# Finding the Yellow Brick Road: Part 3, The Wizard of Oz

WARREN B. POWELL

*Department of Civil Engineering and Operations Research*
*Princeton University*
*Princeton, New Jersey 08544*

DONALD E. MAYORAS

*Mau Trucking, Incorporated*
*90 Jacobs Addition*
*Ida Grove, Iowa 51445*

**Sometimes, there is more truth in fiction than in publicly available accounts. This story by Warren B. Powell, a leading researcher of real-time routing and scheduling models and frequent consultant to the motor carrier industry, and Donald E. Mayoras, a senior transportation executive and president of a trucking company, provides a view into the discussions many carriers undertake prior to adopting an optimization model. The account is too long to appear as a single article and instead is being published in serial form. This is the third installment.**

The cast: Dan Manning, President; Tom Gorman, Chief financial officer; Matt Peterson, Vice-president of operations; Ken Richards, Vice-president of sales; Bill Johnson, Vice-president of management information systems; John Breswick, Director, driver management; Michelle Corwin, Director, customer service; Larry Michaels, Director, driver dispatch; and Walter McCormick, University professor and industry expert.

*Dan Manning has concluded a meeting of his management team in which they discussed a variety of problems the carrier was facing. The company has been struggling to improve profit margins, with problems ranging from driver turnover to regions where they could not seem to balance the amount of freight going in and out. The discussion moved toward technological solutions, ranging from expensive satellite communication systems that would provide close contact with drivers to advanced optimization models. In view of the expense of satellite communications, Dan gave the nod to look into some of the new optimization systems that had become available.*

*Four weeks later, the same group has reconvened to hear a presentation by Professor Walter McCormick, an expert in advanced technologies for motor carriers. At Tom Gorman's request, he has come to Allegheny Motor Carriers to talk about satellite communications and optimization models for truckload motor carriers. Tom is hoping that Walter will help con-*

Copyright © 1997, Institute for Operations Research and the Management Sciences
0092-2102/97/2701/0143$05 00

INDUSTRIES—TRANSPORTATION
DECISION ANALYSIS—APPLICATION

INTERFACES 27: 1 January–February 1997 (pp. 143–154)

Copyright © 2001 All Rights Reserved

*vince Matt and Bill to implement an optimization model for dispatching. John and Ken want to learn what it takes to implement a satellite communications system. Dan is in favor of anything that will halt the company's earnings decline but is doubtful that technology alone will solve the problem. The group looks at the speaker at the end of the table, who is arranging his slides next to the overhead projector. Dan stands up to give the introduction.*

"I appreciate everyone coming back so soon after our last quarterly meeting. Today Dr. Walter McCormick is with us to help us sort through these new technologies. After his talk, you will have time to ask questions."

Walter positioned his first slide on the projector and turned to the group.

"I appreciate your inviting me here to describe the exciting advances that have been made recently in information technologies for the motor carrier industry. I would like to discuss two types of technologies that are having an impact on truckload motor carriers; hardware-based and software-based technologies. Hardware-based technologies allow you to collect data and communicate with your drivers. Software-based technologies effectively allow you to use the enormous volumes of data that we now have access to.

"There are three principal technologies for collecting and transmitting data. The first is mobile communications, which allow two-way communication with a truck on the road. A number of technologies have been tested over the years, including one that bounces signals off bursts of meteors in the atmosphere. The one that has emerged as the dominant technology, at least for the moment, uses a radar unit on the top of the cab to communicate via satellites back to a central facility. The driver and dispatcher conduct two-way communications via a unit in the cab that can display short messages. Although this technology doesn't permit voice communication, it allows the driver to send requests and receive instructions. The carrier is also able to obtain information about the condition of the truck, including fuel, oil, speed, and even tire wear.

