheduling subproblem is a topological network because it has no directed cycles. The shortest path between two nodes in a topological network can be found in the linear order of the total number of segments in the network (i.e., O(m)). Since there are no cycles, by the time we reach a node, we would have examined all the possible paths to get to that node. As a result, a node does not need to be examined more than once. Consequently, the running time of the algorithm will be significantly less than that of the regular shortest path algorithms.

In the network formulation described above, the effect of economies of scale in start-up cost at the beginning of a relay job is included (e.g., arc c in Figure 3), but the effect of economies of scale in transportation cost is not included. The transportation cost is the cost of transporting the steel gang and equipment when relaying rail in the relay-route. The economies of scale in transportation can be achieved if we can skip relaying (i.e., avoid the need for transporting the steel gang and equipment) at the relay-route in some years. If we skip relaying in a year, the transportation cost for that year will be zero. This can be done by relaying all the segments that must be relaid by a certain year early (i.e., before it becomes a MUST) and hence avoiding the need for transportation for that year.

To determine whether it is economical to skip relaying in certain year, we need to run the shortest path algorithm for the modified network for each of the possible and feasible skipping combinations (at most, 31 times when the length of the planning horizon equals five years). The modified network is obtained if all the arcs beginning and/or ending at the skipping year are deleted. We then compute the total cost for five years by adding the corresponding amount of total discounted transportation cost to the shortest path cost of each possible and feasible combination. The relay schedule corresponding to a relay-year combination that minimizes the total cost will be the optimal five-year rail relay schedule.

Some additional features can also be incorporated by modifying the network and/or by modifying certain arc costs. Techniques have been developed to incorporate the possibilities of coordinating the rail relay schedule with other track maintenance program and the possibility of relaying a segment more than once during the five-year period (Acharya 1990).

CASE STUDY

In the previous two sections, we have developed models of a computerized scheduling system that generates a rail relay schedule for a five-year period. In this section, we will run the system for a significant length of track using the real data from the Burlington Northern Railroad and show that the system can be used in practical application, in solving rail relay scheduling problems in the railroad industry. We will also analyze and evaluate the results from the system.

It is important to demonstrate that the scheduling system described in this paper can be used in solving real rail scheduling problems in the railroad industry. We demonstrate this in this section by presenting the results from a case study which required running the scheduling system for a significant length of track using real railroad data.

Input Data

The rail relay scheduling system has been successfully run to develop optimal five-year rail relay schedules for 13 subdivisions covering about 1,850 miles (i.e., 3,700 rail miles) of track for Burlington Northern Railroad. This amounts to about 20 % of the mainline tracks of BN. In order to avoid the possible confusions with a large amount of data and results, we will use 6 out of the 13 subdivisions for our case study (Table 1).

The traffic density of the six subdivisions ranges from very low to very high (i.e., the annual traffic ranges from less than 10 MGT/year to greater than 50 MGT/year). The distribution of cumulative traffic also varies considerably among the six subdivisions (Table 1). The distribution of the other statistics, such as bolted or welded and new or secondhand, is also different among the six. For example, some subdivisions contain mainly bolted rails and others contain mainly welded rails. Historically, some subdivisions were mainly relaid with new rail and others with both new and secondhand rail. Since the distributions of cumulative traffic are different, the distributions of the total head and gage face wear are also different among the six subdivisions (Table 1). We believe that, collectively, these six subdivisions represent the typical range of subdivision types that any railroad is likely to have.

Case Study Results

Results from the CM:

The results from the CM include the five-year categorization, the possibility of transposition, and rail relay types for each rail segment. The categorization result of a subdivision reflects the overall condition of the subdivision. The greater the percentage of OK and MAYBE segments in the subdivision, the better its overall condition. The case study results from the CM for each of the six subdivisions are summarized in Table 2.

