Optimization in the Rail Industry

Alexandra M. Newman
Department of Operations Research
Naval Postgraduate School
Monterey, CA 93943

Linda Nozick
School of Civil and Environmental Engineering
Cornell University
Ithaca, NY 14853

Candace Arai Yano
Department of Industrial Engineering and Operations Research
University of California
Berkeley, CA 94720-1777

September 13, 2002

1 Introduction

Deregulation from the Staggers Rail Act along with improved technology and productivity has led to a 50% increase in rail freight transportation since 1980. Railroads transport over 40% of the ton-miles of intercity freight, including 70% of new automobiles, 40% of farm products and 65% of coal in the U.S. (Association of American Railroads, 1999). With increased profit incentives and better availability of computer systems, railroads have begun to adopt optimization-based decision support with some success.

We focus on problems that are specific to the rail industry because of special features of infrastructure, operations, or cost structures, and that are amenable to solution by optimization methods. We first describe these industry-specific characteristics and then discuss problem areas and commonly-used modeling constructs.

1.1 Special Characteristics of the Rail Industry

The nature of travel and freight handling are the most distinctive characteristics of rail transportation. Trains operate on limited-capacity tracks over long distances, and their hauling capabilities depend on the assigned set of locomotives. Moreover, railcars, containers (traveling on *stack cars*) and truck trailers (traveling on *flatcars*) can be transferred between trains only at rail yards.

Due to the high cost of installing and maintaining track ($9 billion is spent annually in the U.S. on maintenance alone), many railway networks consist of single tracks with periodic sidings, where trains may wait while others pass. This necessitates detailed *meet-pass plans* that specify when and where trains meet and pass or overtake one another, and related train *timetables*.

Train crew costs exhibit strong economies of scale; they depend on the number of crew-shifts and are fairly insensitive to the number of transported railcars. Locomotive operating costs include a fixed-charge component for their use, and variable (per ton-mile) costs for the transported freight. Different combinations of locomotives, or *consists*, may be assigned to a train. These assignments must satisfy constraints on minimum tractive power or number of locomotives, that depend on the topography of the rail segment. Train scheduling problems are complicated by fixed charges for each train and for each additional locomotive. Moreover, the possibility of assigning multiple locomotives to a train results in motive planning problems quite different from those encountered in truck or air transportation. The long travel distances involved in rail transport also increase the importance of minimizing equipment deadheading and empty car movement because of both direct costs and opportunity costs of idling expensive equipment (e.g., $1 million locomotives or $65,000 railcars) for long periods.

Both space for storage and capacity to handle railcars and containers are limited at rail yards, which have various track configurations and use different mechanisms for sorting or *classifying* railcars. Typically, each track segment is allocated to a *block*, which consists of a group of cars traveling together for one or more portions of the journey. *Train makeup* refers to the assignment of these blocks to trains. The layout of the yard and the capacity of the handling equipment constrain the number of blocks, and the workload imposed on the yard affects in-transit delays. Containerized freight may be handled differently because of its fragility. The containers, rather than the stack cars, may be rearranged, but doing so imposes additional workload on expensive, specialized cranes which are often the bottlenecks at intermodal rail terminals.

Figure 1 depicts a simple rail network. Trains may carry grain and coal to urban areas and manufacturing sites. Container ships may generate intermodal traffic bound for domestic sites while an automobile factory may provide intermodal exports. The trains may visit rail yard(s) where cars are reclassified. Sidings allow trains to meet and overtake along single-tracked lines. Signals at sidings and intersections govern traffic flow.

1.2 Rail Decision Problems

We review several major categories of rail optimization models. Section 2 covers infrastructure planning models concerned with the design of a railway network, locations of track sidings, and track maintenance and improvement. These are strategic decisions that require a long-term, systemwide view. Problems of sizing fleets of lo-

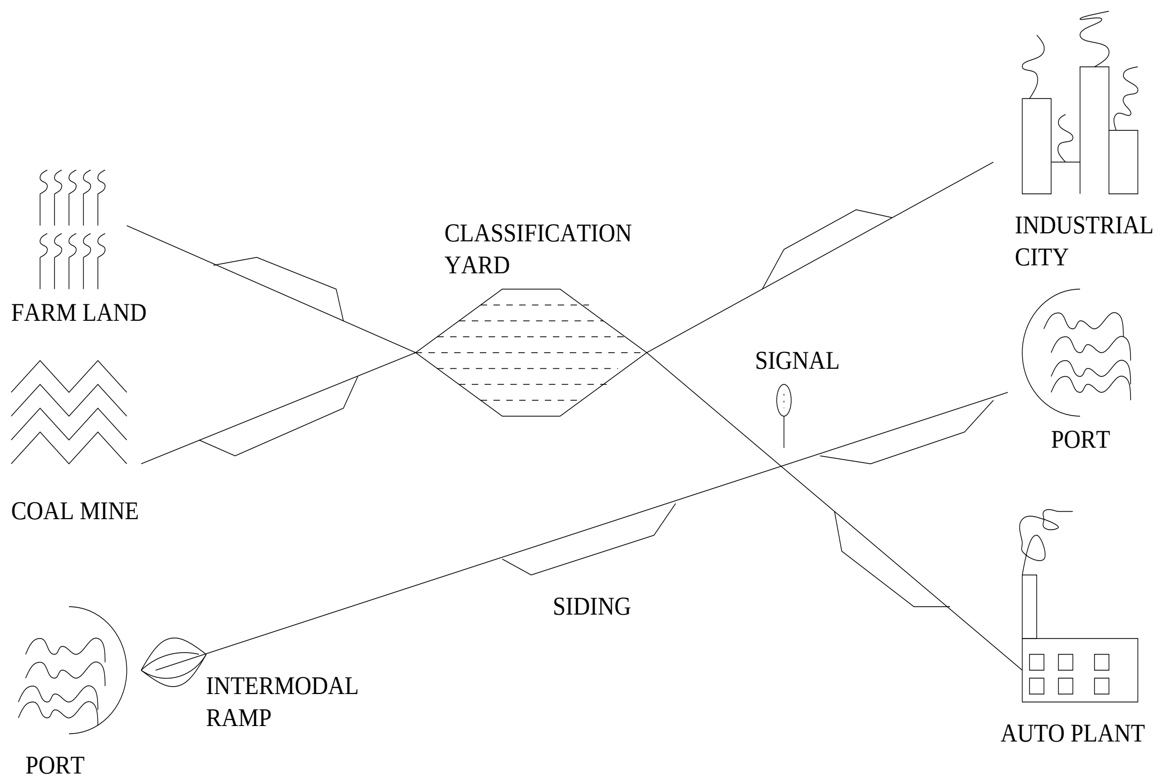

Figure 1: Rail Network

comotives and rail cars, which are covered in Section 3, also are strategic decisions. However, they commonly are addressed using more detailed data on demand, train schedules and motive power constraints because these factors strongly influence fleet requirements.

