Interfaces
Vol. 40, No. 4, July–August 2010, pp. 291–302
ISSN 0092-2102 | EISSN 1526-551X | 10 | 4004 | 0291

DOI 10.1287/inte.1100.0495

Capacity Management on Long-Distance Passenger Trains of Indian Railways

Raja Gopalakrishnan
Western Railway, Churchgate, Mumbai 400020, India, gopalakrishnan1975@gmail.com

Narayan Rangaraj
Industrial Engineering and Operations Research, IIT Bombay, Powai, Mumbai 400076, India,
narayan.rangaraj@iitb.ac.in

In this paper, we discuss a model to allocate train capacity among multiple travel segments on an Indian Railways train route with several stops. We use a linear programming model and data preprocessing and postprocessing to determine the optimal capacity allocation on multiple travel legs. The model uses a simple, effective capacity management tool. For historical and social reasons, Indian Railways splits its train capacity based on user and type of travel. Determining the optimal split of such capacity is nontrivial. We deal with the specific issue of the spatial split of seats in an end-to-end travel segment within a given travel class. This addresses the needs of end-to-end and en route travelers. When we applied our model to 17 Indian Railways trains, we achieved increases of 2.6 to 29.3 percent in revenue, 6.7 to 30.8 percent in load factors, and 8.4 to 29 percent in passengers carried.

Key words: rail transportation; linear programming, applications; capacity management; revenue management.
History: This paper was refereed. Published online in *Articles in Advance* May 24, 2010.

Indian Railways (IR) operates more than 1,600 long-distance and intercity trains and carries more than 7 million passengers daily (Indian Railways 2007). IR trains include three classes of coaches—air-conditioned (AC) reserved, non-AC reserved, and non-AC unreserved. Reserved tickets are booked through IR's passenger reservation system (PRS), which reserves a specific seat per booking.

The problem we address in this paper is that of allocating seat capacity in a given class of a train to multiple travel segments, including segments on which en route passengers (i.e., those who are not traveling end-to-end) travel. The significant travel requirements of en route passengers, including those who travel to and from intermediate stations, make this a critical issue for IR. We confine our study to reserved seats booked through PRS.

The basic problem is part of the revenue management of a network of products or services. This paper analyzes specific data requirements for such an application and reports the tangible benefits accrued, i.e., increases in numbers of passengers, load factors, and revenue. The underlying model is the deterministic linear programming (DLP) formulation of the capacity-allocation decision (Talluri and van Ryzin 2004); however, our model has a different objective. Similar to the work of Ciancimino et al. (1999), we describe a rail-operation application in which the model yields tangible benefits. Hersh and Ladany (1978), Glover et al. (1982), Dror et al. (1988), and Rozite et al. (2005) address similar problems. Most of these works deal explicitly with demand variability as part of the model; however, they have revenue objectives and (or) have a limited number of origins and destinations in their transport networks. We propose a different optimization objective—one that is consistent with IR's obligation as a public carrier with service provision as an objective. Within this framework, our approach has enhanced IR's capacity management and revenue. We also discuss the detailed considerations of preprocessing and postprocessing the IR-specific data required by this basic model.

The problem of allocating capacity in multiple-travel-leg situations is seen in rail travel and in air transport and is conceptually similar to the production planning-problem in which common resources produce a variety of end products.

We have organized the rest of this paper as follows. *Segmenting a Train's Capacity* addresses understanding how IR segments the capacity of its trains. *Implementation* includes the main linear programming model and its data requirements and constraints; this section also includes a description of the model's implementation. The *Results and Discussion* section describes the results of using the model on various IR trains and also includes a detailed example. In the *Conclusion* section, we discuss some directions for future work. Appendix A describes the various seat patterns that are possible on a train and the objective of maximizing the number of passengers with confirmed accommodations that the train can carry. Appendix B lists the details of the PRS and the various quotas.

Segmenting a Train's Capacity

A train can make many stops en route from its origin to its destination. Consequently, passengers can buy many combinations of station-to-station tickets. Tickets are booked both from a train's origin to its destination and to or from intermediate stations on its route. Because passengers detrain and board en route, multiple passengers can occupy a given seat during a train's complete journey. The seat might also be vacant for some segments. A vacant seat does not earn any revenue. Whenever a seat to or from an intermediate station is booked, a partially vacant seat is created. In the event of an abnormally high booking for intermediate station(s), many partially vacant seats are created; although these might not be filled, subsequent attempts by passengers to book end-to-end trips could be denied, thus leading to suboptimal utilization of the train's capacity. In addition, heavy cancellations of intermediate tickets near the date of a trip could cause additional losses. Conversely, if the proportion of seats meant for end-to-end trips increases, it may cause denial of confirmed seats to or from intermediate stations; seats are confirmed against the vacant end-to-end seats only at the time of final manifest preparation (called charting in PRS) at the origin station. Manifest preparation is the process of final assignment of all available seats to passengers by confirming some or all of the train's wait-listed passengers by assigning any vacant seats to them; this process uses quotas other than the quotas by which these passengers were booked. A passenger who is denied a confirmed seat at the time of booking might opt to use another mode of transport. In the event of lack of demand for end-to-end seats, the train could theoretically run empty.

