ARTICLE IN PRESS

Transportation Research Part E xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

Estimating flow times for containerized imports from Asia to the United States through the Western rail network

Robert C. Leachman [a,*], Payman Jula [b]

[a] Department of Industrial Engineering and Operations Research, University of California at Berkeley, Berkeley, CA 94720-1777, USA
[b] Technology and Operations Management, Beedie School of Business, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6

ARTICLE INFO

Article history:
Received 21 November 2010
Received in revised form 6 June 2011
Accepted 14 June 2011
Available online xxxx

Keywords:
Congestion Analysis
Railroad
Rail terminal
Intemodal freight transport
Containerized import
Queuing

ABSTRACT

Queuing models are introduced for estimating container dwell times at rail intermodal terminals and transit times through rail line-haul corridors. These models are statistically calibrated on industry data. The intent of these models is to estimate changes in container flow times stemming from changes in infrastructure, staffing levels at terminals, or import volumes passing through given infrastructure. Flow times estimated for individual line segments are aggregated to provide estimates of the total transit time from West Coast rail ramps to inland destination ramps for imports moving from Asia to the Continental United States in marine as well as domestic containers.

1. Introduction

The substantial growth in waterborne, containerized imports from Asia to the Continental United States experienced until 2007 (i.e., before the onset of the subsequent economic recession) strained the capacities of West Coast ports and landside channels to inland markets. At times, "melt-downs" were experienced at certain West Coast ports and on certain Western rail lines that triggered major shifts in the port and channel volumes of such imports. In response to this trade growth, there have been major expenditures by public agencies and private carriers to expand infrastructure, continuing at the present time. In some cases, new user fees or container fees have been introduced or proposed to pay for such improvements. Rates charged by railroads have in some cases escalated dramatically.

Some of the melt-down events came as a surprise to industry managers and governmental officials. We believe this reflects a lack of practical analytical tools that can be used to predict container dwell and transit times as a function of volume, infrastructure and staffing. While there is much useful literature on simulation models and queuing formulas for operational analysis of individual transportation links and terminals, to our knowledge there is little research on practical congestion analysis of large rail networks to support strategic planning. We aim to fill that need in this article. We provide a queuing model for rail intermodal terminals, the first such model to the best of our knowledge. We develop a new queuing formula for individual rail line-haul segments and combine in a practical way the results of application of the formula to individual rail links into estimates of the total transit times in transcontinental rail container corridors. Combined with results of queuing analyses of terminals, estimates of total container flow times from arrival at origin ramp to train arrival at inland destination ramp are readily developed for multiple transcontinental corridors. To the best of our knowledge, this is the first time a practical queuing-theoretic approach has been developed for estimating transit times through large, general freight rail networks.

* Corresponding author. Tel.: +1 510 642 7054; fax: +1 510 642 1403.
E-mail address: leachman@ieor.berkeley.edu (R.C. Leachman).

1366-5545/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tre.2011.07.002

Our motivation is the planning for continued growth of containerized imports from Asia to the Continental United States passing through West Coast ports and distributed across the Continental United States. The queuing models we propose are statistically calibrated on industry data for these import flows. There are many terminals and landside channels for which container flow times must be estimated. The desired accuracy of total-channel container flow times is on the order of days.

The structure of this article is as follows. We first provide an overview of the sequence of queues that imports moving in rail channels must negotiate. Next, we review the relevant literature. We then proceed to the development of our proposed queuing models and illustrate their application.

2. Supply-chain strategies and relevant queues

The majority of containerized imports from Asia to the Continental United States are retail goods or goods nearly ready for retail sale. The uninformed observer might think that all such containerized imports from Asia would move intact in marine containers from port of entry to distribution centers serving the regions of ultimate consumption, but such is not the case. In 2006, about 30% of such imports were unloaded from marine containers within the hinterland of the port of entry and re-shipped in larger domestic containers immediately or after import warehousing for some time, and this share has been steadily rising for more than a decade (Leachman, 2010).