Section 3 also surveys models for short- to medium-term locomotive, railcar, and container repositioning. Some are tactical models that rely on steady-state analysis and suggest long-run rebalancing strategies, but most are designed to consider the state of the system and detailed demand forecasts, train schedules, etc., to determine when and how locomotives, and empty railcars and containers should be repositioned.

Section 4 discusses *train scheduling* and *freight routing.* Tactical models in this section are concerned with the *service frequency* of direct and indirect trains, blocking patterns (i.e., paths specifying intermediate transfer locations) to be utilized, and allocation of freight to the blocking patterns. As in other tactical models, these decisions are based on average costs and demand rates. Operational models focus on constructing more detailed train schedules and/or freight routing to satisfy demands while considering other factors (such as due dates, blocking plans, and locomotive availability). The level of precision may extend to minute-to-minute timetables which are often determined concurrently with meet-pass plans.

Space limitations preclude us from covering models of rail yard operations; most are descriptive models based on queueing concepts. Related discussion appears in Petersen (1977a, 1977b) and Turnquist and Daskin (1982). We also omit real-time decisions for train scheduling, freight routing and equipment repositioning. Most work in these areas relies on short-term planning approaches for decision-support, and does not address the real-time problem directly. More commonly, real-time decisions are based on judgment and experience. We omit railcar loading and unloading. Interesting applications of optimization to these areas appear in Bard (1997) and Vasko (1994). Finally, we do not discuss crew scheduling. Models for rail crew management and relevant references can be found in Caprara et al. (1997).

Several survey papers cover topics that we discuss. Assad (1980) provides excellent background on institutional aspects of and operating policies for rail transportation.

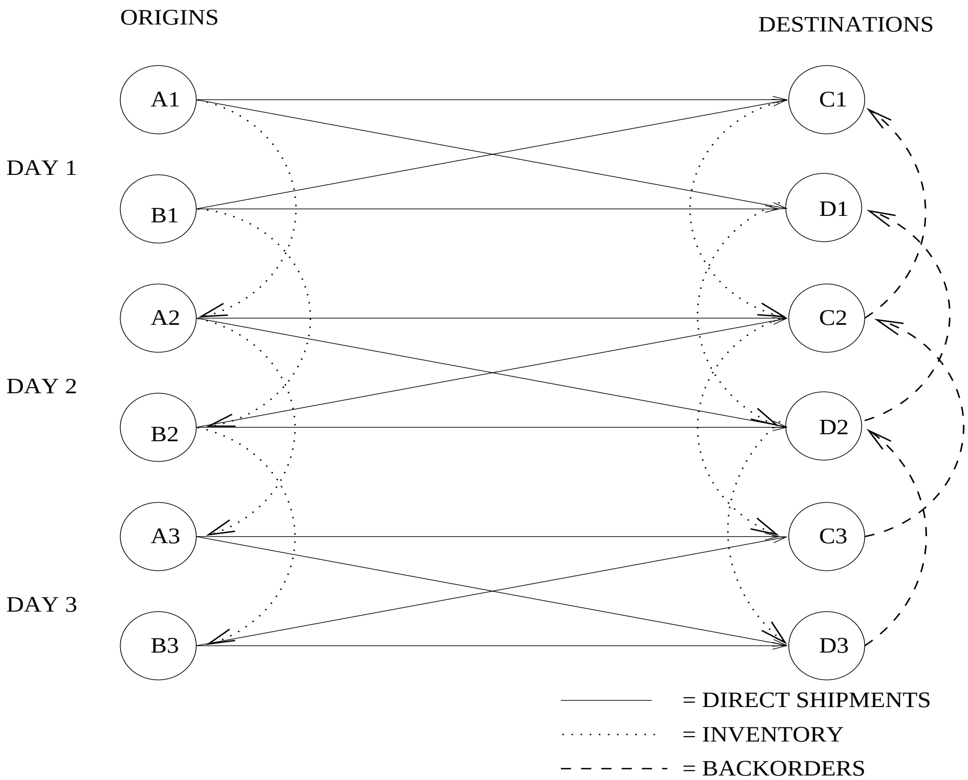

Figure 2: Time-Space Diagram

We cite more focused surveys within the corresponding section of this chapter. We have attempted to describe representative models that collectively provide the reader with both a broad perspective of decision problems and an understanding of how researchers and practitioners have addressed them using optimization models.

1.3 Modeling Constructs

Several modeling constructs are commonly used for rail planning and scheduling problems. One important construct is a *time-space network*; a simple example with instantaneous travel times appears in Figure 2. Each node represents a time-location pair. Arcs connect nodes if it is feasible for a railcar or locomotive to move from one time-location pair to the other. Arcs from one period to the next at the *same location* represent holding inventory and "reverse" arcs from one period to a previous period represent backordering of demand for applicable equipment.

Time-space networks lead naturally to problem representations as single or multicommodity networks. When train capacity and timing are given, the network contains one capacitated arc per train and the flows represent railcars, locomotives and/or containers. The "commodities" in the network may be distinguished by their origin, destination, priority, and their physical attributes (e.g., coal versus automobiles). The latter distinction is important due to differences in the rail cars used for various goods. When the train service and/or timing are to be decided, the resulting models become *network design problems*. In some cases, arc capacities may also be decided (e.g., by selecting the number of locomotives). Some fleet sizing problems may be modeled by adding arcs for external supply of capital equipment to the basic time-space network. Research on single and multicommodity network flow and design problems has advanced rapidly in the past decade; many algorithms applicable to rail planning and scheduling problems can be found in Ahuja et al. (1993) and Magnanti and Wong (1984).

The rail industry commonly uses *string diagrams* to represent detailed scheduling problems in continuous time and space domains. In string diagrams, the locations of trains, usually along a single rail line, are plotted, with the x-axis representing time and the y-axis representing the distance from one end of the line. Thus, the slope of line segment corresponds to the velocity of the train, and horizontal segments represent waiting at a siding or a rail yard. Note that string diagrams are designed for representing trains, not railcar, container, or locomotive movements.

