Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

EVALUATING STRATEGIES TO IMPROVE RAILROAD PERFORMANCE -- A SYSTEM DYNAMICS APPROACH

Jack B. Homer

Homer Consulting
36 Covington Lane
Voorhees, NJ, 08043, U.S.A.

Thomas E. Keane

Norbridge, Inc.
30 Monument Square
Concord, MA, 01742, U.S.A

Natasha O. Lukiantseva
David W. Bell

CSX Transportation
Jacksonville, FL, 32202, U.S.A.

ABSTRACT

This paper describes the context for and creation of a strategic planning model of a major U.S. railway. The key factors in choosing a system dynamics approach are presented, and the synergy between the new model and the existing suite of planning applications is highlighted. The overall structure of the model is reviewed, including the key reinforcing and balancing loops, and model creation issues such as level of aggregation are discussed. The methodology followed in the data collection and calibration phases is described in detail, and samples of calibration metrics and sensitivity testing parameters are provided, as well as sample model output. Lastly, potential future uses of the model are noted.

1 BACKGROUND

Built largely in the last century, then regulated into financial distress, freight railroads are the stuff of American legend. Since 1980, when the passage of the Staggers Act deregulated the industry and enabled railroads to price competitively, they have gradually reshaped themselves into leaner more cost-effective operations.

As the initial wave of cost-cutting that took place in the 80’s ran its course, railroads have looked towards mergers and acquisitions for further savings and elimination of duplicative overhead functions. As a result, the number of U.S.-based major “Class 1” freight railroads has declined from 35 in 1980 to 12 in 1993 to 5 today. In just the last five years, Union Pacific has acquired first the Chicago & Northwestern, then the Southern Pacific. The two major Canadian railroads, Canadian National and Canadian Pacific, each gobbled up a U.S. line, and Burlington Northern acquired the Santa Fe. Most recently, CSXT and Norfolk Southern jointly acquired Conrail.

Against this backdrop of cost-focused management, the economy improved in the early 1990s, long-haul trucking firms recognized a creeping labor shortage that impacted their ability to retain drivers, and rail intermodal traffic – the placement of truck trailers and ocean shipping containers onto rail flatcars – continued to win supporters. It was time to take advantage of the financial upturn and decade of cost-cutting to invest in growing the business!

But what to invest in? The average Class 1 Railroad locomotive was already 15 years old. Most Class 1’s hadn’t bought freight cars in over a decade, and the average age of the railroad-owned freight car fleet had crept up to be just shy of 20 years old. Track investments in the '80s had been mostly cost-focused, with significant efforts to replace traditional jointed track with welded rail, and extensive programs to replace wooden with concrete rail ties in certain high-stress locations. Few if any locomotive engineers (drivers) or conductors had been hired in the prior decade, and the average age of the workforce at the typical Class 1 had passed 50.

To take volume off the highways and move it by rail (in traditional railcars or as intermodal traffic), any investment would need to be evaluated according to its impact on on-time performance and service. These were the historical Achilles’ heel of railroads and the strength of the trucking industry.

CSXT derived a metric which attempted to measure its reliability and embarked on a series of studies and initiatives to determine what could be done to improve performance. The results were unsettling – no single course of action would have much effect. As soon as one new computer system could be built to address one aspect of the operation or a new batch of locomotives could be acquired, a subsequent bottleneck would surface to prevent the kind of performance improvement that was desired.

Existing analytical tools were unable to offer further insights. It was in this setting, in late 1997, that the decision was taken to try to develop a strategic planning model that might be able to assist the railroad in evaluating various potential investments and policy changes.

2 MODELING APPROACH

2.1 Other Planning Tools

CSXT employs a wide array of operations research (OR) tools for the regular planning and scheduling of operations on a weekly to bimonthly basis. These are used to optimize the use of particular resource subsystems, such as locomotives or freight cars or crew, based on assumptions about the near-term demand for and availability of the target resource, assumptions about the availability and performance of other impinging resources, and current operating policies and standards.