"Satellites are not the only way to communicate with drivers, although at the moment this is the only proven technology for long-haul trucking. Another form of mobile communication relies on ground-based cellular telephone networks. The drawback of this approach is that there are gaps in coverage, especially in remote areas away from metropolitan centers. We are not sure how this technology will evolve. Some carriers are still sitting on the sidelines waiting to see how it will shake out.

"The second technology is automatic vehicle location, or AVL. By bouncing signals off of three satellites, a driver can determine the exact latitude and longitude of the vehicle. This information is not very useful to the driver, but by using mobile communications, the driver can transmit his location back to the carrier. AVL, combined with mobile communications, allows carriers to track their entire fleet at all times. For some applications, you can combine AVL with in-vehicle navigation systems, which provide drivers with specific routing instructions. This can be helpful for driving around metropolitan areas, especially with drivers who do not know the area. But these navigation systems do not know about traffic conditions—they work off of



Copyright © 2001 All Rights Reserved

static road networks.

"The third technology, which is extremely young and will not be fully developed for at least a decade, is advanced traffic information systems or ATIS. The idea here is to provide information to drivers on accidents and congestion so that they can choose the best routes. The appeal of this type of information is pretty obvious for certain areas of the country, but we will have to wait a few years before making any decisions about these systems. The problem is that we simply don't have the road sensors to detect traffic conditions. For the moment, let's stick with what is available now.

"I would like to say a few words about software technologies. Mobile communications and vehicle location can produce a tremendous volume of data that is difficult to use effectively. Also, two-way communication allows you to send instructions to drivers, giving you much more control over your fleet than you ever had before. But now you have to ask, can you actually take advantage of all this data?

"The answer to this is optimization models, which are software-based technologies that have grown out of the field of operations research. Optimization models are useful when we are trying to improve productivity but are juggling a lot of balls at the same time. Optimization can help a truckload motor carrier in two areas.

"First, in sales and marketing. Several carriers are using a special type of optimization model to simulate their operations and determine the profitability of lanes and shippers. These network models can help a carrier balance its network, identify its most profitable lanes, and weed out shippers who are not pulling their weight. They usually run these models quarterly and use them to find out where they are overbalanced and which lanes are losing money. They then try to adjust their pricing or their allocation of their capacity. These models also tell you how to allocate your sales efforts to get the maximum reward. Some systems can also tell you how large your fleet should be. If you have questions about this type of capability, I can answer them later.

"I'd like to turn now to optimization models for real-time dispatching. These have been around for over a decade, but only a handful of carriers use one. To see the basic idea behind these models, assume we have a region with four drivers and six loads, and we have to find the best assignment of drivers and loads. For the moment, assume that we are concerned only with minimizing empty miles."

He put up a transparency that illustrated the situation.

"Here we have a problem of assigning four drivers to cover six loads. Remember that your dispatchers can't see this picture, and the real problem is much larger. By working on one driver or load at a time, you might come up with this solution," and he put up a slide labeled "Manual solution (Figure 1)." "As you can see, we have four drivers, but we have covered only three loads. We put driver B on Load 1, and now driver A can't get to anything in time."

He took down that transparency and put up another labeled "Optimal matching" (Figure 2).

"If you buy a commercial load-matching package, you could easily get this solution. These packages use a network solver to