The string diagram in Figure 3 illustrates a meet-pass plan and timetable for two southbound trains (I and II) and a northbound train (III) over three hours. There are three sidings, A, B, and C. The pass that occurs between trains II and III at siding C is feasible. However, a collision would then ensue between trains I and III which could be avoided if train I waits at siding B.

2 Infrastructure Design and Maintenance

U.S. railroads are making significant investments in both track and rail yards. Burlington Northern Santa Fe Railway (BNSF) has improved and added track along its main-

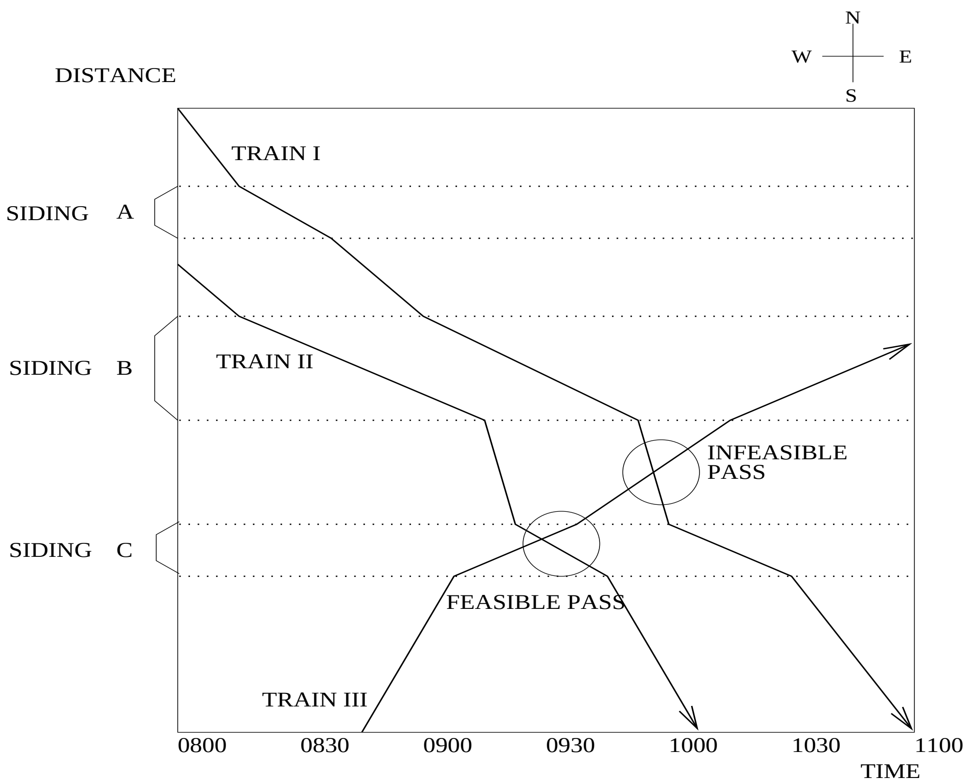

Figure 3: String Diagram

line in five states in the past year. Union Pacific has expanded its lines in Iowa and recently upgraded a bridge crossing in Louisiana. Norfolk Southern opened new intermodal terminals in North Carolina and, jointly with Kansas City Southern Railway, in Texas. Despite this growth, few optimization models have been developed for rail infrastructure planning.

Higgins et al. (1995) address the problem of determining the location of sidings on a single rail line to minimize total tardiness for a planned train schedule. Construction costs are not considered. The solution method involves iteratively optimizing the train schedule (meet-pass plan, departure and arrival times of trains) for revised siding locations, and optimizing the locations of sidings for a revised train schedule. They report that modifying siding locations leads to a substantial reduction in tardiness.

The ambitious task of planning expansions of the Brazilian Railway is addressed by Crainic et al. (1990). Much of their work involves parameter estimation. The optimization model itself seeks to minimize the average cost per unit time of loaded and empty freight flows, expressed in tons, for a given network configuration. The model permits multiple transportation modes (e.g., different track gauges) and freight transfers between modes. The authors present an application of the model for proposed rail lines.

Improved track maintenance reduces travel times, which increases labor productivity and customer service, and reduces the probability of derailment. LeBlanc (1976) develops a nonlinear optimization model in which investments in track improvements have decreasing marginal returns for improving train speed and decreasing operating costs. The problem is to simultaneously select investment levels (including implied abandonments) for each track segment and freight routing to minimize overall costs. For a particular relation between the investment cost and the variable transportation cost on each segment, he shows that the problem can be solved optimally.

Re-laying track involves moving used rail to a lower-traffic or lower-speed location where it can still be used effectively. Acharya et al. (1989) address the problem of determining when stretches of track should be re-laid after an expert system has identified candidates for replacement. There are economies of scale from re-laying nearby stretches of track, even if immediate replacement is not required. Operating costs, future maintenance and derailment costs, the salvage value of the re-laid rail, and the total value of the rail are considered. Constraints ensure that track is replaced before its wear limits have been reached. Labor, operating budgets, and equipment use may be limited. This constrained shortest path problem is solved via Lagrangian relaxation of the capacity constraints.

A discussion of issues and models for rail maintenance planning, including the scheduling of rail repair facilities, appears in Genser (1982). Ferreira and Murray (1997) highlight the need for methods to maximize net financial benefit considering design standards, maintenance expenditures, and the capability of a rail segment to handle increased loads or train speeds.

3 Fleet Sizing and Repositioning

Minimizing empty car and locomotive repositioning contributes to reducing the ownership and operating costs for locomotives and railcars. In this section we address these issues for a fixed train schedule. The integrated problem of scheduling trains and routing loaded and empty railcars is discussed in Section 4. We first concentrate on repositioning and then discuss fleet sizing.

3.1 Fleet Repositioning

For a fixed train schedule, fleet management entails the assignment of loaded and empty railcars and locomotives to trains to satisfy shipment requirements and empty car demands. Most research on fleet management, however, takes the freight movements as given and optimizes empty car and/or locomotive repositioning.

Turnquist and Markowicz (1989) address the problem of minimizing the total cost of moving, holding inventory and backordering of empty cars. Their model permits multiple car types and limited substitutability among them. The problem is formulated as a single-commodity minimum-cost network flow problem. A decision-support tool based on this model was used regularly by CSX from 1990 to 1996.