The distributions of the percentage of the total length of MUST category rails in Years 1 through 5 in Table 2 are quite different among the six subdivisions. The reason for the difference is that some subdivisions, such as Subdivisions C and D, are high traffic density subdivisions (see the MGT of Subdivisions C and D in Table 1). They have a high percentage of MUST segments throughout the five-year program period. When looking at the change in the total length of MUST category rails over the five-year period, we can see that the condition of Subdivisions C and D is deteriorating in a somewhat faster rate than the other four subdivisions. The reason for this faster rate of deterioration is that the annual traffic on these two subdivisions is very high (i.e., the majority of the rails have annual MGT $\geq$ 50, see Table 1). In contrast, for other subdivisions, such as A, B, E, and F, the annual traffic density is relatively low and hence the total length of MUST category rails increases relatively slowly during the five-year program period. Therefore, the deterioration rate of the four subdivisions is relatively slow. As we noticed from Table 1, we can also see from Table 2 that most of the rails in all of the six subdivisions are in the OK category, which means most of the rails are in good condition. However, if we look at the percentage of the length of rail that remains in good condition over the five-year program period (i.e., the percentage of OK condition rail in Year 5), the percentages range from 82% (Subdivision D) to 99% (Subdivision A) among the six subdivisions. Therefore, the differences in the deterioration of rail condition among the six subdivisions are significant.

The results from Table 2 itself could also be very useful for the rail managers. They can see how fast the condition of different subdivisions deteriorate over a five-year period. By looking at the categorization results such as shown in Table 2, rail experts can get a rough estimate of rail relay need for the five-year program period.

Results from MORRS:

The results from MORRS for the six subdivisions described in Table 1 are tabulated in Table 3. Each of the six subdivisions are geographically distant from each other. Therefore, each of the six subdivisions are in separate relay-routes. As a result, we can run MORRS for each subdivision (i.e., each relay-route) independently.

The transportation cost associated with each relay-route is $12,000 per scheduled year. The start-up cost used was $1,313 and the market prices of secondhand rail generated were $200/ton, $300/ton, and $150/ton for Code 3, Code 4, and Scrap rail, respectively. In this paper, we use two groups of secondhand rail: 1) Code 3, if both head height and gage face wear are less than 1/4 of an inch, and 2) Code 4, if both head height and gage face wear are less than 3/8 of an inch and if wear on at least one of the two is between 1/4 and 3/8 of an inch. The costs used in this case study are reasonable costs but not necessarily the costs that BN would use to develop their rail relay schedules. All the costs shown in the Table are discounted costs with discount rate = 10%.

As shown in Table 3, the percentage of the total length of rail scheduled to be relaid during the five-year period is very small (about 1.1%). Relaying one percent of BN's track miles would mean relaying about 250 miles of track, which would amount to about 50 million dollars. Since Subdivisions C and D are high traffic density subdivisions and their condition deteriorate faster than the other four subdivisions, their percentage of the total length of rail scheduled to be relaid during the five-year period (4.7% and 3.9%, respectively) is higher than that of the other four subdivisions.

TABLE 1

Basic Characteristics of Case Study Input Data

Description	Length in Rail Miles						
	Subdivisions						
	A	B	C	D	E	F	Total
Total Miles of Data Available:	202	211	88	46	368	211	1126
Joint Type							
Bolted:	151	123	1	3	220	11	509
Welded:	51	88	87	43	148	200	617
Existing Rail							
New:	202	188	86	46	364	148	1034
Secondhand:	0	23	2	0	4	63	92
MGT/year							
≤ 10:	9	0	1	0	368	0	378
10 - 30:	193	210	0	0	0	211	614
30 - 50:	0	1	7	0	0	0	8
≥ 50:	0	0	80	46	0	0	126
Cumul'tve MGT							
< 200:	21	70	2	1	291	211	596
200 - 450:	49	102	38	20	77	0	286
450 - 600:	123	39	16	17	0	0	195
600 - 800:	9	0	21	5	0	0	35
800 - 1000:	0	0	3	3	0	0	6
> 1000	0	0	8	0	0	0	8
Total Miles of Wear Data Avail.	202.2	211.0	88.3	46.3	367.8	211.2	1126.8
Head Ht. Wear (in inches)							
< 0.20:	197.5	207.9	76.6	42.4	324.3	198.8	1047.5
0.20 - 0.38:	3.9	3.0	11.2	3.2	42.3	11.7	75.3
0.38 - 0.50:	0.1	0.1	0.5	0.8	1.2	0.7	3.4
0.50 - 0.70:	0.8	0.0	0.0	0.0	0.0	0.0	0.8
≥ 0.70:	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gage Face Wear (in inches)							
< 0.20:	202.2	211.0	88.3	46.3	357.1	208.8	1113.7
0.20 - 0.50:	0.0	0.0	0.0	0.0	10.7	2.4	13.1
0.50 - 0.75:	0.0	0.0	0.0	0.0	0.0	0.0	0.0