Copyright © 2001 All Rights Reserved

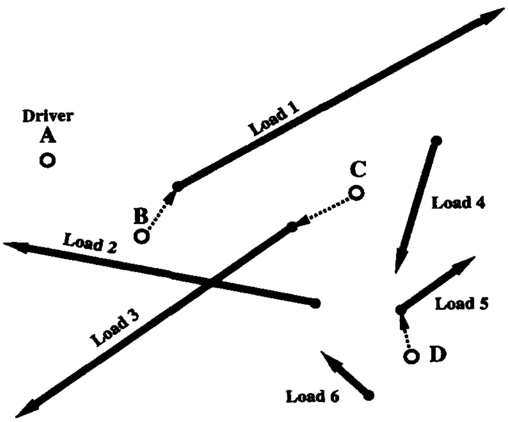


**Figure 1: Manual assignments, made over the course of the day, can leave loads uncovered with unassigned drivers.**

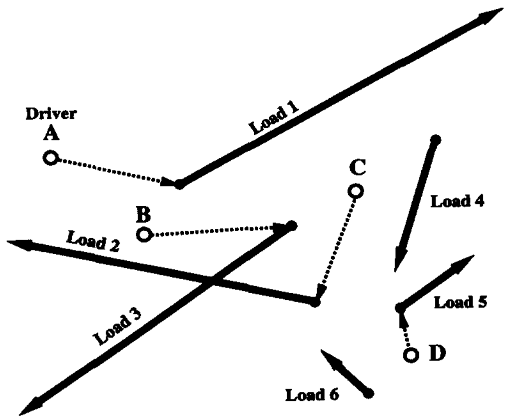


**Figure 2: An optimal matching system will match each of four drivers to one of the six loads available.**

find the optimal match of drivers and loads. In this case, the system is smart enough to know that it is better to put driver A on load 1, and driver B on load 3. Driver C is switched to load 2, probably because it is longer and more profitable."

He paused as everyone looked over the solution and then continued. "The limitation of optimal matching systems is that they can assign only one load to each driver, so our four drivers are covering only four of the loads."

"But once you assign each driver to a load, you can block the assignment and let the model assign the driver to a second load," suggested Matt. "Besides, we don't like to assign a driver to two loads. Something may change, and we may want to change his assignment."

"That's right," responded Walter, "but you still have to plan the entire tour even if you want to do only one assignment. Just because you assign the driver to two loads, doesn't mean you have to tell him or even commit to it."

He turned back to the slide. "Look at what would happen if you locked in these assignments and then considered the next assignment of each driver. Driver D could handle load 4 and, if you're lucky, could then cover load 6, if there was enough time."

"What's wrong with that?" asked Tom.

"Well, let's take a look at an even better answer," and he switched to a third slide, labeled "Optimal tours" (Figure 3). "In this solution, the system was smart enough to know that driver D should first do load 6, which is short, and then pick up load 2. This allows driver C to cover both loads 4 and 5."

"How is that better than our first solution?" Tom pursued.

"If you use the first solution, you have to cover loads 4, 5, and 6 with one driver. Even if this were possible, you would have to drive empty from load 5 to 4 and then from the drop off of load 4 down to load 6. In my answer, you have the empty only from the end of load 4 to the pickup of



Copyright © 2001 All Rights Reserved

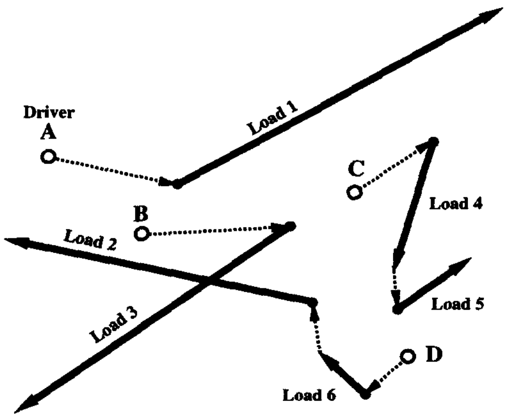


**Figure 3: Optimal tours assign the four drivers to all six loads.**

load 5. In most cases, the savings in empties might be only five or perhaps 10 percent, but that adds up to a lot of money."

Everyone paused to think about it. Matt had to admit that he didn't think his planners would find this answer, especially under pressure. They just didn't pay that much attention to short empties as long as they were covering loads.

"Do the commercial packages give us the right answer?" asked Bill.

"The most popular packages marketed today use an optimization technology called networks. A network is a special type of linear program that can be solved very quickly. Also the codes to solve these problems are extremely simple and widely available. The down side is that they can solve only special types of problems. If you want to match drivers and loads, with at most one load per driver, that's a network. If you want to find tours, that's not a network. To solve these problems optimally, you have to go to a different class of solvers, but they are much slower and, at the moment, can't be used in real-time applications."