An interesting repositioning problem pertains to railroad autoracks, which are multi-level railcars that carry automobiles. An empty autorack generally had been returned to its last loading location, a practice that resulted in many empty car miles. In 1982, all railroads that move automobiles entered into a pooling agreement under which an unloaded autorack is repositioned to an economically efficient location, possibly another auto manufacturer's site.

Sherali and Suharko (1998) develop a decision support system which is used daily by RELOAD, a central management group that controls autorack repositioning. It incorporates two discrete-time optimization models. The first model seeks, in effect, to minimize the maximum tardiness in filling demands for each autorack type at each location. Constraints ensure that each automaker's cumulative car-day utilization is consistent with its contribution to the pool. Each automaker's prioritization of its plants is incorporated via priority weights. Travel time uncertainty is considered by using chance constraints to exclude repositioning options that are unlikely to meet the target transit time. This problem can be decomposed by autorack type and solved by standard network flow algorithms. For the second model which also considers blocking, the authors evaluate several heuristic procedures based on priority schemes and approximate dual prices. They also present a method to modify the solution to the first model to account for blocking.

The need to select among available locomotive consists for assignment to trains is a major distinction between railcar and locomotive management problems. Ziarati et al. (1997) seek to optimize the schedule for locomotive movements between "power change points" (where locomotives may be transferred between trains) to support a one-week train schedule, requirements of "outposts" for locomotives handling local pick-up and delivery, and scheduled maintenance. The sub-network for each locomotive type is a single-commodity time-space network, with costs for hauling and deadheading. Bundle constraints limit the total locomotives of each type, and additional constraints ensure that assigned consists satisfy power requirements. The authors develop a branch-and-bound procedure with lower bounds derived using Dantzig-Wolfe decomposition. There is a subproblem for each locomotive type and the master problem links the types. The authors test the procedure on a data set from Canadian

National North America for one week in 1994 (about 2000 trains) and report the potential for substantial savings.

Fleet management problems typically are formulated as large, integer programs that become extremely difficult to solve if there are complex constraints. Some researchers have developed approximate decomposition approaches as a step toward surmounting these challenges.

Powell and Carvalho (1998a) consider the problem of routing vehicles and assigning full or empty loads to them to maximize profit over a finite horizon. Each full load earns revenue only if delivered within a specified time window, and costs are incurred for both loaded and empty vehicle movements. They develop an approximate decomposition method for their dynamic program. The state of the system is defined by the vector of empty-vehicle inventories by location and the set of uncovered tasks. They assume that the "value-to-go" for additional vehicles at each time and location is a linear function, with the estimated incremental value as its coefficient. With this, the optimal solution for each state is determined easily using a greedy procedure: available vehicles should be dispatched or held in inventory in descending order of the incremental benefit of the available assignments. The coefficients are updated and the process repeats until the objective value converges.

Powell and Carvalho (1998b) apply the aforementioned approach to container and flatcar management. They also implement it within a decision support system for locomotive fleet management to be launched at the Norfolk Southern in 1999. A forecasting module predicts loads on outbound trains for the next 14 days. Using these forecasts, a scheduling system determines the assignments of locomotives to trains and estimates the incremental value of each locomotive type at each yard in each period. A real-time heuristic is employed to select a consist for each train during the upcoming 24 hours, considering factors such as repositioning costs and the value of each consist at the train's destination.

In reality, supplies and demands of railcars and travel times are stochastic. Jordan and Turnquist (1983), Crainic et al. (1993), and Beaujon and Turnquist (1991; see Section 3.2), among others, model the stochastic aspects of this problem. For surveys on fleet management and empty flows, see Dejax and Crainic (1987), Haghani (1987) and Cordeau et al. (1998). Recent references also appear in Holmberg et al. (1998).

3.2 Fleet Sizing

Beaujon and Turnquist (1991) develop a multi-period stochastic program to simultaneously determine fleet size and routing decisions to maximize expected revenues from moving loads less costs of empty and loaded car movement, ownership, lease and shortages. Demands for loaded movements are stochastic and non-stationary. Travel times are represented by probability mass functions on the integers. The authors approximate the inventory and backorder costs as deterministic, nonlinear functions of the mean and variance of on-hand inventory or backorders at the corresponding time and location. They propose a heuristic which iterates between solving the underlying network flow problem for assumed variances (from the last iteration, if available), and computing updated variances resulting from the network flow solution. In computational experiments, the procedure produces solutions superior to those obtained from heuristics based on deterministic travel times.

Sherali and Tuncbilek (1997) address the problem of minimizing the number of additional autoracks required to satisfy demand in all time periods and locations. They propose a network flow model with time-varying demands, and with recirculation of flows from sink to source to account for end-of-horizon effects. The problem is decomposed heuristically into overlapping time intervals. This model is used annually by RELOAD to develop recommendations for autorack purchases and the apportionment of costs among the automakers.

Little research has been done on locomotive fleet sizing. Gertsbach and Gurevich (1977) address the problem of minimizing the size of a fleet of homogeneous locomotives to cover a fixed schedule by assigning a "chain" of transport segments to each locomotive. Their procedure relies on properties of good chains in constructing an optimal solution and is applicable to both finite-horizon and periodic, repeating schedules. Florian et al. (1976) appears to be the only paper that permits nonhomogeneous locomotive consists. The objective is to minimize the total cost of capital investment and maintenance for the locomotives. The model is an aggregate one which ensures a balance of locomotive flows among locations, but does not account for the effects of the train schedule and connections on fleet requirements.

Table 1 characterizes the articles reviewed in this section, including the decisions, major assumptions, objective function, and important constraints. Note that self-evident constraints, such as conservation of flow, are omitted. For reviews of fleet management and sizing, including empty car repositioning, see Dejax and Crainic (1987) and Haghani (1987).

4 Scheduling and Routing

In this section, we discuss *train scheduling* and *freight routing*, including *blocking* and *train makeup* decisions. These problems are addressed, both in research articles and in practice, at different levels of demand aggregation across time. The most aggregate models use average demand rates over an appropriate time horizon. At an intermediate level, customer shipments may be aggregated within each origin-destination pair, but not across time, leading to time-varying demands. At the most detailed level, the current system status is considered for real-time decision-making (not discussed here).

We first discuss meet-pass planning and closely-related timetabling problems, followed by routing problems for conventional railcars, where blocking decisions are integral, and for intermodal traffic where blocking is not required. We then continue with joint train scheduling and freight routing problems, and finally discuss more comprehensive models.