In specific instances, because of operational constraints such as terminal capacity or the need to connect all parts of the country using rail services, trains might start from terminals at which the demand is low. For example, on an IR train that connects the important temple towns of Okha on the west coast and Puri on the east coast of India, the majority of the passenger traffic is to or from the major en route cities (e.g., Ahmedabad, Surat, and Nagpur). In such instances, allocating a substantial end-to-end seat quota is detrimental because most train traffic originates at intermediate stations; thus, the possibility that passengers requesting tickets at intermediate stations will be wait-listed would be high. Therefore, a train's capacity must be distributed among various intermediate stations by allocating specific quotas to ensure that the twin objectives of maximizing the number of confirmed seats and increasing the seat utilization are met.

We refer to the intermediate stations at which specific inventories of train seats are allocated as remote stations. We also define a major station as one that belongs to a set of stations comprising the origin, destination, and all remote stations; all the other stops on a train for which ticketing is permitted are minor stations. During each journey, manifests are prepared both at the origin and at each remote station. PRS limits the number of remote stations per train to a maximum of six. These specific seat inventories in the form of spatial quotas are allocated for the classes available on the train. Within the spatial quotas, user-specific and time-specific quotas may also be allotted. Appendix B describes IR's quotas.

Trains that traverse long distances have many demands between en route stations; therefore, segmenting a train's capacity in a fair and effective manner is necessary.

Seat Splits

Conventionally, IR allocates spatial quotas to remote stations; however, its process has been to allocate the seats based on a one-leg (i.e., origin to destination) or two-leg (i.e., origin to a remote station and remote station to destination) trip. This type of quota allocation

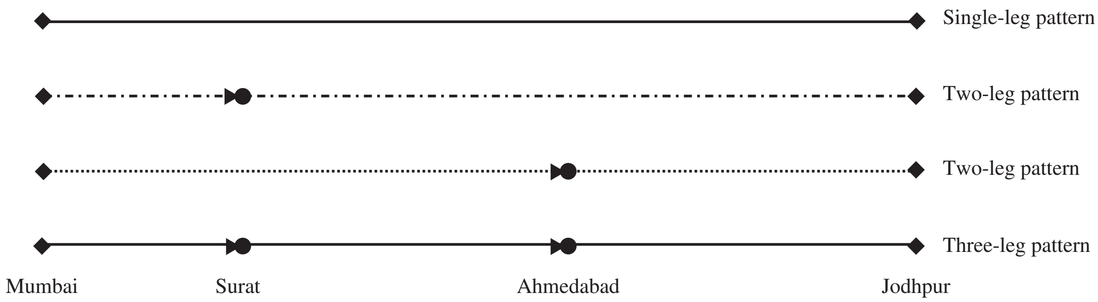

Figure 1: The schematic illustrates seat patterns on a train from Mumbai to Jodhpur.

is reasonable if the entire traffic pattern to (from) an intermediate station is from the origin or to the destination. However, many trains have substantial traffic between intermediate stations.

Figure 1 illustrates a train traveling from Mumbai to Jodhpur and passing through the important cities of Surat and Ahmedabad. Substantial traffic travels not only between Mumbai and Jodhpur, Surat and Jodhpur, and Ahmedabad and Jodhpur (routes that are origin-to-destination, origin-to-remote station, or remote station-to-destination pairs), but also from Surat to Ahmedabad, two remote stations. In IR's traditional two-leg seat-split pattern, an Ahmedabad–Jodhpur ticket is booked against the Ahmedabad–Jodhpur quota and a Surat–Ahmedabad ticket is booked against the Surat–Jodhpur quota. Because confirmed seat numbers are issued at the time of booking, two different seat numbers might be assigned for the Surat–Ahmedabad and Ahmedabad–Jodhpur legs; note that one seat would have been sufficient to clear both demands, thus leaving one full seat to clear a demand for a Surat–Jodhpur seat. Correcting this by creating a new three-leg pattern (i.e., Mumbai–Surat–Ahmedabad–Jodhpur) would cause the Surat–Ahmedabad seats to be booked against the Surat–Ahmedabad leg of the trip, which can be used again at Ahmedabad to offer a confirmed seat to any subsequent booking to Jodhpur. In addition, IR fares are structured such that the fare per kilometer is inversely related to the distance traveled; therefore, the revenue from two tickets (i.e., Surat–Ahmedabad and Ahmedabad–Jodhpur) is greater than the revenue generated from a direct ticket covering the same distance (i.e., Surat–Jodhpur).

To generalize, consider a train that originates at station S and terminates at the station L with four stops and two designated remote stations (Figure 2). In the case of two legs splitting, the seats are defined in three ways: S–L, S–B–L, and S–D–L. Let there be demand for all possible pairs of stations on the train. Table 1 illustrates the quotas that IR uses to book the station-to-station tickets.