"So how do other companies solve this problem?" continued Bill.

"For the most part, they follow Matt's first suggestion. They assign a driver to one load and then take where he is at the end of the load and assign him again. This approach can give good answers under the right circumstances but can also give answers that are not only well below optimal but worse than what a dispatcher could achieve working manually."

"You mean a human being can outperform a computer?" Tom asked incredulously.

"When you have a problem for which you can guarantee a mathematical optimum," Walter explained, "the human beats the computer only because he knows things the computer doesn't. But most practical problems are so complex that we just don't have algorithms that produce the mathematical optimum fast enough, even when the data is perfect. Assigning a driver to a single load is one of the few problems that can be easily solved to optimality. Virtually all other problems are problems for which we can find a mathematical optimum but not fast enough for real-time applications. Algorithm designers have to find methods that provide the best solution they can find within five to ten seconds."

Walter paused again and then put up another slide.

"Unfortunately, all of these pictures represent pretty simple problems that computers can handle. The real problem is much more complicated." He pointed at the slide labeled "The real problem" (Figure 4). "The hard part about a dispatch problem is not what you see but what you don't see.



Copyright © 2001 All Rights Reserved

In practice, a planner might hold off on assigning driver B to load 3 because there is a shipper close by who calls in every morning with two or three loads. Since this is an important shipper, he is going to hold driver B for another hour or so, until he knows what this shipper is doing. If he needs driver B for this shipper, he'll shift driver C onto load 3, and then call up another off-duty driver that the computer doesn't even know about. But the dispatcher might know that this driver is just taking some personal time and can be called on in a pinch. Driver D is also being held for the same reason, and if he is needed for a new load, the dispatcher will probably try to route driver D to load 5, after which he can handle one of the new loads, while whoever handled load 4 will now have to come down and do load 6 and then load 2. Most likely, a shuttle driver will grab load 2 and bring it into the terminal, since it has some slack in the delivery appointment."

Bill had been leaning back in his chair staring at the ceiling and now leaned forward. "I am trying to figure out how this system would work. If the optimization works on the whole country at once, do you have a single user running the model and telling everyone else what to do? We have 22 customer service reps and 20 dispatchers running the company. How do they use the model?"

"That's a very good question." Walter responded. "An optimization model outputs two sets of answers. The first is called the 'primal solution,' which is what load a driver is assigned to. The second answer is called the 'dual solution,' which gives the