4.1 Meet-Pass Planning and Timetabling

Most timetabling models involve single-track paths shared by multiple trains, often with bi-directional traffic. Brännlund et al. (1998) address a timetabling problem on a single rail line, assuming a constant velocity for each train. The track is divided into segments, each with a capacity of one train. The goal is to maximize profits from the selected train itineraries, where profits depend on the departure time and enroute delay. The problem is solved via a Lagrangian heuristic in which the segment capacities are relaxed. See Carey and Lockwood (1995) for related work.

Sauder and Westerman (1983) develop a meet-pass planning system for a set of intersecting tracks at Southern Railway (a predecessor of Norfolk Southern Railway). Their model seeks to minimize total weighted delay, where tardiness is weighted according to the train's priority, and

Paper/year	Decisions	Major Assumptions	Objective Function	Constraints
Fleet Management and Fleet Sizing				
Turnquist & Markowicz (1989)	Empty car routing	-Backorders allowed -Fixed train schedule -Partial substitutability of car types	Minimize railcar transporta- tion, holding & back- order costs	-Supply & demand of cars by time & location
Sherali & Suharko (1998)	Empty car repositioning	-Backorders allowed -Fixed train schedule -Travel times uncertainty considered	Minimize maximum weighted tardiness	-Consistent cumulative car-day use vs. pool contribution
Ziarati et al. (1997)	Locomotive scheduling	-Known train makeup -Fixed train schedule	Minimize locomotive transportation costs	-Power requirements -Scheduled maintenance -Shop capacities -Locomotive demands at outposts
Powell & Carvalho (1998a)	Scheduling of empty & loaded movements	-Fixed train schedule -Single equipment type	Maximize revenue less vehicle movement costs	-Only loaded movements within time window earn revenue
Powell & Carvalho	Flatcar/ container routing	-Fixed train schedule -Multiple equipment	Maximize profit (as above)	-Container assignments to flatcars
Jordan & Turnquist (1983)	Empty car distribution	-Stochastic empty car supplies, demands & travel times	Maximize expected revenue less car trans., holding & shortage costs	-Car availability
Crainic et al. (1993)	Empty container distribution	-One container type -Demands, supplies partially dynamic & uncertain	Minimize sum of operating & ex- pected holding & leasing costs	-Demand satisfaction -Limits on container movements between depots
Beaujon & Turnquist (1991)	Fleet sizing & vehicle routing	-Dynamic & stochastic demands -Travel times uncertain	Maximize revenue less expected car movement, use & shortage costs	———
Sherali & Tuncbilek (1997)	Fleet sizing	-Fixed train schedule -Time-varying demands	Minimize fleet size	-Demand satisfaction
Gertsbach & Gurevich (1977)	Fleet sizing, locomotive scheduling	-Fixed train schedule -Homogeneous fleet	Minimize fleet size	-Arrival & departure time
Florian et al. (1976)	Fleet sizing, locomotive scheduling	-Fixed train schedule -Multiple locomotive types	Minimize capital & maintenance costs	-Power requirements

Table 1: Literature Classification: Fleet Management and Fleet Sizing

delays within the scheduled arrival time have unit weight. The optimal solution is obtained by enumerating all meets and passes. The meet-pass plans are integrated with a real-time simulator to support on-line decisions. In 1.5 years of monitored performance, delay declined by 15% with steady or increasing volume. Savings in fuel, crew and cost of capital were estimated at over $300,000 annually during the early 1980's.

Jovanovic and Harker (1991) develop the SCAN-I model to construct timetables and meet-pass plans over a set of interlinked traffic lanes, with a focus on robustness against travel time randomness. To assess whether a timetable is feasible under deterministic assumptions, they employ a branch-and-bound variant called a *process-interaction* simulation which proceeds forward in time, resolving conflicts between trains sequentially, while avoiding deadlock. Backtracking occurs when a feasible continuation cannot be constructed. Whenever a schedule is identified that is feasible for the deterministic travel times, simulation is used to estimate the probability that the schedule is achievable for random travel times. The SCAN-I model has been used by BNSF for special studies.

See Kraay et al. (1991) for a timetabling approach allowing variable train velocity, which permits more flexibility in scheduling and fuel cost reductions.

4.2 Routing

The need to make blocking decisions complicates the routing of conventional railcars. Virtually all blocking models are based on aggregate, deterministic, steady-state assumptions. The first successful solution of a blocking optimization model appears to be that of Bodin et al. (1980) who utilize commercial software to solve their mixed-integer program. Van Dyke (1986) provides an excellent discussion of practical issues related to blocking.

Newton et al. (1998) seek to minimize total mileage, handling and delay costs subject to on-time delivery constraints that vary with priority class, and constraints on blocks and railcar handling at the yards. For each origin-destination pair, a set of blocking patterns and the railcar flow on each must be decided. They devise a branch-and-price procedure in which column generation is used to construct and price candidate paths. The binary blocking decisions are handled via branch-and-bound. The model is being extended and tested with a view toward implementation at CSX.

Kwon et al. (1998) formulate a multicommodity flow problem to determine car routes assuming the train schedule, blocking plan, and block-to-train assignments are given. The goal is to minimize late delivery penalties while ensuring demand is met, cars are appropriately assigned to blocks, blocks are appropriately assigned to trains, and train capacity restrictions are enforced. They use a column generation approach to solve realistic problem instances. This model has proved useful for modifying train schedules when the initial train schedule does not provide adequate customer service.

All major North American railroads currently use MultiRail software, which, among other things, assigns railcars to existing blocks. The objective is to minimize variable transportation and handling costs incurred for all transit segments, but costs may be modified to reflect routing preferences. Block and yard capacities are not explicitly modeled, allowing car routing to be determined using a shortest path algorithm (van Dyke, 1999).

Blocking decisions are not essential for intermodal freight because less reclassification occurs en-route. Furthermore, intermodal goods are promised more rapid delivery, so due dates are important and steady-state models are no longer adequate. Nozick and Morlok (1997) address a finite-horizon, discrete-time problem of minimizing the total variable cost of moving loaded and empty trailers and flatcars *given a fixed train schedule* while satisfying due dates. They develop a procedure that involves iteratively solving a linear programming relaxation and rounding some of the resulting fractional values until a feasible integral solution is found. The heuristic is shown to provide good results.