We can see that to book one passenger for an S–B, B–D, or D–L leg requires two seats—one for the S–B

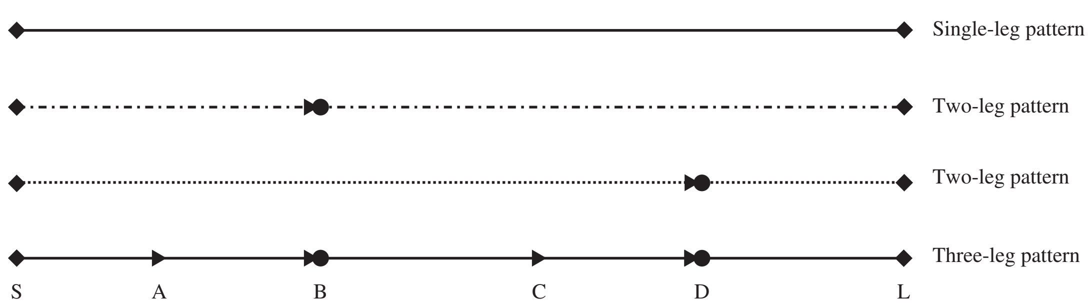

Figure 2: The schematic illustrates a train with A, B, C, and D as stops and B and D as remote stations.

Demand	Pattern used
S–A, S–B, A–B, B–C, B–D, B–L, C–D, C–L	S–B–L
S–C, S–D, A–C, A–D, D–L	S–D–L
S–L, A–L	S–L

Table 1: The table illustrates the use of two-leg seat patterns for different demands; we assume a one-leg pattern to be the default pattern.

and B–D legs using the S–B–L pattern and one for the D–L leg using the S–D–L seat pattern; however, one seat would have been sufficient to book S–B, B–D, and D–L tickets. Similarly, whenever intermediate-to-intermediate station tickets are booked, inefficiencies arise. In addition, because two separate seats are used instead of one, the possibility of requiring a waiting list earlier than necessary increases. To maintain the first-come, first-served (FCFS) booking priority, once an end-to-end or remote-to-destination waiting list starts, any bookings against partially vacant seats are not assigned confirmed seats. Giving a confirmed seat against a partially vacant leg would result in a violation of the booking priority if a passenger who wants to travel the complete leg has a higher waiting-list priority and a cancellation of the partially booked seat occurs. Although many of these passengers might be given seats at the time of manifest preparation, the number of wait-listed passengers would be higher.

This problem can be solved by splitting the seats using multiple-leg seat patterns. Table 2 illustrates the quotas that are used to meet the various station-to-station demands.

Implementation

To implement the model described in Appendix A, processing the relevant data from PRS and moderating the outcome of the LP to provide an implementable outcome was necessary, as we describe in the following sections.

Demand	Pattern used
S–A, S–B, A–B	S–B–L, S–B–D–L
S–C, S–D, A–C, A–D	S–D–L
S–L, A–L	S–L
B–L	S–B–L
B–C, B–D, C–D	S–B–D–L
D–L	S–B–D–L, S–D–L

Table 2: The table illustrates the use of multiple-leg seat patterns for different demands.

Preprocessing of Data

The model uses two types of input data: the selection of remote stations and the demand d_{jk}. The remote stations are train specific and are chosen based on parameters (e.g., the number of persons boarding and detraining at the station) that are available in the PRS data warehouse. Other considerations in selecting remote stations are the importance of the city, the time of arrival at the station, and the availability of a PRS manifest-preparation facility. The model restricts to six the maximum number of stops of a train that can be selected as a remote station in the IR reservation system.

Therefore, we can define at most $2^6 = 64$ different seat-split patterns on a train. Each seat pattern can have a definition that involves one to seven legs of travel. Thus, this model can include up to 64 decision variables.

The demands d_{jk} are estimated using the station-to-station booking data available in the PRS data warehouse. This includes both passengers who have traveled and wait-listed passengers who have not canceled their tickets within five days of their departure date. Many passengers with a wait-listed ticket board the train hoping that they can be accommodated against the train's no-shows. Thus, the station-to-station data also capture some amount of the latent demand in the form of waiting lists.

As we see above, each station-to-station ticket must be booked against one of the quotas q_{jk}. A train may have many stops; however, at most six stops can be defined as remote stations in PRS for the purpose of allocating quotas.

We compress the station-to-station booking data into a major station-to-major station demand d_{jk}. To do this, we note that a station-to-station ticket that is booked on a train can be one of four types: major-to-major, major-to-minor, minor-to-major, or minor-to-minor ticket. To arrive at the major-to-major compressed demand matrix, demand from and to the minor stations must be added to the major-to-major station demand. We do this using the following steps.

Step 1. Retain the demand from major-to-major stations.

Step 2. Aggregate the demand from a major station i to a minor station j with the demand from i to the first major station after j.

Step 3. Aggregate the demand from a minor station i to a major station j with the demand from the first major station before i to station j.

Step 4. Aggregate the demand between minor stations i and j with the demand from the first major station before i to the first major station after j.

To illustrate these steps, consider a train with three stops between the origin and destination; the stops are numbered as 0 (origin), 1, 2, 3, and 4 (destination). The following matrix represents the station-to-station demand.

$$\text{Demand} = \begin{pmatrix} 0 & d_{01} & d_{02} & d_{03} & d_{04} \\ 0 & 0 & d_{12} & d_{13} & d_{14} \\ 0 & 0 & 0 & d_{23} & d_{24} \\ 0 & 0 & 0 & 0 & d_{34} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

If station 2 is identified as remote, the major stations are $\{0, 2, 4\}$ and the minor stations are $\{1, 3\}$. The compressed matrix is the 3×3 matrix represented below.