For example, on a monthly/weekly time horizon, the train schedule is developed. In that process, a number of other tools are employed, including an optimal blocking model and a block-to-train assignment model, that are designed to efficiently utilize the existing capacity of the track and terminal network. On a weekly basis, locomotive assignment tools distribute locomotive resources to meet the train schedule.

On a daily basis, a train is built when a block of cars with the same intermediate destination is identified. Crews are assigned to trains when a train is built and locomotive power is available for the train. And, based on constantly changing demand, the dynamic car scheduling system assigns empty cars to destinations every 15 minutes.

2.2 Limitations of Operational Tools

Despite the many and considerable benefits they provided, the limits of these planning tools were acknowledged. In general, the argument is that tools intended to optimize individual subsystems cannot reliably anticipate the performance of the overall system, even when these tools are used in such a way that one tool's results feeds into another tool's assumptions. Thus, CSXT's existing tool set was unable to accurately assess how vulnerable railroad performance was to unexpected changes in resource demand or availability, say, in terms of the likely magnitude and duration of impact on service reliability. Perhaps even more importantly, the existing tools could only partially predict the potential synergies and system-wide benefits that might be derived from enhancing particular resource levels or modifying operating policies. When the whole is greater than the sum of the parts, traditional OR tools are generally inadequate for the task of strategic analysis (Forrester 1968, Homer 1999).

2.3 A Complementary Addition

A number of strategic modeling methods, both deterministic and stochastic, exist for addressing the combined consequences of multiple variables. System Dynamics (SD) is unique among these for its ability not only to project ultimate consequences, but also to anticipate the difficulties – often underestimated, ignored, or unforeseen by other methods – a system may face in "getting from here to there." Such transitional effects are the combined result of a system's feedback loops, nonlinear relationships – e.g., saturation and/or threshold phenomenon, and time delays.

Indeed, it was precisely these aspects of rail network operations that had made CSXT's service performance so intractable to analysis. The dominance of feedback loops and nonlinear relationships is apparent in a discussion of congestion. As one location in the network becomes clogged, locomotives become caught up and thus unavailable to meet their assignments, crews fail to reach their destination within the Federally-mandated on-duty time limit and have to be taxied away from the train to be replaced by relief crews, cars miss their connections, and trains elsewhere begin to be affected.

SD offered a way to investigate relationships between various performance measurements in light of this domino effect of operational problems.

We designed the model to be complementary to existing tools. Thus, as shown in Figure 1, the SD model is at a strategic level – the issues cross over departmental boundaries, the variables are aggregates rather than individual trains, terminals, sections of track, etc., and the time horizon is months or years rather than days or weeks. This is in contrast with the narrower concerns of the existing tools used for subsystem optimization. In this integrated planning framework, the higher level models can be used to establish the ground rules or operating assumptions to be input to the lower level models.

Strategic > 60 Days	System Dynamics Model
Operations Planning 7-60 Days	Locomotive Scheduling Tools Crew Scheduling Model Curfew Planning Tool Car Blocking Block-to-Train Assignment Model Terminal Simulation Model Train Scheduling Tool
Tactical < 7 Days	Dynamic Car Planning System Locomotive Distribution System Crew Assignment Tool Train Dispatching System

Figure 1: The SD Model Complements the Existing Suite of Operational and Tactical Models

2.4 The SD Modeling Process

The SD railroad performance model described in this paper is the product of the combined effort of outside consultants and railroad management and staff. A first version of the model was developed in the first half of 1998, and a second version in the first half of 1999. This sort of sustained and iterative effort is typical of SD modeling efforts. Strategic SD models almost invariably end up raising important new questions which are not adequately answered by existing theory and data, thus generating the need to develop and validate new theory and often new data (Homer 1996).