4.3 Combined Scheduling and Routing Models

Morlok and Peterson (1970) introduce one of the first models for concurrent routing and scheduling decisions. The objective is to minimize the sum of fixed costs for trains, variable costs for transportation, handling and storage of freight, and opportunity costs of using rail equipment, while providing on-time deliveries of time-sensitive goods. Each potential train has a departure time, routing, set of stops, and an upper limit on cars. Decisions are which trains to operate and which freight to assign to each train. The authors apply branch-and-bound to solve a small instance of the resulting multicommodity network design problem.

Gorman (1998) treats the discrete-time problem of simultaneously deciding train service on all possible non-stop links and freight allocation. The model seeks to minimize the cost of labor, locomotive utilization, fuel, freight handling, and the opportunity cost for equipment use subject to on-time delivery, train capacity, rail yard handling and aggregate track capacity constraints. The proposed tabu-enhanced genetic search procedure is tested on small problems. Because of difficulties making widespread changes, a heavily constrained version of this model is used by BNSF for evaluating new routing and scheduling plans. A similar problem is addressed by Newman and Yano (1998), who present both centralized and decentralized approaches for solving the problem.

Huntley et al. (1995) devise a procedure for scheduling trains and routing cars to transport the one billion bushels of grain handled by CSX Transportation each fall. Each potential train is defined by a non-stop route between two locations and a departure time. Routing decisions are made for batches of cars with the same origin, time of availability at origin, and destination. The objective is to minimize a complex nonlinear cost function, including labor, fuel, freight car rental, and locomotive capital expenditures, while ensuring that no cars miss train connections, and no trains exceed car-carrying capacity. Starting from a simple initial solution, simulated annealing is used to search for improved solutions by considering addition or deletion of stops or changing train departure time. The company reports substantial savings from using this procedure.

Marín and Salmerón (1996) address an aggregate, steady-state freight planning model in which train routes (including stops), their frequency, and the number of cars using each service are determined. Costs include a fixed charge for each train, handling and delay costs, and costs of investments in additional trains. Constraints are imposed on the number of cars transported on each track segment, the number of cars using each yard, and the number of trains. They suggest heuristics in which service frequency decisions are handled by simulated annealing or tabu search and freight routing is addressed using a network flow model.

For surveys of rail scheduling and routing for freight, see Cordeau et al. (1998) and Assad (1980).

4.4 Integrated Models

More comprehensive models can provide better-coordinated decisions. The resulting size of such models, however, dictates that detailed decisions (e.g., blocking or train timing) and local constraints (e.g., yard capacity) must be ignored or modeled approximately.

Keaton (1989) examines the problem of simultaneously deciding which pairs of terminals are provided direct train service and its frequency, car routing, and allocation of blocks to trains, to satisfy constant demand rates among the terminals. The objective is to minimize the average cost per unit time incurred from a fixed cost for each unlimited-capacity train, variable costs for the use of cars, and delays for classification and train assembly. Constraints are imposed on the number of blocks formed at each yard but delays due to block formation are not modeled. Keaton develops a Lagrangian procedure, relaxing constraints on the number of blocks, as well as a simpler heuristic in which, starting from a feasible solution, train frequencies are reduced and some connections are eliminated.

Haghani (1989) considers the discrete-time problem of simultaneously determining train service, train makeup, and loaded and empty car movements. The objective is to minimize the sum of fixed charges for trains, variable car movement, classification, delay and empty car backorder costs, and penalties for undelivered goods. Constraints ensure that the allocation of locomotives to each route is adequate to handle the flow of loaded and empty cars. The solution procedure involves heuristically rounding the train variables. An example illustrates that the integrated approach provides better customer service at a lower cost than a sequential solution approach.

Abel et al. (1981) develop a model for scheduling sugar cane harvesting times at growers and scheduling trains to transport it to mills in Queensland, Australia. Only one route exists between each grower and its associated mill, but a train need not visit all growers on its route. The goals, in ranked order, are to minimize total operating cost, minimize “staling” of sugar cane, minimize locomotive and railcar capital costs, and minimize the number of shifts required at the mills. The authors develop a heuristic to find the lowest-cost set of locomotive trips while satisfying demand. The timing of the trips is then determined considering the mill operating schedules. Mill labor cost considerations are handled by imposing constraints to ensure that sugar cane arrivals at the mills allow them to operate continuously while open. Equipment is added only if needed. The authors report that the solution provides improvements in operating cost, cane age and capital requirements.

Tables 2 and 3 characterize the major articles reviewed in this section.

5 Conclusions and Directions for Further Research

Our review of many dozens of papers revealed areas where opportunities exist for development of new frameworks and paradigms. An important opportunity lies in the development of *formal* methods for linking decisions at different levels of the hierarchy. For example, aggregate, steady-state scheduling and routing models specify train service frequency and average freight flow patterns, but no systematic procedures exist for making short-term train scheduling and freight routing decisions as demand evolves using the results from the aggregate model as guidelines or constraints. Scheduling and routing models do exist for short-term, discrete-time problems with time-varying demands, but they do not account for any longer-term considerations.

We similarly found few papers that incorporate the effects of short-term and local dynamics upon longer-term, aggregate decision models. Turnquist and Daskin (1982) and Powell and Carvalho (1998a) take important steps, but more research is needed.

Another opportunity lies in extending models to handle system-wide circulation of locomotives and railcars. Nearly all discrete-time, finite-horizon scheduling and routing models terminate in states of the system that may be undesirable in the longer term. Optimization methods can handle minimum cost circulation problems for periodic, repeating, plans, but the problems become much more difficult when demands are not periodic. Likewise, considerable opportunity exists to account for randomness. Most of the models described above are deterministic, and because optimization methods identify “extreme point” solutions, they are often not robust to uncertainty.