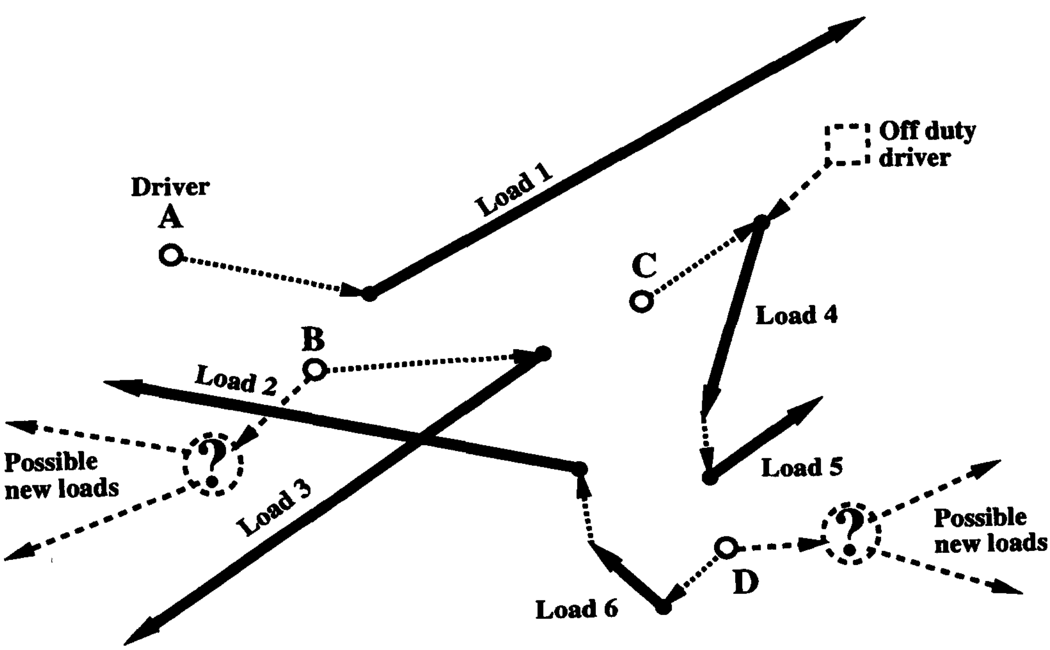


**Figure 4: The real problem is much more complicated, including factors that are not immediately obvious or known.**



Copyright © 2001 All Rights Reserved

value of assigning a driver to a load, taking into account the impact such an assignment has on the entire network. You can think of the primal solution as the answer, and the dual as the reason behind the answer. In a production environment, you output the dual solution to a database. Then, each user can reconstruct what the answer should be, without actually knowing the answer. Even better, it is possible to know not only the best assignment but the second best, third best, and so on.

He looked around and saw the confusion on everyone's faces.

Walter put up a picture of a set of drivers and a set of loads (Figure 5). He started jotting numbers alongside each driver and load.

"After the model has optimized, it produces a set of numbers, one for each driver and one for each load. Let's call these the 'values' of each driver and each load. The research community calls them dual variables, but the reason for this is pretty complicated. In my example, I'll draw the best assignment of each driver to a load using a solid line . . ." he drew several lines on the drawing, ". . . and I'll draw other possible assignments using dashed lines. I'm using two different types of dashed lines, but that's just so I can keep track of things. Notice that driver A is only 12 miles from load 3, but we have assigned him to load 1, which is 18 miles away. Let's see what would happen if we assigned him to load 3 instead. This would cost us an extra 12 miles but save us the 18 miles to load 1. Now, we'd have to take driver C off load 3 and put him on load 4, for an extra five miles. Next we would have to take driver D off load 4 and put him on load 1, for an extra 20 miles. Now all the loads are still covered, but we have run . . ."

He added up the numbers, . . . "an extra 19 miles overall. You would think that to find this number, you would have to optimize the problem all over again. But we have a shortcut. There's a formula we can use. Take the cost of assigning driver A to load 1, 12 miles, plus the value of driver A, minus the value of load 1. This is 12 plus 22 minus 15, or 19 miles. So, with this little calculation, we can find the impact of reassigning driver A on all the other drivers in the system. I like to call this the total system contribution."

Walter stopped for a minute and looked around the table. The puzzled looks were not encouraging.

"The important thing to realize is that after the optimization, all we do is store these value numbers for each driver and load. Then, your application package can read in these numbers and evaluate the value of assigning a driver to a load. You don't really even need to know what we think is the best assignment. For example, let's say you know just these value numbers, and you want to know what to do with driver A. Looking at loads 1, 2, and 3, you would get total system contributions of . . ." he jotted the calculations down on the overhead.

Load 1: $18 + 22 - 40 = 0$
Load 2: $25 + 22 - (-15) = 62$
Load 3: $12 + 22 - 15 = 19$

"So, we would choose load 1, with load 3 the second choice, and load 2 the third choice. Notice the big cost of assigning driver A to load 2–62 extra miles. You wouldn't know that just looking at driver A, but when you consider the impacts on all the other drivers, it's pretty expensive.



Copyright © 2001 All Rights Reserved

Most important, the calculation is pretty simple. You can do it using your own application package."

Bill's face lit up. "You mean, this model could run as a black box, punch out a bunch of numbers, and then we would do all the calculations?"