Several previous applications of SD to transportation system performance modeling do exist, exploring the dynamic interplay of transportation supply and demand (Wright 1978, Stephanede 1981, Gottschalk 1983, Abbas 1990). These models serve as useful background, but none specifically addresses the question of on-time performance, and the various sources of delay affecting it, which was a primary focus of our railroad modeling.

The process of SD model development and validation has been discussed at length elsewhere, but some aspects bear repeating here. To be an effective strategic tool, an SD model must be able to reproduce relevant aspects of past history, but the model must also produce plausible outputs in the face of test conditions, even those that depart radically from historical experience. Also, its equations and parameter values must have a firm foundation in reality, being drawn directly from detailed records or recollections wherever possible, rather than from unconstrained "model tuning" to match history (Forrester and Senge 1980, Graham 1980). A process of partial-model analysis often contributes to the realism and robustness of the model's individual sectors (Morecroft 1985), and can be useful for doing a constrained form of model tuning to estimate parameters when detailed data are not available (Homer 1983).

Every strategic SD model, no matter how sophisticated, contains elements of uncertainty, including imprecise parameter values and unpredictable sources of disturbance, e.g., weather. Thus, strategy options should be considered in the context of sensitivity tests or scenario alternatives that establish not just a single baseline result, but rather a whole range or envelope of possibilities that accurately reflect the magnitude of uncertainty.

2.5 Model Structure

An overview of the model's causal logic is presented in word-and-arrow form in Figure 2. The railroad's operational and financial performance derives ultimately from three things: customer shipment demand, the operating plan for the scheduling and routing of trains and rail cars, and the physical resources available. From an operational standpoint, the focus is on car cycle time, expressed as a number of days from load to load, or in relative terms as on-time delivery percentage. From a financial standpoint, the focus is on earnings growth or simply on unit cost. For a given level of owned resources, cycle time and unit costs tend to rise or fall together, according to the adequacy of those resources to handle customer demand.