Paper/year	Decisions	Major Assumptions	Objective Function	Constraints
Timetabling				
Brännlund et al. (1998)	Detailed train itineraries	-Single track -Fixed train velocity	Maximize value of selected itineraries	-Train capacity on each segment
Carey & Lockwood (1995)	Train paths & schedules	-One-way traffic -Fixed velocity	Minimize deviations from preferred departure time	-Arrival & departure time windows
Sauder & Westerman (1983)	Train timetables, meet-pass plans	———	Minimize total weighted delay	-Maximum train velocity -Siding length
Jovanovic & Harker (1991)	Detailed train timetables, meet-pass plans	-Fixed velocity -Travel times uncertain	(High probability of feasibility)	-Deadlock avoidance -Siding capacity
Kraay et al. (1991)	Train velocity profile & meet-pass plan	-Single track -Variable velocity	Minimize fuel consumption, deviation from schedule	-Maximum train velocity
Blocking				
Bodin et al. (1980)	Car classification strategy for all yards	-No freight priorities -Constant demand	Minimize transportation, classification & delay costs	-Demand satisfaction -Yard & block capacity -Block formation & strategy
Van Dyke (1999)	Assignment of cars to blocks	-Existing blocking plan -Train schedules fixed -Constant demand	Minimize transportation & handling costs	-Demand satisfaction
Newton et al. (1998)	Blocking patterns, assignment of cars to blocks	-Blocks with different priority classes -Constant demand	Minimize mileage, handling & delay costs	-Demand satisfaction -Number of containers classified & blocks built at each rail yard
Kwon et al. (1998)	Car routing	-Train schedules, block definitions, block-to-train assignments fixed -Time-varying demand	Minimize late delivery penalties	-Train capacity -Correct car-to-block, block-to train assignments

Table 2: Literature Classification: Scheduling and Routing, I

Paper/year	Decisions	Major Assumptions	Objective Function	Constraints
Intermodal Routing				
Nozick & Morlok (1997)	Assign equipment & loads to trains	-Fixed train schedule -Known demands	Minimize costs of repositioning & satisfying demand	-Demand satisfaction -Fleet size -Terminal capacity
Scheduling				
Morlok & Peterson (1970)	Train schedule & freight routes	-Deterministic demand & transit times	Minimize fixed, variable operating, storage & opportunity costs	-Delivery windows -Power requirements -Train capacity
Gorman (1998)	Train schedules & freight routes	-Deterministic transit times -Time-varying demand -Limited feasible routes	Minimize fixed & variable operating & opportunity costs	-Freight due dates -Train, yard & line capacity
Newman & Yano (1998)	Train schedules & freight routes	-Deterministic transit times -Time-varying demand	Minimize fixed & variable operating & storage costs	-Freight due dates -Train & line capacity
Huntley et al. (1995)	Train schedules & routes for car "batches"	-Deterministic transit times -Constant demand -No handling costs	Minimize fixed, & variable trans., capital & car rental costs	-Demand satisfaction -Train capacity
Marín & Salmerón (1996)	Train routes & car routing frequencies	-Deterministic transit times -Constant demand	Minimize fixed & variable trans., holding, handling, & investment costs	-Demand satisfaction -Yard, line & train capacity, -Limited cars on track
Integrated Models				
Keaton (1989)	Frequency of direct train service, block allocation, car routes	-Constant demand -Block formation delays ignored	Minimize fixed, variable car use & delay costs	-Demand satisfaction -Yard & train capacity
Haghani (1989)	Locomotive & loaded car routing & empty car distribution	-Constant demand	Minimize trans., congestion, classification, undelivered goods & backorder costs	-Tractive power
Abel et al. (1981)	Train & harvesting schedule, fleet size	-Perishable freight	Minimize operating costs, staling loss & capital costs	-Mill operation -Equipment availability -Train capacity

Table 3: Literature Classification: Scheduling and Routing, II

There is opportunity to develop models with greater realism and solution methods that facilitate more effective implementation. Martland and Sussman (1995) point out that even fairly realistic optimization models must be customized to suit a railroad's operating policies and user capabilities.

Research areas that have become popular in other transportation industries are beginning to touch the rail industry, including demand management, yield management and crew scheduling. Although there are many dozens of articles on optimization in the rail industry, the use of optimization in the industry is young, leaving many doors open for further contributions.

Acknowledgment

We gratefully acknowledge information, comments and suggestions provided by Carl Van Dyke, Multimodal; Ajith Wijeratne and Erick Wickum, Norfolk Southern; Todd McClain, Burlington Northern Santa Fe; Michael Bodden and David Hunt, ALK Associates; Harry Newton, U.S. Air Force; and Carl Martland, Joseph Sussman and Cynthia Barnhart, Massachusetts Institute of Technology.

References

Abel, D.J., K.P. Stark, C.R. Murry and Y.-M. Demoulin. "A Routing and Scheduling Problem for a Rail System: A Case Study." *Journal of the Operational Research Society* 32.9, 1981, 767-774.

Acharya, D., C. D. Martland and J. M. Sussman. "Using Expert Systems and Optimization Techniques for Rail Relay Scheduling." *Proceedings of the ASCE First International Conference on Advanced Technologies in Transportation Engineering*, San Diego CA, Feb., 1989, 378-383.

Ahuja, Ravindra K., Thomas L. Magnanti and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications.* Englewood Cliffs, NJ: Prentice Hall, 1993.

Association of American Railroads web site: http://www.aar.com.

Assad, A.A. "Models for Rail Transportation." *Transportation Research* 14A.3, 1980, 205-220.

Bard, J.F. "An Analysis of a Rail Car Unloading Area for a Consumer Products Manufacturer." *Journal of the Operational Research Society* 48.9, 1997, 873-883.

Beaujon, G.J. and M.A. Turnquist. "A Model for Fleet Sizing and Vehicle Allocation." *Transportation Science* 25.1, 1991, 19-45.

Bodin, L.D., B.L. Golden, A.D. Schuster and W. Romig. "A Model for the Blocking of Trains." *Transportation Research* 14B.1-2, 1980, 115-120.

Brännlund, U., P.O. Lindberg, A. Nou and J.-E. Nilsson. "Railway Timetabling using Lagrangian Relaxation." *Transportation Science* 32.4, 1998, 358-369.

Caprara, A., M. Fischetti, P. Toth, D. Vigo and P.L. Guida. "Algorithms for Railway Crew Management." *Mathematical Programming* 79.1-3, 1997, 125-141.

Carey, M. and D. Lockwood. "A Model, Algorithms and Strategy for Train Pathing." *Journal of the Operational Research Society* 46.8, 1995, 988-1005.

Cordeau, J.-F., P. Toth and D. Vigo. "A Survey of Optimization Models for Train Routing and Scheduling." *Transportation Science* 32.4, 1998, 380-404.

Crainic, T.G., M. Florian and J.-E. Leal. "A Model for the Strategic Planning of National Freight Transportation by Rail." *Transportation Science* 24.1, 1990, 1-24.

Crainic, T.G., M. Gendreau and P. Dejax. "Dynamic and Stochastic Models for the Allocation of Empty Containers." *Operations Research* 41.1, 1993, 102-126.

Dejax, P.J. and T.G. Crainic. "A Review of Empty Flows and Fleet Management in Freight Transportation." *Transportation Science* 21.4, 1987, 227-247.