"That's right. And you may not want to use the optimization's choice of the best assignment of a driver to a load for reasons that only you know. But at least you will know the impact of your decision on all the other drivers. And you can do it . . ."

"That means we can evaluate other possible assignments without having to call the optimization model." Bill finished Walter's sentence. Turning to the rest of the group, he added. "This way, the optimization can run, give us these value numbers, and then I just have to prepare screens that rank loads based on this formula. That's even faster than what I do now, since I have to do a lot of distance calculations to find the closest loads."

John had listened carefully to the entire presentation, and while he didn't follow all of it, he had picked up what the optimization was trying to do. But there were still some big hurdles, and he was worried by Bill's enthusiasm. "Walter, I see how the optimization will allow us to see what assignments will best minimize empty miles, but that's not the most important part of driver assignment. Every one of those drivers is a human being with different needs and priorities. I am sure that it is nice to know what assignment gives us the fewest empties, but we might want to assign a driver to the fifth or sixth closest load, since that is the right load for that driver. I can't see how this system would be useful."

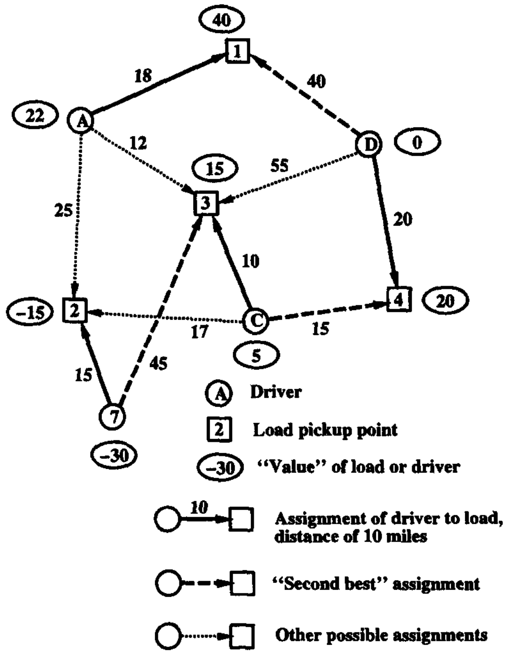


**Figure 5: An optimization model produces two sets of answers: the primal solution—what loads drivers are assigned to—and the dual solution, which gives the value of assigning each driver to each load, taking into account the impact of the assignment on the entire network.**

Walter responded patiently. "You're absolutely right, and the system will be useful only if you are using the best assignment most of the time. I used miles in my example, but in practice you would use dollars, and these dollars would include various bonuses when an assignment achieved a certain objective, like getting a driver home. Of course, there are a lot of issues." He put up a new transparency (Figure 6). "I like to call it the balancing act. You have to get drivers home, meet shipper commitments, minimize empties, and get tractors into maintenance."

"You mean, if your load 2 there got driver A home, we could put a thousand



Copyright © 2001 All Rights Reserved

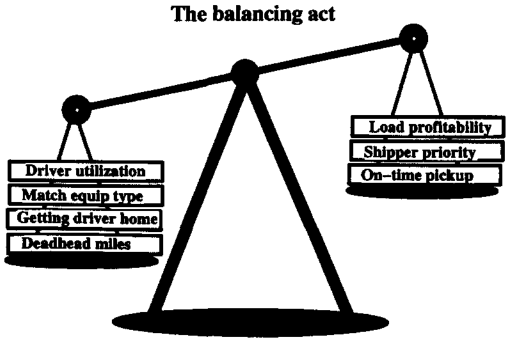


**Figure 6: The assignment of a driver to a load involves much more than minimizing empty miles—it requires taking into account a number of soft issues that reflect driver satisfaction and customer service.**

dollar bonus on that assignment?" John interjected.