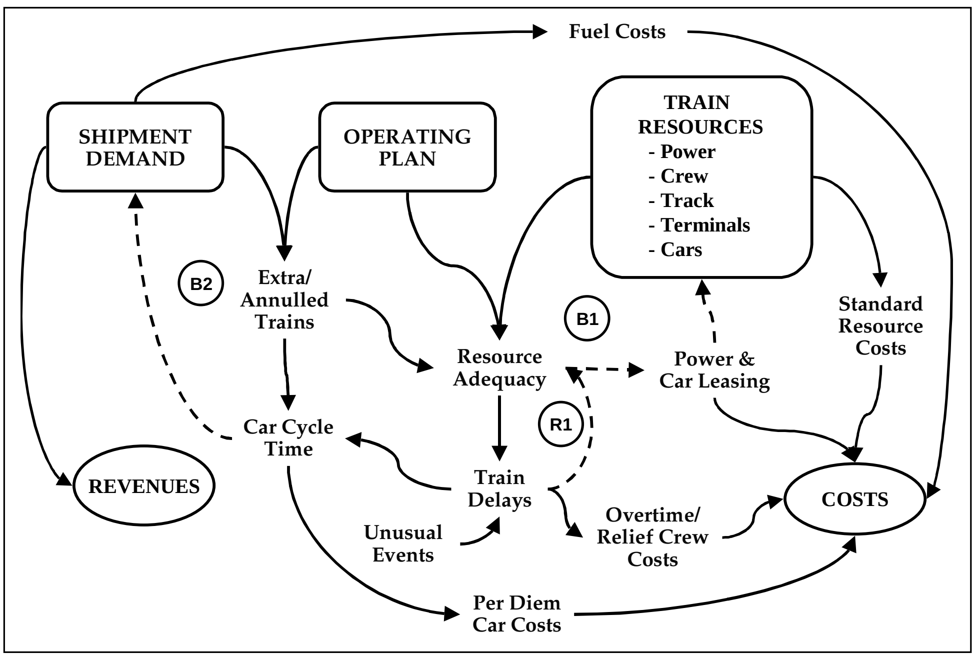

Figure 2: Overview of Model Causal Structure

2.5.1 Train Delays – A Vicious Cycle

Let us take a closer look at car cycle time – one of the key drivers of asset productivity for the railroad. While the planned cycle time is a function of train velocity, distance traveled, and time spent in the terminals and at customers, cycle time variations are driven in large part by train delays. Train delays affect average velocity and also affect the fraction of cars that miss connections due to late arrival (and thus have to wait for the next available train going their way.) Cars may also miss connections when terminals become congested with cars or trains (inadequate terminal resources), or when a car is bumped from its originally scheduled train as a result of too many cars already on board (extra train required) or too few (original train annulled).

Train delays are a chronic problem and have a self-perpetuating or self-reinforcing tendency, symbolized in Figure 2 as Loop R1. Trains absorb more hours of resources (i.e., power/locomotives, crews, track, and terminals) when they are slow or delayed than they would otherwise. This extra absorption of resource hours makes the railroad less flexible than the operating plan assumes – e.g., there are fewer spare locomotives or idle crew available to be used if necessary, and thereby makes the railroad likely to suffer further delays.

2.5.2 Matching Resources to Demand

Train delays may start as a result of unusual events, such as severe storms, equipment breakdowns or accidents, or, more commonly and persistently, when the demand for one train resource or another exceeds or presses too hard against its supply. In the case of inadequate power or too few cars, the railroad has the short-term ability to lease resources to alleviate the problem (see Loop B1 in Figure 2), albeit at additional cost. Note that owned resources are considered a given in the railroad model, because they are determined by annual budgets and cannot be adjusted quickly based on need. Another way that the railroad may prevent or alleviate a demand-supply mismatch is to adjust the train schedule, which is a central component of the operating plan. The intent of such schedule adjustment is to minimize the required number of extra and annulled trains, which not only cause missed connections but also put a disproportionate strain on train resources. These unscheduled trains tend to increase the need for costly locomotive repositioning, may create unexpected crew demand, and may lead to problematic bunching of train arrivals at terminals.

2.5.3 Matching Demand to Performance

If a resource demand-supply imbalance is not corrected through power or car leasing or through train schedule adjustments, and if such an imbalance leads to long car cycle times, then another balancing mechanism may come into play: a limit on actual shipments (see Loop B2 in Figure 2). Potential shipments are considered a given in the railroad model, reflecting basic considerations of feasibility, urgency, and price that determine a customer's standard choice of transportation mode and carrier. But if extended cycle times constrain the railroad's capacity to ship, customer orders may become backlogged and even cancelled, customers may switch to other carriers, and the railroad is likely to lose some potential shipments as a result. From a financial standpoint, this means that poor operational performance may hurt not only costs but also revenues.

2.5.4 Level of Aggregation

Although the SD model represents the railroad in an aggregated way as befits its strategic purpose, a considerable amount of structural detail was required to flesh out the causal relationships shown in Figure 2. The model contains about 450 active equations and is organized into ten interconnected sectors: Shipments, Train starts, Train delays, Car cycle time, Cars, Crew, Power, Terminals, Track, and Financials. The model does not depict individual geographical areas or corridors, but instead describes variables in terms of network-wide totals or averages.

On the other hand, the model does use business segment-specific variables in many areas. We currently have four segments: Merchandise, Unit (primarily coal), Automotive, and Intermodal. These segments have distinct characteristics that result in different levels of operational and financial performance, and that place different levels of demand on the train resources they share. For example, while all merchandise cars (and a fraction of automotive and intermodal cars) make connections and change trains one or more times to get from their origins to their various destinations, cars on unit trains are assumed not to make connections. Another important distinction is that unit trains are unscheduled, and thus put somewhat more strain on power and terminal resources, while all other train types are scheduled.