This is significant for our purposes because domestic containers and marine (international) containers are generally handled by US railroads at separate intermodal terminals and in separate trains. Trains handling domestic containers are generally assigned higher horsepower and more priority by the railroads. Thus flow times for imports in the direct-shipment-of-marine-box import channels and in the trans-load-to-domestic-box channels, and the sensitivities of these flow times to growth and to changes in infrastructure or service levels, are different. Moreover, a shift in market shares between direct-shipment and trans-loaded supply chains results in different strains on the elements of the rail network.

Imports moving across the Rockies intact in marine containers move under what is termed inland point intermodal (IPI) service, whereby a trans-Pacific steamship line sells transportation door-to-door from Asian factory to inland USA distribution center and subcontracts with a US railroad to operate trains of double-stacked marine containers from rail intermodal terminals at West Coast port cities to distant inland rail intermodal terminals. The marine containers may be loaded into the railroad's double-stack well cars at on-dock rail terminals within the port complex, or the boxes may be drayed over streets and highways from port terminals to "off-dock" (i.e., remote) rail terminals for loading into railroad well cars.

Imports trans-loaded to domestic containers move under a new bill of lading after trans-loading, and the steamship line is not involved in the onward transportation. Such imports are drayed out of the port terminal over city streets and highways and are tendered to the railroad at an off-dock rail intermodal terminal as domestic freight after trans-loading to domestic containers.[1] Domestic containers are double-stacked in a different-sized railroad well car designed to accommodate the larger domestic containers.

The series of rail-related queues experienced by waterborne containerized imports under the alternative strategies are summarized as follows:

Port terminal queues:

- Loading containers into railroad well cars coupled into strings of well cars on parallel tracks at on-dock terminals (on-dock rail IPI movement only).
- Shuttling strings of loaded well cars from on-dock loading tracks to a nearby staging railyard, and assembling the strings of well cars into longer trains at the staging railyard (on-dock IPI movement only).

Queues experienced outside the port but still in the general vicinity of the port of entry:

- Loading domestic containers into railroad well cars at off-dock rail terminals (trans-loading supply chains) and loading marine containers into railroad well cars at off-dock terminals (for IPI movements not loaded on-dock).

Queues in line-haul movement:

- Delays to line haul movement of double-stack container trains (both marine-box stack trains and domestic-box stack trains) for meets with opposing rail traffic on single-track lines, and for following and overtaking slower rail traffic moving in the same direction.

The magnitude of delays and the sensitivity to total rail import volumes are different for the various types of queues. Considering our purposes, we develop analytical queuing models for the queues that we perceive to be very sensitive to import volume and to have flow-time impacts in aggregate that are measured in days. We approximate the delays stemming from

[1] On-dock terminals are integrated into ports. The container flow times through such terminals from vessel arrival until double-stack trains are loaded or until container drays are dispatched are the subject of a different article, Leachman and Jula (in press). Off-dock terminals are operated by railroads. The container flow time from arrival of container dray until departure of double-stack train from off-dock terminals is an important subject of this article.

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002

other queues we perceive to be relatively insensitive to import volume with fixed time lags. Our classification of queues in this regard is discussed as follows.

2.1. Rail queues inside the port terminals

2.1.1. Loading outbound drays and railroad well cars at port terminals

Containers unloaded from a vessel are placed in stacks in a staging area of the dock. These stacks comprise large cubes of containers. Containers must be extracted from the cube and placed on chassis for outbound drays or placed in railroad well cars (if the terminal has on-dock rail service). At larger terminals, separate cubes are maintained for containers to be loaded in rail well cars and for containers to be drayed off the dock. A separate cube also is maintained for export containers (to be loaded on vessels). In such cases, each cube has a separate workforce and in effect is a separate queuing system. Unfortunately, container dwell-time data disaggregated to provide statistics solely on import containers leaving the terminal via rail is not readily available at most terminals we have visited. However, more aggregate statistics on total container dwell times from vessel arrival until container departure on dray or train are routinely compiled by most port terminal operators. Generally, at USA West Coast port terminals, the IPI volume is a fraction of the outbound dray volume. For this reason, a queuing model jointly applied to containers moving in both outbound drays and IPI service must be developed. That is outside the scope of this paper but we develop such a model in Leachman and Jula (in press).