"Yes, and that would certainly force the model to assign driver A to load 2. But that means that, at a dollar a mile, you would be willing to literally drive a thousand miles to get that driver home. That's your choice, but you need to be sure that is what you really would do. More realistically, think about how many extra miles you would be willing to drive to get this guy home. And you don't pick this number on the fly, you let your people set it, and you review it a couple of times a year. That way, you make those trade-offs on a consistent basis. If you decide over time that you are not getting drivers back home, then you can increase the number of miles you are willing to run to achieve that objective."

Matt broke in. "So, we really have to put a value on some of these choices."

"That might be a good exercise for us." Dan added.

Michelle had been sitting quietly trying to absorb the discussion and didn't want to ask a dumb question. But there was something nagging at her about the types of questions being asked. She took advantage of a lull in the conversation.

"We spend a lot of time thinking about what shippers will do. How do these models take into account the loads that haven't been booked into the system?"

Walter's eyes lit up. "That's an excellent question, and this is where the different modeling systems really differ. The simplest strategy is to just ignore forecasted loads and let the system react to them as they come in."

"Does that work?" Michelle asked skeptically.

"Not really. You might have a driver sitting at a shipper's yard, where he knows a load will be moving out in a few hours. But because it hasn't been booked into the database, the optimization model might assign that driver to another load near by. This can look really dumb to a driver or knowledgeable planner.

"The next step up is to forecast loads. This means combining actual booked loads with forecasted loads."

"You mean, we have to be able to predict the loads that are going to be booked?" Michelle asked.

"That's right." responded Walter.

"But how do we know what they are going to do? We can't tell who is going to book with us, except for the major accounts, and even for those accounts we don't know how many or where they are going?" Michelle sounded a little exasperated.

"You're absolutely right," Walter agreed. "That is what we call a deterministic model, and it has a lot of problems. We call it deterministic, because we assume that all



Copyright © 2001 All Rights Reserved

the forecasts are exactly right. These models really have a problem with accounts that book one or two loads per week, or an average of, say, 0.2 loads per week. We can't put two tenths of a load in these models, unless we want to assign two tenths of a driver to the load.

He paused and looked around and then continued. "The next step is a stochastic model."

He stopped abruptly when Larry burst out laughing, finally asking, "What in the world is . . . stochastic?"

Walter thought briefly and then offered an explanation. "Stochastic is like random. When shippers call in randomly over time, we call these phone calls stochastic."

Larry didn't want to go down this road and quickly nodded his satisfaction with the answer. Michelle's interest, however, was climbing.

"What does a . . ." she paused to check herself, ". . . stochastic model get you?"

Walter answered patiently. "A stochastic model recognizes that the demand forecasts are not perfect. If a shipper books an average of, say, two loads per day, a stochastic model understands that on any given day, those two loads might be anywhere between zero and five. A stochastic model estimates probabilities for each possible outcome and then takes an average."

Michelle continued to press Walter. "I'm going to assume that this is a better model. Why wouldn't we use this approach?"

"The mathematics behind these models becomes very difficult, and they can't be solved in a production setting without making some approximations." Walter explained.

"Such as?"

"Well, one popular model lumps all forecasted loads in a single region together. In this approach, you would forecast, say, the total loads in northern Illinois, and a separate forecast for the total loads in Ohio."

"But how can you do that?" Michelle objected. "The accounts are all different!"

"I know, that's the problem."

Frank jumped in at this point, and asked, "But why can't you just use small regions?"

"Good question," Walter nodded approvingly. "The problem is that once you lump loads into these regions, the model assumes that a driver in one region can't handle a forecasted load in another region. That approximation isn't too bad if the regions are big, but becomes a real problem if they become small."

Walter looked around the room, and everyone's faces looked downright glum.

Ken tried to summarize, "So, what is the best approach?"

Walter tried to be as honest as he could. "The simple answer is, the jury is still out. The topic is an active area of research. A model that does no forecasting can still be useful, as long as you understand that you have to do the forecasting yourself, and override the model when necessary. However, there are some questions you can ask of a system. For example, how does the system adjust forecasts during the day as new loads are booked? Are forecasted loads lumped into regions, or can the system forecast individual accounts? And the hardest question of all: can it forecast capacity?"