3 MODEL CALIBRATION

The railroad model contains, in addition to its equations, some two hundred parameters – constants, time series inputs, and nonlinear functions – requiring data collection and calibration. The majority of these parameters were estimated directly from available information sources, often based on straightforward statistical averages and regressions. Indeed, much of the model-building effort was spent in gathering and analyzing data to support the model's structural assumptions and for the purpose of parameter estimation.

3.1 An Iterative Process

Like the modeling process generally, the data collection effort was and is an iterative process. We began, based on a draft version of the model, with a structured database listing all of the desired information, the format it was desired in, the unit of measure, the model sector it would be used in, and likely sources. Every few weeks we would compile a "Hot List" of desired data clarifications and additions, and attempt to resolve as many items as possible. Some required custom query development.

For example, one of the simplifying assumptions in the draft model was that freight cars carrying automobiles (the model's Automotive business segment) did not make connections at terminals from one train to another, but rather traveled directly from origin to destination on one train. Further discussions and observations suggested that this assumption was overly simplistic. The model structure was therefore changed and data collected to estimate the actual number of connections being made by Automotive rail cars.

SD model review discussions quite often reveal that a given variable may be causally influenced by other variables not originally considered. For example, we originally postulated in the model's crew sector that greater train arrival delays would lead to more crew 'deadheads' requiring taxi rides to transport the crews. There are two reasons for this. First, crews arriving late are sometimes unable to be ready in time for their next planned trip, and may need to be transported back to their home terminal. Second, crews that are underway and delayed can be stranded by hours-of-service limitations prior to reaching their assigned trip end points, and then must be removed from the train, replaced by a 'relief crew,' and transported to a rest location. To initially calibrate this assumed function, we compared costs for crew deadheads from a financial database to a custom query run to report train arrival delays.

In reviewing this assumed relationship with members of the project steering group, we learned that crew deadheads can also occur when a terminal or service region is temporarily short of its own regular road crew, requiring other crews to be taxied in from outside the area to help out. From a modeling standpoint, such spot shortages are more likely to occur during certain seasonal periods when road crew are less available than usual. Based on this discussion and additional data analysis, the model now includes two different effects on crew deadheads, one from late arriving crew and one from unavailable departing crew. (These overlay a base amount of deadheads that are simply a result of train imbalances.)

3.2 Partial-Model Calibration

Despite the wide array of data sources at the railroad, we did not always have sufficient information to estimate a particular parameter with the necessary precision. In such a case of missing data, the parameter was adjusted within its range of plausibility during the process of partial-model testing. In this process, a model sector, or portion thereof, is tested in isolation from the rest of the model to determine whether it can produce appropriate outputs in response to historical or other possible inputs (Homer 1983). For example, the model's car cycle time sector contains a nonlinear function that relates missed car connections back to car arrival and processing delays and the planned dwell time per connection. This function was first estimated roughly by considering the logic of normal and extreme conditions. It was then adjusted until the car cycle time sector in isolation, using historical time series on delays and planned dwell time (data feeds from outside the sector), could closely reproduce corresponding historical data on missed car connections (validation data internal to the sector).

3.3 Full-Model Testing

After the completion of all partial-model testing, the model's sectors were linked together to verify that the full model could do a satisfactory job of matching a year's worth of actual history. In fact, the full model produces a very good to excellent fit to history for some two dozen different model variables, in many cases at the level of individual business segments. For example, Figure 3 below shows the fit of average departure plus line-of-road ("LOR") delay hours per train start produced by the model (solid line) with the actual collected data (dotted line). Note that there are some Christmas/New Year (week 51-52) delays in the actual data that the model does not reproduce.