TABLE 2

Results from the Categorization Model

Description	Cumulative Length in Rail Miles						
	Subdivisions						
	A	B	C	D	E	F	Total
Total Length Analyzed:	202	211	88	46	368	211	1126
MUST							
Year 1:	0.90	0.10	1.15	0.87	0.62	0.40	4.04
Year 2:	1.51	0.11	1.61	1.10	0.91	0.73	5.97
Year 3:	1.51	0.34	2.76	1.70	0.97	1.17	8.45
Year 4:	1.90	0.39	3.83	1.75	1.40	1.24	10.51
Year 5:	1.91	0.55	4.14	1.94	1.79	2.21	12.54
SHOULD							
Year 1:	0.67	0.20	1.01	0.86	1.02	0.76	4.52
Year 2:	0.40	0.28	2.12	0.70	1.06	0.51	5.07
Year 3:	0.45	0.20	1.29	0.25	1.42	1.08	4.69
Year 4:	0.09	0.54	1.42	0.33	1.63	1.95	5.96
Year 5:	1.52	0.96	2.01	1.41	2.06	2.20	10.16
MAYBE							
Year 1:	0.42	0.85	3.50	0.36	21.26	2.04	28.43
Year 2:	1.53	1.35	2.83	1.54	21.75	3.17	32.17
Year 3:	1.78	1.50	4.83	5.14	21.71	3.82	38.78
Year 4:	2.60	1.20	5.56	5.72	22.85	4.16	42.09
Year 5:	1.15	0.90	5.66	5.11	22.58	5.02	40.42
OK							
Year 1:	200.33	209.88	82.64	44.26	344.93	207.98	1090.02
Year 2:	198.89	209.30	81.75	43.01	344.10	206.77	1083.82
Year 3:	198.59	208.98	79.42	39.26	343.71	205.10	1075.06
Year 4:	197.74	208.90	77.49	38.60	341.95	203.80	1068.48
Year 5:	197.74	208.60	76.50	37.90	341.39	201.75	1063.88

TABLE 8

Results from MORRS

	Subdivisions						
	A	**B**	**C**	**D**	**E**	**F**	**Total**
Total Length Analyzed:	202	211	88	46	368	211	1126
	Length in Rail Miles to be Relaid						
Year 1:	1.9	0.6	1.8	1.1	1.1	1.2	7.7
Year 2:	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Year 3:	0.0	0.0	2.3	0.7	0.0	0.0	3.0
Year 4:	0.0	0.0	0.0	0.0	0.8	0.0	0.8
Year 5:	0.0	0.0	0.0	0.0	0.0	0.9	0.9
Total:	1.9	0.6	4.1	1.8	1.9	2.1	12.4
%:	0.9	0.3	4.7	3.9	0.5	1.0	1.1
Start-up Cost ($)							
Year 1:	6,566	9,193	7,879	5,253	9,193	18,385	56,469
Year 2:	0	0	0	0	0	0	0
Year 3:	0	0	10,853	5,426	0	0	16,279
Year 4:	0	0	0	0	10,853	0	10,853
Year 5:	0	0	0	0	0	4,485	4,485
Transp. Cost ($)							
Year 1:	12,000	12,000	12,000	12,000	12,000	12,000	72,000
Year 2:	0	0	0	0	0	0	0
Year 3:	0	0	9.917	9,917	0	0	19,834
Year 4:	0	0	0	0	9,015	0	9,015
Year 5:	0	0	0	0	0	8,196	8,196
Secondhand Rail Gener'td (in Tons)							
Code 3:	0	0	9	0	0	0	9
Code 4:	0	44	149	42	0	0	235
Scrap:	190	15	301	149	200	222	1077
Total Budget Used ($):	135,472	56,086	357,874	165,843	127,093	164,398	1,006,766
Obj.Func.($)	70,873	43,645	226,778	117,483	84,651	90,567	633,997