Ferreira, L. and M.H. Murray. "Modelling Rail Track Deterioration and Maintenance: Current Practices and Future Needs." *Transport Reviews* 17.3, 1997, 207-221.

Florian, M., G. Bushell, J. Ferland, G. Guerin and L. Nastansky. "The Engine Scheduling Problem in a Railway Network." *INFOR* 14, 1976, 121-138.

Genser, R. "Maintenance Planning in Rail Transportation." *European Journal of Operational Research* 11.1, 1982, 35-41.

Gertsbach, I. and Y. Gurevich. "Constructing an Optimal Fleet for a Transportation Schedule." *Transportation Science* 11.1, 1977, 20-36.

Gorman, M.F. "Operating Plan Model Improves Service Design at Santa Fe Railway." *Interfaces* 28.4, 1998, 1-12.

Haghani, A.E. "Rail Freight Transportation: A Review of Recent Optimization Models for Train Routing and Empty Car Distribution." *Journal of Advanced Transportation* 21.2, 1987, 147-172.

Haghani, A.E. "Formulation and Solution of a Combined Train Routing and Makeup, and Empty Car Distribution Model." *Transportation Research* 23B.6, 1989, 433-452.

Higgins, A., L. Ferreira and E. Kozan. "Modeling Single-Line Train Operations." *Transportation Research Record* 1489, 1995, 9-16.

Holmberg, K., M. Joborn and J.T. Lundgren. "Improved Empty Freight Car Distribution." *Transportation Science* 32.2, 1998, 163-173.

Huntley, C.L., D.E. Brown, D.E. Sappington and B.P. Markowicz. "Freight Routing and Scheduling at CSX Transportation." *Interfaces* 25.3, 1995, 58-71.

Jordan, W.C. and M.A. Turnquist. "A Stochastic, Dynamic Network Model for Railroad Car Distribution." *Transportation Science* 17.2, 1983, 123-145.

Jovanovic, D. and P.T. Harker. "Tactical Scheduling of Rail Operations: The SCAN-I System." *Transportation Science* 25.1, 1991, 46-64.

Keaton, M. H. "Designing Optimal Railroad Operating Plans: Lagrangian Relaxation and Heuristic Approaches." *Transportation Research* 23B.6, 1989, 415-431.

Kraay, D., P.T. Harker and B.T. Chen. "Optimal Pacing of Trains in Freight Railroads: Model Formulation and Solution." *Operations Research* 39.1, 1991, 82-99.

Kwon, O.K., C.D. Martland and J.M. Sussman. "Routing and Scheduling Temporal and Heterogeneous Freight Car Traffic on Rail Networks." *Transportation Research* 34E.2, 1998, 101-115.

LeBlanc, L.J. "Global Solutions for a Nonconvex Nonconcave Rail Network Model." *Management Science* 23.2, 1976, 131-139.

Magnanti, T.L. and R.T. Wong. "Network Design and Transportation Planning: Models and Algorithms." *Transportation Science* 18.1, 1984, 1-55.

Marín, A. and J. Salmerón. "Tactical Design of Rail Freight Networks - Part I: Exact and Heuristic Methods." *European Journal of Operational Research* 90, 1996, 26-44.

Martland, C.D. and J.M. Sussman. "Perspectives on Rail Systems Modeling: What Works and What Doesn't." Presented at the Spring 1995 Meeting of the Institute for Operations Research and the Management Sciences, April 1995.

Morlok, E.K. and R.B. Peterson. "Final Report on a Development of a Geographic Transportation Network Generation and Evaluation Model." *Journal of the Transportation Research Forum* 11, 1970, 71-105.

Newman, A.M. and C.A. Yano. "Centralized and Decentralized Scheduling for Intermodal Operations." Working Paper, Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA, 1998.

Newton, H. N., C. Barnhart and P. H. Vance. "Constructing Railroad Blocking Plans to Minimize Handling Costs." *Transportation Science* 32.4, 1998, 330-345.

Nozick, L.K. and E.K. Morlok. "A Model for Medium-Term Operations Planning in an Intermodal Rail-Truck Service." *Transportation Research* 31A.2, 1997, 91-107.

Petersen, E.R. "Railyard Modeling: Part I. Prediction of Put-Through Time." *Transportation Science* 11.1, 1977a, 37-49.

Petersen, E.R. "Railyard Modeling: Part II. Effect of Yard Facilities on Congestion." *Transportation Science* 11.1, 1977b, 50-59.

Powell, W.B. and T.A. Carvalho. "Dynamic Control of Logistics Queuing Networks for Large-Scale Fleet Management." *Transportation Science* 32.2, 1998a, 90-109.

Powell. W.B. and T.A. Carvalho. "Real-Time Optimization of Containers and Flatcars for Intermodal Operations." *Transportation Science* 32.2, 1998b, 110-126.

Sauder, R.L. and W.M. Westerman. "Computer Aided Train Dispatching: Decision Support Through Optimization." *Interfaces* 13.6, 1983, 24-37.

Sherali, H.D. and A.B. Suharko. "A Tactical Decision Support System for Empty Railcar Management." *Transportation Science* 32.2, 1998, 306-329.

Sherali, H.D. and C.H. Tuncbilek. "Static and Dynamic Time-Space Strategic Models and Algorithms for Multilevel Rail-Car Fleet Management." *Management Science* 43.2, 1997, 235-250.

Turnquist, M.A. and M.S. Daskin. "Queueing Models of Classification and Connection Delay in Railyards." *Transportation Science* 16.2, 1982, 207-230.

Turnquist, M.A. and B.P. Markowicz. "An Interactive Microcomputer-Based Model for Railroad Car Distribution." Working Paper, Cornell University, Ithaca, NY, 1989.

Van Dyke, C.D. "Dynamic Management of Railroad Blocking Plans." *Transportation Research Forum* 27, 1986, 116-121.

Van Dyke, C.D. "Algorithm-Based Blocking Plan Analysis." Working Paper, MulitModal Applied Systems, 1999.

Vasko, F.J., J.A. McNamara, D.D. Newhart and F.E. Wolf. "A Practical Solution to a Cargo Loading Problem at Bethlehem Steel." *Journal of the Operational Research Society* 45.11, 1994, 1285-1292.

Ziarati, K., F. Soumis, J. Desrosiers, S. Gelnis and A. Saintonge. "Locomotive Assignment with Heterogeneous Consists at CN North America." *European Journal of Operations Research* 97.2, 1997, 281-292.