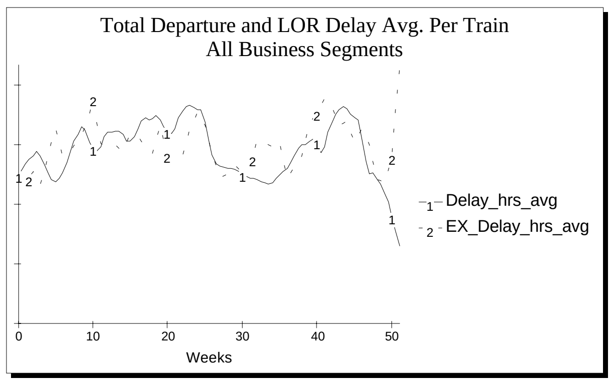

Figure 3: Model Fit for Average Total Train Delay

An example of the full-model and partial-model fit to history is shown in Figures 4a and 4b, where on-time delivery performance is the variable of interest. The full model (Figure 4a) captures the general shape and range of historical behavior, but produces output that is smoother than the actual data, and misses the plunge in on-time performance that occurred around Week 10.

In contrast, running the car cycle time sector in isolation as a partial model produces output nearly as spiky as reality, including the Week 10 plunge (Figure 4b). We can see one of the reasons for the difference by looking back at Figure 3. There we can see that the overall train delays produced by the model are somewhat smoother than the exogenous data. In particular, there are spikes in actual delays around weeks 10-11 and weeks 33-36 that the model does not fully portray. Feeding the actual rather than the model-produced train delays into the cycle time sector results in the difference between Figures 4a and 4b.

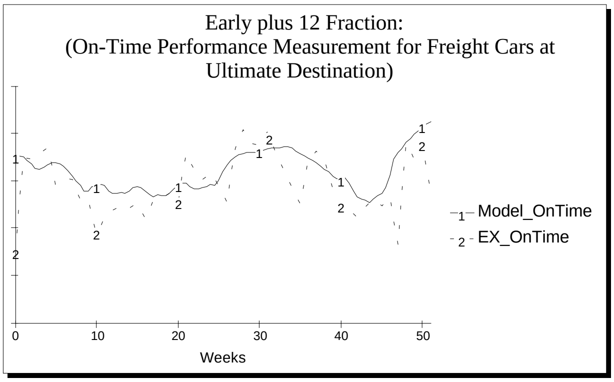

Figure 4a: Full-Model Fit for On-Time Performance

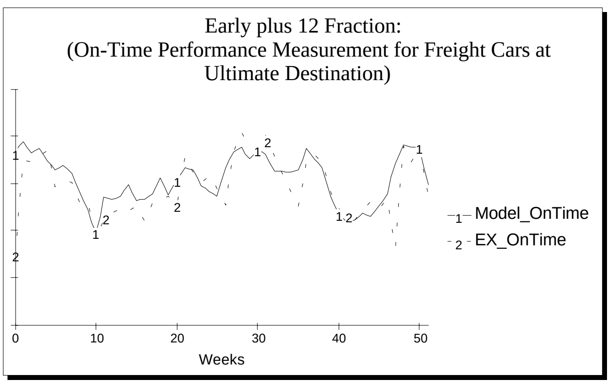

Figure 4b: Partial-Model Fit for On-Time Performance

In general, one should expect a partial model to do better than a fully connected SD model in reproducing the detailed spikes of history and not just the smooth trends. This is because the drivers of behavior in social systems are always a combination of the predictable and the unpredictable, the usual and the unusual. For example, the most important variable affecting on-time delivery performance is train delays. In the full model test, train delays are generated deterministically, whereas in the partial model test, they are taken directly from actual data. As noted previously, train delays originate mostly from resource shortages but also in part from unusual events; the former are mostly predictable while the latter are mostly unpredictable. In looking back at the actual history leading up to the Week 10 performance plunge seen in Figures 4a and 4b, one unusual event not represented in the model stands out: the introduction of a new network-wide program intended to improve productivity at terminals. This program unfortunately led to numerous train delays. As a result, the program was phased out after a relatively short time.

While we have not captured in the model's causal structure the impact of every operational policy change known to have occurred, it is important to understand the factors underlying those data swings that remain outside the model. Only by doing such investigation can one unearth those "usual" factors that repeatedly and systematically affect system behavior, as opposed to those unusual factors that one can never predict.