All the costs shown in the Table are discounted costs with discount rate = 10%

All six subdivisions have a certain length of rails to be relaid in Year 1. There are no rails scheduled to be relaid in Year 2. For some subdivisions such as Subdivisions A and B, there are no segments to be relaid in the remaining four years as well. However, for the other four subdivisions (i.e., Subdivisions C, D, E, and F), there are segments scheduled to be relaid in one more year between Years 3 and 5 (i.e., in addition to relaying in Year 1). For example, for Subdivision D, in addition to 1.1 miles scheduled to be relaid in Year 1, there are 0.7 miles of rails scheduled to be relaid again in Year 3.

In this section, the results from the rail relay scheduling system were analyzed, and the results obtained made sense. The rail relay schedules produced by MORRS has been compared with those produced by REPLACER, developed on a knowledge system environment as a part of the same project. As in the CM, REPLACER includes a set of rules designed to identify segments that satisfy certain criteria for replacement (Mishalani 1989). REPLACER also uses part of the categorization results and generates rail relay schedules for the first analysis year only. The percentages of matches between the rail relay schedule produced by MORRS and that of REPLACER for the first year for the above described six subdivisions ranged from 99.5% to 66.7%. The reason there are some differences are due to differences in the approaches used and the factors considered by the two models. For example, MORRS examines the temporal and cumulative effects of multiple factors in rail scheduling over the entire subdivision over five-year period, whereas some of these effects are not captured in REPLACER.

The methodologies used and the kind of results produced from MORRS and REPLACER were presented to BN officials. BN officials preferred the results from REPLACER over MORRS for the first analysis year because: 1) it was easier to follow the methodology used in REPLACER (i.e., rules), and 2) it provided reasoning or justification for including each segments in the schedule. In addition, their priority at this initial stage is not in obtaining an optimal or close to optimal rail relay schedule but in obtaining a reasonably good schedule from a system that uses simple and understandable logic so that rail managers from different divisions feel comfortable in using it. However, they preferred MORRS over REPLACER for the purposes of rapid analysis, strategic planning, and predicting resource requirements for a multiple number of years. This helped us understand their priorities, which was very helpful in further improving the models. For instance, after we found that the officials liked the reasoning or justification mechanism included in REPLACER, we also devised a technique to incorporate the reasoning or justification mechanism in MORRS.

The running time of MORRS for a typical subdivision was about 2 minutes, whereas the running time of REPLACER was 30 to 40 minutes. Later on, REPLACER has also been run for a mountainous subdivision of BN. Due to the complexity of the subdivision, many more homogeneous segments were created and as a result, REPLACER took hours to run. The running time of MORRS would not be affected much by such factors. Therefore, the ability of doing rapid analyses also makes MORRS more attractive than REPLACER.

We demonstrated that the rail relay scheduling system could be used to solve a real-world problem. In fact, once we have the data ready, running the model is not a major problem. In the next section, we summarize the results from sensitivity analyses of the rail relay scheduling system.

SENSITIVITY ANALYSES

The previous sections analyzed the results from the rail relay scheduling system for six subdivisions. As a part of this research, sensitivity analyses with respect to rail wear limits, transportation cost, start-up costs, prices of secondhand rail, and expected defect repair and derailment costs were carried out. We found that it costs more to operate a subdivision with tighter wear limits than with relaxed wear limits. And, beyond certain wear limits, it will also cost more for the railroad with too relaxed wear limits because of too frequent service failures and derailments. Therefore, in order to minimize the overall cost, the railroads should use an optimal wear limits which will be somewhere in between the "tight" and "relaxed" wear limits. Such an optimal wear limits may also be different for different subdivisions.