Compressed demand D

$$= \begin{pmatrix} 0 & d_{01} + d_{02} + d_{12} & d_{03} + d_{04} + d_{13} + d_{14} \\ 0 & 0 & d_{23} + d_{24} + d_{34} \\ 0 & 0 & 0 \end{pmatrix}.$$

The compressed demand matrix D is used to solve the LP and arrive at the optimal quotas between the major stations. We then use these quotas to define the train in PRS.

Postprocessing of Data

The result of the LP above is q_{jk} values, which we translate to actual seat allocations on trains. This has the following two characteristics before a workable result can be obtained.

1. The model has not explicitly captured the variability in demand. We made heuristic adjustments to the model to account for the variability that we see in practice. Informal estimates from IR regarding variability (because of seasonality and other causes) are as high as 30–40 percent. The full extent of variability is difficult to estimate because capacity shortages lead to unfulfilled demand on some segments. To account for variability, the slack capacity, if available on a train, is allocated to seat splits based on the judgment of IR management. Practically speaking, end-to-end seat-split allocation is most convenient because the reservation system would actually permit the last-minute allocation of such a seat for split demand, but not vice versa. Seat allocation and manifest preparation would pose a constraint in the latter option. As we have mentioned, the volumes are now so high that seat allocation at the time of actual boarding is not a viable option. In the model, the end-to-end seat split appears in exactly one pattern (the single split). Because the LP that we have proposed has a minimization objective, an end-to-end seat split will always be allotted a quantity equal to the demand. Other seat splits might actually receive a slightly larger allocation because they could appear in multiple patterns. Finally, intermediate-station demands could also be met by other trains running on that part of the network. With the above considerations in mind, the major adjustment we made to the outcome of the LP was to heuristically allot the remaining train capacity to various patterns to allow for demand in excess of the stated value and to allot a large part of the remaining capacity to end-to-end seat splits.

2. In IR, an unavoidable element of PRS is that seats are allocated at the time of booking (except for first class and first class, air conditioned, which account for a very small percentage of the total bookings). This means that the quota allocations determined above must be translated into seatwise definitions of seat splits, so that the reservation system can be used in seat-allocation mode. As of now, each seat must be labeled individually because of the design of the relevant seat-inventory database. This IR information system design issue, which must be resolved, hampers the efficient, large-scale implementation of this logic, especially if the solution involves multiple seat-split patterns.

Figure 3 shows a flowchart of the procedure, including data preprocessing and postprocessing.

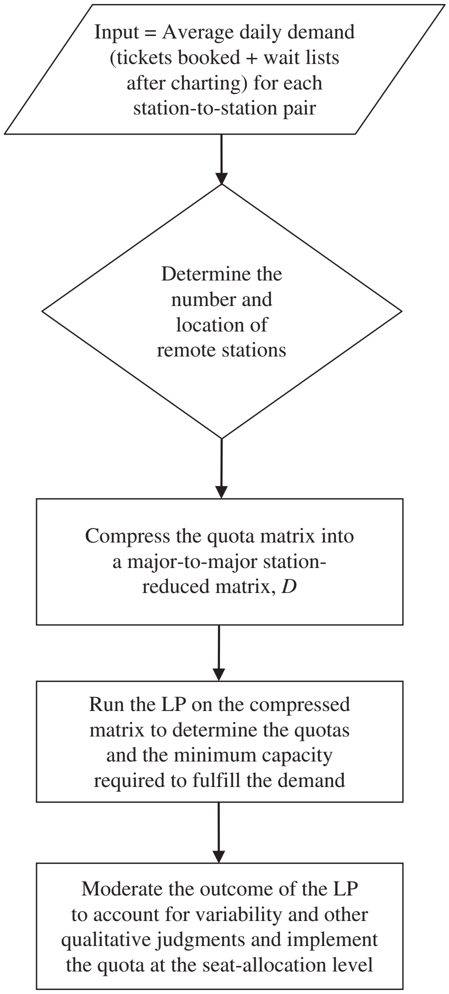

Figure 3: This flowchart of the LP illustrates the case of unconstrained capacity.

Results and Discussion

Detailed Example

Tables 3 to 7 summarize the results of using our application to profile an intercity day train with station A as its origin, station D as its destination, and nine stops en route its 500-kilometer journey. In its earlier train profile (i.e., seat-quota definition) in PRS, only one intermediate station, station B, was defined as a remote station. However, in analyzing the station-to-station ticket sales for this train, we found that many passengers entrained and boarded at station C. Hence, we added station C as a remote station. In discussions with the frontline sales personnel, we found that many passengers wanted to travel from station C to station D at the last minute; most were given wait-listed tickets because of the PRS seat-allocation logic—even when many partially vacant seats from C to D were available. The model suggested a quota of 273 seats between A and C, which was equal to the average demand from A to C. However, the demand would be lower than the average on some days;

From/To	A	B	D	Total
A	0	36	835	871
B	0	0	36	36
D	0	0	0	0
Total	0	36	871	907

Table 3: This table shows the seat quota for each pair of major stations before optimization (February–March 2007).

From/To	A	B	C	D	Total
A	0	70	273	487	830
B	0	0	2	27	29
C	0	0	0	59	59
D	0	0	0	0	0
Total	0	70	275	573	918

Table 4: This table shows the average number of passengers who traveled per trip for each pair of major stations before optimization (February–March 2007).