4 SCENARIO TESTING

Inputs to the model that may vary from one scenario to another include assumptions about potential shipment demand, train resources, and the operating plan, specified over the model's time horizon of one to two years.

Within the SD model proper, potential shipments have both trend and seasonality components that can be applied to specify future demand. In addition, a linked spreadsheet input module provides further flexibility in creating alternative demand scenarios. For example, one may use this spreadsheet to reshape seasonal peaks in sub-segments of shipment demand, such as grain or coal.

One may specify changes in any modeled train resource, i.e., crew, power, terminal, track, and/or cars, but must also to specify the ongoing cost (e.g., maintenance and depreciation) impacts of such changes. In regard to the operating plan, one may change planned train velocity, planned cars per train, planned dwell time per connection, miles per shipment, the number of connections per shipment, and/or the fraction of shipments for which the railroad provides local train service at origin or destination. One may also modify the priorities that affect the way different business segments are treated when train resources are short, which may lead to significantly more train delays for one business segment than for another.

4.1 The Impact of Key Assumptions

Figure 5 demonstrates the impact of a sizeable increase in Merchandise business segment volume beginning in week 22. The increase in this case is set large enough to produce a significant degradation in service. The solid line represents our baseline assumption, as described in section 2.5.3, that some customers will cancel their shipments if service is not sufficiently prompt. While no one ever wants to lose business due to poor service, it was important for us to understand the degree to which cancellations may actually benefit on-time performance during and after a period of resource inadequacy. In Figure 5, such a period occurs during weeks 35-45, after which demand slackens during weeks 46-51 and performance recovers. These graphs indicate that cancellations are quite important during the rebound period, but assist performance only moderately prior to that time. This analysis also underscores the need for sensitivity testing, and the usefulness of presenting an envelope of possible results, as described earlier in section 2.4.

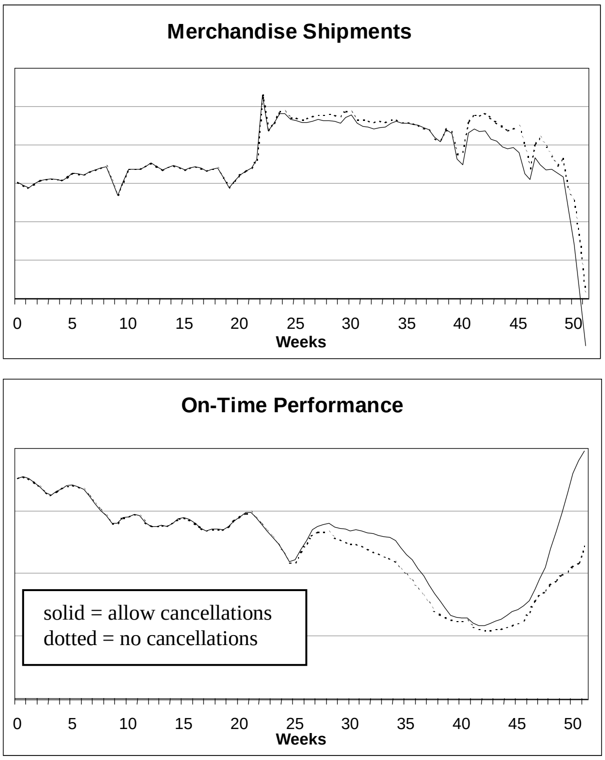

Figure 5: The Effect of Cancellations

5 PLANS FOR ONGOING USE

The SD model has gained acceptance as a strategic planning tool among top CSXT management. Already, the model is being used to probe a variety of management concerns that have arisen as the integration of Conrail with CSXT proceeds. It is also being used in conjunction with the budgeting cycle. Typical inquiries involve predicting the performance impact of various demand scenarios, or evaluating alternative capital investments..