The rail relay schedule developed by the system was sensitive to the magnitude of the transportation cost. If the transportation cost is high, more segments are relaid together in one year (although some of these segments may not need to be relaid yet) and hence the subdivision is visited less often for relaying. Conversely, if the transportation cost is low, the majority of segments relaid in any year are of the MUST category and the subdivision is visited almost every year for relaying.

The schedules produced by the system were also sensitive to the magnitude of the start-up cost. For instance, when the start-up cost is high, more contiguous segments are relaid in one year (again some of these segments may not need to be relaid yet) and when the start-up cost is low, only the segments with the MUST category are relaid in any year.

The prices of secondhand rail also affected the rail relay plan produced by the scheduling system. For example, if the prices of secondhand rail are low, segments are not relaid until necessary, unless economies of scale can be achieved by relaying early with adjacent segments or subdivisions. If the prices of secondhand rail are very high, segments are then relaid early in order to benefit by generating secondhand rail.

The schedules generated by the scheduling system were also affected by the magnitude and the rate of increase of the expected future maintenance and derailment costs. If the expected annual maintenance and derailment costs are relatively low, the optimal rail relay schedules obtained by increasing the magnitude and the annual increment rate of the maintenance and derailment costs within a certain limit are similar. However, if the maintenance and derailment costs are relatively high, more segments are relaid early to avoid annual increases in maintenance costs due to the poor condition of rail.

FINDINGS AND CONCLUSIONS

We have developed a rail relay scheduling system that uses the rail profile data obtained from the electronic rail profile measurement technique along with other rail-related data as input and produces a close to optimal rail relay schedule. We ran the scheduling system using real railroad data for a significant length of track, which demonstrated that the rail relay scheduling system can be used by railroads to develop their annual rail relay schedules.

The categorization of segments involved human judgments based on rules of thumb applied by rail experts in developing their rail relay schedules, and on the predictions of future conditions based on a limited amount of historical condition data. An expert system with the capability of incorporating rules of thumb and with some computation capability was believed to be an appropriate framework for categorization. Based on the prediction of future condition and on condition limit parameters set by the railroad, each segment was categorized as MUST, SHOULD, MAYBE, or OK for each year of the planning horizon.

The results showed that expert systems and network optimization techniques can be used effectively in tandem to approach complex problems such as the rail relay scheduling problem described in this paper. Expert systems provide the flexibility to incorporate policies and rules of thumb for limiting the size of the problem. Network optimization techniques provide the computational power to examine a vast number of options in order to identify the "best" solution. The key is to structure the problem so that each approach complements the other and produces a result which would be very difficult to achieve using either approach by itself. The potential benefits obtained are financially substantial and operationally important.

Since BN's rail wear and rail statistic data were never used before in such an extensive manner, the most amount of time was spent in assuring the quality of the data. One of the indirect benefits gained by BN after we began using the rail-related data for running the rail relay scheduling system was in improving the quality of these data that could be used for other purposes as well. Before this, there was no strong incentive or need to keep rail-related data of high quality.

The next step in the research is to further evaluate the system results and identify areas for improvements. Possible future research areas are to incorporate prioritization and reasoning mechanism in MORRS. With the help of priorities, rail managers can easily identify relay projects that can be deferred in case of limited resources. And, with the help of reasoning for each relay projects, it will be easier for managers to understand and justify relaying of segments. Basic rules that can be used to develop a reasoning mechanism from the reports produced by MORRS has been developed (Acharya 1990). Finally, it would be interesting to understand the effect of different lengths of scheduling period to the rail relay plans. With this, possibly more can be understood with respect to how should the credit for the terminal value of rail be computed.