From/To	A	B	C	D	Total
A	0	70	273	488	831
B	0	0	43	27	70
C	0	0	0	316	316
D	0	0	0	0	0
Total	0	70	316	831	1,217

Table 5: This table shows the seat quota suggested by the model for each pair of major stations using February–March 2007 data.

From/To	A	B	C	D	Total
A	0	56	192	623	871
B	0	0	25	31	56
C	0	0	0	217	217
D	0	0	0	0	0
Total	0	56	217	871	1,144

Table 6: This table shows the actual seat quota defined in PRS for each pair of major stations using the model and postprocessing data (February–March 2008).

From/To	A	B	C	D	Total
A	0	69	239	582	890
B	0	0	4	35	39
C	0	0	0	132	132
D	0	0	0	0	0
Total	0	69	243	749	1, 061

Table 7: This table shows the average number of passengers who traveled per trip for each pair of major stations after optimization (February–March 2008).

on those days, some end-to-end passengers traveling from A to D might be denied seats although seats were vacant as the train traveled between A and C and C and D. Hence, we used a more conservative quota; we reduced it to approximately 70 percent of the suggested quota—an allocation of 192 seats between A and C. Similarly, we reduced the quota of seats on the A–B route to 56 seats from the suggested 70 (a reduction of 80 percent), although this represented a substantial increase from the 36 allocated earlier. In addition, we found that although the average seat demand between A and B was 70, the average seat demand between B and D was only 36; i.e., approximately 40 seats were vacant between B and D. Therefore, we split the new quota of 56 seats between A and B into two patterns; we allocated 25 seats to B–C–D and 31 seats to B–D; the same B–C–D seat pattern was available for booking at C, thus increasing the quota from C to D. We expected that the B–C leg would be patronized poorly, as it had previously been. As a result of redistributing the quotas, the average number of seats booked between C and D rose from 59 to 132, an increase of 124 percent.

Application on Western Railway

On Western Railway, a large operating zone of IR headquartered in Mumbai, we changed the profile, i.e., the various quotas on the train, of 50 long-distance trains using this multiple-leg LP model; the results we achieved were encouraging. In Tables 8 and 9, we summarize the increases in earnings, load factors (i.e., number of passengers divided by capacity, as a percentage), capacity, and passengers for 17 trains that we used our model to profile.

The following conditions apply to specific trains listed in Tables 8 and 9.

- Train xx33: IR added competing trains in similar time slots and with similar train composition; nevertheless, passengers and revenue increased.
- Train xx65: During the review period, this train's route changed to bypass a major en route station.

		Revenue (in Indian Rupees)			Load factor		
Train no.	Period of review	Before	After	Change (%)	Before	After	Change (%)
xx03	Oct07–Oct08	282,657,760	327,173,320	15.75	184.21	190.87	6.66
xx37	Nov07–Nov08	67,204,799	76,523,042	13.87	115.86	139.35	23.49
xx35	Nov07–Nov08	36,715,220	41,020,730	11.73	121.26	148.93	27.67
xx33	Dec07–Dec08	36,186,597	39,487,865	9.12	123.13	151.82	28.69
xx09	Jan08–Jan09	168,195,463	201,370,269	19.72	109.94	127.14	17.20
xx10	Feb08–Feb09	215,063,818	241,937,639	12.50	124.93	135.09	10.16
xx65	Feb08–Feb10	17,853,341	18,318,168	2.60	136.07	148.04	11.97
xx69	Mar08–Mar09	29,319,224	36,838,306	25.65	171	188.22	17.22
xx13	Mar08–Mar09	35,495,831	41,478,224	16.85	136.36	145.77	9.41
xx17	Mar08–Mar09	86,360,929	104,481,112	20.98	128.37	142.53	14.16
xx01	Mar08–Mar09	32,921,257	41,465,771	25.95	115.05	145.61	30.56
xx05	Mar08–Mar09	31,846,876	41,191,803	29.34	113.83	138.82	24.99
xx33	Mar08–Mar09	93,050,197	101,384,831	8.96	108.4	123.03	14.63
xx34	Mar08–Mar09	92,895,154	104,475,423	12.47	113.2	128.47	15.27
xx11	Apr08–Apr09	43,028,686	48,744,600	13.28	98.31	116.91	18.60
xx12	Apr08–Apr09	54,836,571	60,470,937	10.27	102.65	120.35	17.70
xx80	Dec07–Dec08	152,352,457	186,150,195	22.18	128.29	146.94	18.15
IR average	Apr08–Apr09	$108{,}309 \times 10^6$	$127{,}084 \times 10^6$	17.33	99.72	103.86	4.14

Table 8: This table summarizes the revenue and load-factor changes for the trains we profiled using our OR model. (Note that the load factor can exceed 100 percent because multiple passengers can occupy the same seat during a trip between the origin and the destination.)