"Why do we need to forecast capacity" Michelle took the bait.

"Remember, you are not interested in



Copyright © 2001 All Rights Reserved

forecasting demand, you are interested in forecasting whether you have too much or too little capacity for the demand. This means, you have to forecast both demand and capacity. It is the difference between the two that counts."

"Isn't that just the forecasted demands minus the number of trucks in the region?"

"Remember, you are not forecasting loads now, but in the future. You might need to forecast loads tomorrow or the next day, which means you have to forecast the number of trucks that will be available tomorrow or the next day." Walter explained.

"But don't we just look at how many trucks will be in a region tomorrow or the next day?" Michelle asked.

"How do you know how many trucks are coming into a region?" Walter challenged her. "You know what trucks have been dispatched, but what about the ones that haven't, but which will? And what about the trucks that will be held over in a region? And what about the trucks that will move empty in the future?"

Michelle became a little exasperated. "But we can't predict all that! Why try to forecast empties that we don't even know will happen?"

"Because you might have a lot of loads terminating at an account in Milwaukee, after which they always move empty to Chicago to be reloaded. If you don't forecast these moves, you will always think that Milwaukee has too many drivers, and Chicago too few. If you don't forecast empties, you might think you have booked too many loads into Milwaukee, when in fact Chicago desperately needs those drivers."

Michelle nodded in satisfaction. The others mostly looked relieved. Dan tried to steer the conversation back to more down-to-earth issues.

"Walter, if you implement an optimization model, what percentage of its recommendations could we expect to implement directly?"

"I hate to be vague, but it really depends on how much you tune the model," Walter responded. "If all you do is focus on empty miles, you might get only about 50 percent. It's a useful tool, but if you override the model that much, it becomes just one more report people have to look at. It also means that you really have to review every dispatch, which means you are not saving much time."

"What's the best you can expect?" Dan pressed on.

"There is no reason you can't get a 90 to 95 percent acceptance rate," Walter asserted.

"What's that take?"

"Three things. First, you have to lay out your priorities and trade-offs clearly. You have to decide the relative importance of getting drivers home, serving loads, meeting mileage targets, and so on. Management must participate closely. Sometimes you may not even be aware of some of the things you are doing until a model starts doing them differently. Developing a good model is an ongoing process—it never really stops."

Walter looked around the group and continued. "The second is data. The computer can't make good decisions without good data. Models are basically blind—they see what you feed them. It takes a lot of discipline. Typically, people take shortcuts with the computer, since they keep



Copyright © 2001 All Rights Reserved

some information in their heads. That often meets with some resistance, since they have to do extra work."

"And the third?" Dan pressed.

"You have to have a model that is smart enough to use this information. People are smart, but they can't handle a lot of data. The computer is only as smart as your model, and a simple model makes a stupid computer."

"How do we know if the model is smart?"

Walter flipped off the overhead projector. "There are three levels of testing. First, talk to the people who have to implement the system. Ask them what issues the model considers. Ask them what data it needs. Do they forecast shipper needs? Do they need to know about driver domiciles, and how long the driver has been out? Do they track trailer pools? What information do they need about the shippers? And listen to the types of questions they ask you about your organization. Does it seem as if they need to know how you run your company?

"The second level is a demonstration. Set up a test situation, and see how the system would assign drivers. Try out different scenarios.

"The third level is the implementation. This is the real test. You should be able to get references from other carriers to see how the system is working with their operation. But ultimately, the real test is with you."

Dan looked around the room and then turned back to the speaker. "Walter, thank you. I think we need to think about the issues you have raised." Turning to the rest of the group, "Let's meet back here tomorrow at 8 AM to discuss our options." He thanked everyone and walked back to his office. Tom, Matt, and Bill stayed in the room to ask Walter some more questions.

(To be continued)



Copyright © 2001 All Rights Reserved