Looking past current planning purposes, we see several potentially fruitful uses for the model. One is as a training tool for operating management. The model is quite good at conveying a sense of how the network as a whole responds to increases in volume and to operational policy changes. Especially for those managers new to railroading or for those who have worked only in jobs that are local in nature, the broad network perspective provided by the model will be instructive.

REFERENCES

Abbas, K.A. 1990. A Road Provision Model Using System Dynamics, *Proceedings of the 1990 International System Dynamics Conference*, Chestnut Hill, MA.

Forrester, J.W. 1968. Industrial Dynamics – A Response to Ansoff and Slevin, *Management Science*, 14(9): 601-618.

Forrester, J.W. and P.M. Senge. 1980. Tests for Building Confidence in System Dynamics Models. In *System Dynamics, TIMS Studies in the Management Sciences, Vol. 14*. New York: North-Holland.

Gottschalk, P. 1983. A System Dynamics Model for Long Range Planning in a Railroad, *European Journal of Operational Research (Netherlands)*, 14(2): 156-162.

Graham, A.K. 1980. Parameter Estimation in System Dynamics Modeling. In *Elements of the System Dynamics Method*, ed. J. Randers. Cambridge, MA: MIT Press. Reprinted by Pegasus Communications, Waltham, MA.

Homer, J.B. 1983. Partial-Model Testing as a Validation Tool for System Dynamics, *Proceedings of the 1983 International System Dynamics Conference*, Chestnut Hill, MA.

Homer, J.B. 1996. Why We Iterate: Scientific Modeling in Theory and Practice, *System Dynamics Review*, 12(1): 1-19.

Homer, J.B. 1999. Macro- and Micro-Modeling of Field Service Dynamics, *System Dynamics Review*, 15(2).

Morecroft, J.D.W. 1985. Rationality in the Analysis of Behavioral Simulation Models, *Management Science*, 31(7): 900-916.

Stephanede, Y.J. 1981. Modeling Dynamic Operating and Financial Strategies in Transportation. In *System Dynamics and the Analysis of Change*, ed. B.E. Paulre. Paris: North-Holland.

Wright, R.D. 1978. Industrial Dynamics Implementation: Growth Strategies for a Trucking Firm. In *Managerial Applications of System Dynamics*, ed. E.B. Roberts. Waltham, MA: Pegasus Communications. Originally published 1971 in *Sloan Management Review*, 12(1).\

AUTHOR BIOGRAPHIES

JACK B. HOMER is an independent consultant specializing in System Dynamics modeling. He received a B.S. in Mathematical Sciences and an M.S. in Statistics from Stanford University, and a Ph.D. in Management from M.I.T. Prior to becoming a full-time consultant, he was an assistant professor of Systems Science at U.S.C. He is a founding member of the System Dynamics Society and the 1997 recipient of its Jay W. Forrester Award for his contributions to the field.

THOMAS E. KEANE is a Principal at the consulting firm of Norbridge, Inc. He received a B.S. in Civil and Environmental Engineering, a M.ENG. in Operations Research, and an MBA, all from Cornell University. Prior to working at Norbridge, he modeled distribution networks from a shipper perspective at Cleveland Consulting Associates, and worked at Mercer Management Consulting. He is a member of the Council of Logistics Management.

NATASHA O. LUKIANTSEVA is a Manager -- Operations Research at CSX Transportation. She received a B.S. and M.S. both in Mathematics from Rostov State University in Russia, and an MBA specializing in Quantitative Analysis from the University of Cincinnati. She is a member of INFORMS.

DAVID W. BELL is an Assistant Vice-President with CSX Transportation in Jacksonville, FL. He joined CSX in 1986 and has held a series of positions in CSX's transportation and marketing departments. In his current position Mr. Bell oversees CSX's operations research function and has recently served a leading role planning the integration of a portion of Conrail into the CSX network. Prior to joining CSX, Mr. Bell was employed at TTX Company and at the Milwaukee Railroad. Mr. Bell received his BA and MBA degrees from Northwestern University.