Interfaces 40(4), pp. 291–302, © 2010 INFORMS

		Capacity			Passengers		
Train no.	Period of review	Before	After	Change (%)	Before	After	Change (%)
xx03	Oct07–Oct08	316,090	330,690	4.62	582,259	631,179	8.40
xx37	Nov07–Nov08	107,952	107,952	0.00	124,880	150,434	20.46
xx35	Nov07–Nov08	53,976	53,976	0.00	65,452	80,386	22.82
xx33	Dec07–Dec08	53,976	53,976	0.00	66,452	81,944	23.31
xx09	Jan08–Jan09	287,021	293,590	2.29	315,562	373,284	18.29
xx10	Feb08–Feb09	287,021	293,590	2.29	358,585	396,597	10.60
xx65	Feb08–Feb10	32,344	32,344	0.00	44,012	47,883	8.80
xx69	Mar08–Mar09	34,840	39,312	12.84	59,578	73,994	24.20
xx13	Mar08–Mar09	51,532	55,014	6.76	70,268	80,196	14.13
xx17	Mar08–Mar09	156,624	159,588	1.89	201,055	227,455	13.13
xx01	Mar08–Mar09	52,624	53,636	1.92	60,542	78,099	29.00
xx05	Mar08–Mar09	52,624	53,636	1.92	59,904	74,460	24.30
xx33	Mar08–Mar09	419,818	421,611	0.43	455,068	518,707	13.98
xx34	Mar08–Mar09	419,818	421,611	0.43	475,239	541,627	13.97
xx11	Apr08–Apr09	281,088	279,955	−0.40	276,336	327,303	18.44
xx12	Apr08–Apr09	293,166	292,000	−0.40	300,929	351,428	16.78
xx80	Dec07–Dec08	363,905	383,250	5.32	468,663	563,143	20.16
IR average	Apr08–Apr09	274×10^6	297×10^6	8.35	273×10^6	308×10^6	12.84

Table 9: This table summarizes the changes in capacity and passengers carried for the trains we profiled using our OR model.

- Train xx69: IR made an independent decision to reduce this train's capacity; nevertheless, overall passengers and revenue increased.
- Trains xx13 and xx80: During the review period, IR added capacity; however, the passengers and revenue increased by values that cannot be attributed only to the increased capacity.

The following results apply to the data in Tables 8 and 9.

- In most cases, the trains that we profiled using the model exceeded IR's average numbers for passengers, load factors, and revenue. Note that a load factor greater than 100 indicates usage of the seat for multiple segments of travel.
- Because our algorithm generates minimum seat requirements, we found the seat requirement to be less than the capacity on some trains; however, the model suggested that we should increase capacity on other trains. Based on its recommendations, we detached two coaches from each of the four rakes (i.e., coach groups) of a daily long-distance train, thus saving eight coaches and more than US$1 million.
- Using the multiple-leg seat splits, many passengers were given confirmed seats at booking. The revenues per train trip increased with no capacity additions.
- The number of passengers carried per train trip increased from 8 percent to 29 percent because of better seat-split decisions.

Discussion

Passenger demands used in the model are the periodical averages over a long period. Because PRS includes waiting-list data, the model captures some latent demand. However, the demand data include a slight bias because past-bookings patterns were influenced by the previous quotas specified for each train. IR also limits the number of passengers who can be issued waiting-list tickets. Therefore, passengers who want to book tickets when the waiting-list limit has been reached are denied bookings although capacity might be available at the time of manifest preparation against partially vacant seats on the requested legs.

Conducting a fully controlled experiment in capacity management in a railway environment is difficult because railways typically test multiple simultaneous initiatives. For example, a railroad might enhance capacities on several trains and simultaneously introduce new classes that compete with the existing ones. In addition, prices on passenger segments have not changed for the last five years; however, competing transportation systems, such as airlines and buses, have adjusted fares multiple times. Nevertheless, our results demonstrate a substantial improvement in both numbers of people carried and IR revenue.

Extensions

In this section, we briefly discuss a few extensions to our basic model. A train cannot always accommodate all station-to-station passengers, and limitations on the number and availability of coaches that can be attached on a train impose capacity constraints. In such cases, the objective function must be modified. One option is to make maximizing revenue the objective. However, this would impose a limitation; it would be biased in favor of short-distance passengers because IR charges a higher fare per kilometer for traveling short distances than for traveling longer distances. Another way to address the issue is to restrict specific station-to-station passengers because other equally convenient trains with spare capacity are available. Networkwide constraints (e.g., considering all trains running between the different stations) might have to be evaluated. In addition, prioritized demands could be handled by using a weighted-objective function.

The selection of the remote stations at which the seats have been split is intuitive and based on the station-to-station demand matrix. The choice of remote stations could be optimized. In compressing the demand matrix, we have assumed that each seat demand between two major stations would require a full seat; however, we could also have used partially vacant seats between major stations. As seat-split definitions increase exponentially, using this model to solve problems with more remote stations would not be feasible.

Our model does not consider fares between stations; therefore, maximizing revenue could also be a corporate objective. Because IR is a monopoly service provider, its pricing is partly regulated; therefore, we chose the service constraint of maximizing the number of people carried given a specific capacity as our objective in this exercise. However, IR does practice revenue and peak-load management by pricing; through its *tatkal* product, it charges higher fares for late-decision travel and for peak periods. Incorporating revenue objectives into the optimization model would be a valuable next step for this application.

An important issue that we also do not address is splitting a train's overall capacity across various classes of travel; this topic could be analyzed separately as a part of the overall revenue management function.

Conclusion

The model has helped IR to reduce its seat requirements and has increased the availability of confirmed seats for the various station-to-station demands on several trains. Despite the limitations that we discussed above, the model provides a good foundation for improving the capacity utilization of long-distance passenger trains. It has provided a systematic basis for one aspect of capacity management as part of overall revenue management on long-distance passenger rail services.