ACKNOWLEDGEMENTS

The authors acknowledge the assistance of other members of the project team in the Center for Transportation Studies at MIT especially Rabi Mishalani, Rajesh Dontula, and Mark Dunzo who assisted in conducting the case study. At Burlington Northern many people from different departments especially Jim Eshelby, John Hovland, Gordy Besh, Tom Klarkowsky, Hank Lees, and John Beacom who helped in the development and evaluation of this system.

REFERENCES

1. Acharya, D., *Using Expert Systems and Network Optimization Techniques for Rail RelayScheduling*, Ph.D. thesis, Department of Civil Engineering, MIT, February 1990.

2. Acharya, D., C. D. Martland, and J. M. Sussman, *Using Expert Systems and Optimization Techniques for Rail Relay Scheduling*, Proceedings of the First International Conference on the Applications of Advanced Technologies in Transportation Engineering, American Society of Civil Engineers and the California Department of Transportation, San Diego, CA, February 1989.

3. Goldworks, *User's Manuals*, GoldHill Computers Inc., 1987.

4. Harmon, P. and D. King, Expert Systems, John Wiley and Sons, Inc., 1985.

5. Hay, W. W., Railroad Engineering, John Wiley and Sons, 1982.

6. Martland, C. D., S. McNeil, D. Acharya, R. G. Mishalani, and J. Eshelby, *Applications of Expert Systems in Railroad Maintenance: Scheduling Rail Relays*, Transportation Research-A, Vol. 24A, No. 1, pp. 39-52, 1990.

7. Martland, C. D., S. McNeil, J. E. Eshelby, D. Acharya, and R. G. Mishilani, *Scheduling Rail Relays with the Assistance of an Expert System*, Proceedings of the International Colloquium on Railroad Applications of Knowledge Based and Expert Systems, Vienna, Austria, March 8-9, 1989.

8. Martland, C. D., S. McNeil, and M. Piech, *Applications of Expert Systems in Railroad Maintenance*, Rail Group Working Paper, MIT, June 20, 1988.

9. Mishalani, R. G., *Planning Rail Replacements: Satisfying Multiple Criteria Within a Knowledge System Environment*, Master's Thesis, Department of Civil Engineering, MIT, September, 1989.

10. Morgan, T., *For Roadway Investment, Another Good Year*, Progressive Railroading, March, 1989.

11. Piech, M., *REPOMAN: A Prototype Expert System for Rail Maintenance Scheduling*, Rail Group Working Paper, MIT, July 24, 1988.

12. RT&S, *Using Rail Information*, Railway Track and Structures, pp. 12, January 1988.

13. Tew, G. P., D. C. Davis And P. J. Dwyer, *The Application of Computer Modelling to Track and Vehicle Maintenance Practice*, Third International Heavy Haul Railway Conference, Session III, pp IIIB-51-1 - 10, Oct. 13 - 17, 1986.

14. TRACS, *Users Guide to the AAR's Total-Right-of-Way Analysis and Costing Systems*, Association of American Railroads and Center for Transportation Studies, MIT, March 1989.

15. Uzarski, D. R. and D. E. Poltkin, *Interim Method of Maintenance Management for U.S. Army Railroad*, Transportation Research Record 1177, 1988.

16. Waterman, D. A., A Guide to Expert Systems, Addison-Wesley Publishing Company, 1986.

17. Wells, T. R., *Rail Performance Model: Evaluating Rail*, American Railway Engineering Association, 1983 Annual Convention, 1983.

18. Wujcik, M., *Holding the Line on M/W Spending*, Progressive Railroading, March 1987.

19. Zarembski, A. M., *Track Maintenance Planning and Forecasting System, Phase I: Feasibility*, Final Project Report by Zeta-Tech Associates to the Department of Transportation, DOT SBIR 85-038, April 1986.

ENDNOTES

* Senior Engineer, Research and Test Department, Association of American Railroads

** Director, Center for Transportation Studies

*** Principal Research Associate, Department of Civil Engineering, Massachusetts Institute of Technology.