Appendix A. Optimization Formulation: Unconstrained Capacity

We set up the capacity-allocation problem using the set of major stations, i.e., origin, remote stations, and destination. Let n be the number of remote stations on a train route. The number of possible seat-split patterns is 2^n because each of the n remote stations is either present or not present in each possible seat split. We denote x_i as the number of seats allocated to the ith seat-split pattern and denote the associated quantities q_{jk} (i.e., the quota between any pair major stations j and k) as the decision variables, and we denote d_{jk} as the demand between any two major stations j and k.

All seats on the train belong to one of the possible 2^n definitions of the seat-split patterns. The number of seats on the train is $\sum_{i=1}^{2^n} x_i$. Minimizing the total seats required to fill all possible demands d_{jk} is set as the objective function of the LP formulation. We can define $\frac{1}{2}(n+1)(n+2)$ station-to-station quotas (i.e., $n+1$ quotas from the origin, n from the first remote station, and so on, continuing through all n remote stations). The quota q_{jk} is the sum of those x_i in which the pattern $j-k$ includes the ith seat split. To fulfill all demand for seats between each pair of stations, j and

k, the quota of seats q_{jk} should exceed the demand d_{jk}. Thus, we can summarize the LP as

$$\text{minimize} \quad C = \sum_{i=1}^{2^n} x_i$$

subject to

1. $q_{jk} = \sum_{i=1}^{2^n} x_i \cdot \delta_i$ for all patterns $j-k$, where

$$\delta_i = \begin{cases} 1, & \text{if pattern } j-k \text{ is present in the } i\text{th seat spilt pattern;} \\ 0, & \text{otherwise;} \end{cases}$$

2. $0 \le q_{jk} \le d_{jk}$ for all patterns $j-k$;
3. $x_i \ge 0$ for all seat splits i;

where d_{jk} was calculated from the PRS data warehouse described in the *Implementation* section.

Appendix B. IR's Passenger Reservation System: An Overview

IR runs trains with seats available in reserved and unreserved classes. Most of the coaches on its trains are exclusively of one class type. On most trains, seats can be booked 90 days in advance—the advance reservation period (ARP).

IR trains stop at many en route stations at which passengers who have booked a seat in a particular train class can board or disembark. The seat price depends on the distance traveled. A longer distance is more expensive; however, the price per kilometer decreases as the distance increases. Thus, passengers traveling longer distances pay less per kilometer traveled than passengers traveling shorter distances. IR gives fare discounts to specific riders, e.g., senior citizens and students. In each travel class, IR allocates tatkal seats for last-minute passengers who are willing to pay a premium. PRS manages all reserved-segment ticket bookings.

PRS: Types of Trains

IR follows a business policy of dividing the total capacity of each class in a train into different quotas that are based on spatial, temporal, or user-driven factors. Quotas impose restrictions on seat availability based on station, type of passenger, and time of booking. We define a train as either end-to-end or non-end-to-end. We can define a set of stations near the originating station as the origin cluster. Similarly, we can define a destination cluster as a set of stations near the destination station. An end-to-end train is one in which the predominant traffic is from the origin cluster to destination cluster. By using these definitions, it is possible to restrict ticket sales for those pairs of stations that are not classified as an origin/destination cluster pair. A non-end-to-end train is any train that is not defined as an end-to-end train.

IR ticketing involves an elaborate mechanism of spatial, temporal, and user-specific quotas, as we discuss in this paper. A given class in a train might have some or all of the quotas we describe below.

1. Spatial quotas

(a) General quota (GN): GN represents the number of seats allocated for the originating station against which passengers boarding from the originating station are booked. This quota is accessible until manifest preparation, a process undertaken approximately three to four hours prior to a train's departure from the originating station; at this time, the train's bookings are frozen and the final manifest is prepared. After manifest preparation, all the temporal and user-specific quotas at the station are merged into a common pool. In an end-to-end train, only seats for which the entraining and detraining station belong to origin and destination clusters are booked against this quota.

(b) Remote station quota (RS): Under an RS quota, some train capacity is allocated to important en route stations, other than the origin/destination clusters, that have substantial traffic, i.e., a large number of passengers boarding or disembarking. These are also the stations other than the origin at which separate manifests are prepared. All stations starting from a remote station to a station before the next remote station or the destination (whichever comes first) can access the RS quota. The seats in the RS quota are available for booking until the time of manifest preparation (at the remote station), which is three to four hours before the train's scheduled departure from the remote station. Thus, tickets subject to an RS quota might be available although the train has departed from the originating station.

(c) Pooled quota (PQ): This is a common inventory of seats available for all stations of the train on an FCFS basis. Bookings are allotted under this quota once the GN or RS quota for the station, as the case may be, is exhausted. Remember, stations that cannot access GN or RS quota have access to only this inventory of seats. This common inventory acts as buffer for any demand in excess of the quota allocated in GN or RS.

(d) Outstation quota (OS): Some stops on IR do not have access to PRS; i.e., they require booking tickets manually outside the system. A small inventory of seats is allocated to such stations under the OS quota.

2. User-specific quotas

(a) Lady quota (LD): A small number of seats are allocated exclusively for female passengers.

(b) Handicapped-person quota (HP): A few seats on each train are allocated for physically challenged passengers.

(c) Senior citizen quota (SS): Two seats per coach on each train are allocated for senior citizens.

(d) Defense quota (DF): Most trains have seats that are allocated for military personnel.

(e) Foreign-tourist quota (FT): Many trains, in particular those connecting important metropolitan areas and tourist destinations, have capacity allocated for foreign tourists. These are sold through the IR's international sales agents and are available beyond the time dictated by ARP scheduling.

(f) High-official quota (HO): In all train classes, some seats are allocated for the last-minute travel requirements of high-ranking government officials and their guests. Passengers traveling because of medical or other emergencies are also allotted seats under this quota.

3. Temporal quotas

(a) Tatkal quota (CK): IR's tatkal product allocates seats, which are sold at a premium over the regular fare, for last-minute travelers. A substantial capacity, ranging from 10 to 30 percent of the train, is allocated under CK. The premium varies for lean and peak periods. Tatkal tickets can be purchased two days before the scheduled date of departure (excluding the date of departure) and until the time of the train's manifest preparation.

All seats that remain unutilized under the various quotas are used to clear wait-listed passengers under other quotas at the time of the train's manifest preparation.

General Passenger Booking

A passenger who does not belong to any group that is entitled to a user-specific quota is subject to the train's spatial quotas.

1. In a non-end-to-end train, the booking is sequential—bookings for tickets demanded between stations A and B are mapped to the quota allocated between the remote station prior to station A and the remote station following to station B. If this seat quota has been exhausted, the tickets are booked against the quota between the remote station prior to station A and the next remote station, and so on. This seat allocation continues until the quota between the remote station prior to station A and the destination station of the train is exhausted; at that time, waiting lists are generated.

2. In an end-to-end train, an origin cluster and a destination cluster of stations are defined. Only passengers traveling between the origin and destination clusters are booked against the seats using a GN. Passengers boarding at stations after the origin cluster and before the first remote station are booked in the seats allocated to PQ. For other boarding stations, i.e., the first remote station and beyond, the seat-allocation logic is the same as for a non-end-to-end train. Once the general or remote quota has been exhausted, the other stations also use the PQ.

Pooled Quota: Features

Understanding the PQ benefits and disadvantages is important because this quota plays a pivotal role in addressing the quota-allocation problem. The major advantages are as follows:

1. PQ does not have any spatial or user restrictions and is accessible to all possible station pairs on an FCFS basis.

2. It acts as a common buffer for the spillover after the other spatial quotas such as GN or RS are exceeded. In other words, PQ fulfills the variations in demand that exceed the allocated quota.

3. It helps restrict bookings in an end-to-end train for those pairs of stations that are not part of an origin/destination cluster.

PQ also has a disadvantage. Unlike an end-to-end definition, pooled quotas do not deny short-distance bookings; therefore, the entire PQ could be used by travelers who request bookings for short lead-time travel on end-to-end trains. Passengers seeking bookings against the vacant legs of a seat are issued wait-listed tickets that are confirmed only at the time of manifest preparation. Therefore, a trade-off exists between the advantage of PQ's unrestricted accessibility and the possibility of passengers who are traveling for short distances. This trade-off can determine the number of seats to be allocated to PQ.

Acknowledgments

We acknowledge A. K. Goyal, R. K. Tandon, O. P. Chawla, and V. K. Sharma of IR for their help in conceptualizing this application; V. K. Nikale, J. C. Birdi, Anthony Lobo, and Hemangi Vatkar of IR for their help at various stages of implementation; D. Bijulal and Meghraj Nalge for their help in preparing this paper; and the reviewers for their comments.

References

Ciancimino, A., G. Inzerillo, S. Lucidi, L. Palagi. 1999. A mathematical programming approach for the solution of the railway yield management problem. *Transportation Sci.* **33**(2) 168–181.

Dror, M., P. Trudeau, S. P. Ladany. 1988. Network models for seat allocation on flights. *Transportation Res.* **22**(4) 239–250.

Glover, F., R. Glover, J. Lorenzo, C. McMillan. 1982. The passenger-mix problem in the scheduled airlines. *Interfaces* **12**(3) 73–79.

Hersh, M., S. P. Ladany. 1978. Optimal seat allocation for flights with one intermediate stop. *Comput. Oper. Res.* **5**(1) 31–37.

Indian Railways. 2007. Indian Railways Year Book: 2006–07. Retrieved June 14, 2009, http://www.indianrailways.gov.in/deptts/yearbook/yearbook-0607/passenger-business.pdf.

Rozite, K., N. A. Nechval, E. K. Vasermanis. 2005. Optimal airline seat inventory control for multi-leg flights. *Transport Telecommun.* **6**(1) 114–121.

Talluri, K. T., G. J. van Ryzin. 2004. *The Theory and Practice of Revenue Management.* Kluwer Academic Publishers, Boston.

N. C. Sinha, Chief Commercial Manager, Western Railway, Churchgate, Mumbai–400 020, writes: "This is to certify that Mr. Raja Gopalakrishnan has done innovative work on Capacity Management of Long-Distance Trains of Western Railway. He has done this work during his stint as Deputy Chief Commercial Manager, Western Railway, Mumbai, India.

"The model and software developed have been used in over 50 long-distance trains originating on Western Railway with considerable success.

"This software has also been shared with other zones of Indian Railways for implementation. A workshop has also been conducted to train the personnel of the different zones in implementation of the model."