2.1.2. Assembling trains of well cars loaded at on-dock terminals

On-dock rail terminals operate under more restrictive work-rules than off-dock terminals. The lengths of unloading/loading tracks are generally much shorter than at off-dock terminals. Productivity, measured in terms of lifts completed per day, is much lower than at off-dock rail ramps. This stems from a combination of factors: (1) Port terminals operate fewer hours per week than off-dock terminals, and port terminal crews take more break time during their shifts than crews at off-dock terminals. (2) Railroad switching is required to pull out loaded strings of well cars and replace them with empty strings supporting continued loading operations. But switching movements are forbidden on tracks adjacent to a track holding well cars undergoing loading or unloading operations. (3) If loading operations must pass over the lead track to the terminal, then all switching of the terminal is forbidden during loading operations. During periods of high port terminal volume at the San Pedro Bay ports, the railroads responded to these constraints by limiting the destinations for on-dock rail loading to only high-volume destinations. IPI boxes to lower-volume destinations are required to be drayed to off-dock rail intermodal terminals at such times. The cuts of loaded well cars are assembled into trains for long-distance movement using staging yards located outside the on-dock rail terminals. Typically, a single staging yard serves multiple terminals. While there is some queuing associated with the train assembly activity, the railroad policy to limit the destinations during peak traffic periods renders the time to build trains from loaded well cars quite stable. In effect, much of the queuing associated with high IPI traffic levels is transferred to the outbound dray queue described above. We therefore choose to model the time to build stack trains as a fixed lag.

2.2. Queues outside the port terminals

2.2.1. Loading trains at off-dock rail intermodal terminals

Containers arrive at off-dock rail terminals (rail "ramps") mounted on chassis. The containers are lifted off the chasses and placed in double-stack well cars. Limited terminal acreage or limited crew-shifts of operation juxtaposed with significant traffic volumes can lead to significant container dwell times at the terminal. Data we received from the railroads exhibit significant variation in dwell times at rail ramps. We explicitly model queuing effects at off-dock rail intermodal terminals.

2.2.2. Rail line-haul movement

In certain years during the period 1990–2010, the Union Pacific Railroad experienced very serious melt-downs. Rail lines consisting of a single main line supplemented with passing sidings at various points exhibit instability at high traffic levels. That is, once many of the passing sidings on the line are occupied by trains, trains have a difficult time advancing down the line, and it becomes much more time-consuming for trains to complete runs across crew districts. By Federal law, train and engine crews can only be on duty up to 12 h, whereupon they must be given rest and replaced by a rested crew, regardless of their physical location at the time the hours of service expire. At sufficiently high traffic levels, many crews run out of time before reaching their terminal. This leads to a shortage of crews and a shortage of locomotive units whereby many locomotive units are unmanned on trains parked in sidings because there are no rested crews to man them. Such melt-down conditions are difficult to overcome. Recovery of line fluidity has taken months to more than a year. Both the BNSF and Union Pacific Railroads have made substantial investments in recent years to double-track much of their transcontinental rail lines. These investments make melt-downs much less likely. Nonetheless, large importers we have interviewed experience considerable variance in transit times of double-stack container trains. These variances arise from delays for meets with opposing traffic on remaining single-track line segments, for following and overtaking slower-moving bulk and carload trains, for interference from track maintenance work, and other problems. Trains hauling containers loaded with imports and trains moving in the opposite direction backhauling the intermodal equipment account for a very large share of total rail traffic on North American transcontinental main lines. We therefore explicitly model queuing effects in rail line haul movement.

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002

In subsequent sections, we will develop and calibrate practical queuing models for estimating container flow times through the two most serious rail-related types of queues: Loading containers onto trains at off-dock rail ramps, and train delays in rail line-haul networks.

3. Literature review

Caris et al. (2008) provided an overview of planning decisions and solution methods in the domain of intermodal freight transport systems. They found a lack of research on strategic-level issues (network and facilities investment and configuration, service levels) and tactical-level issues (allocation of resources, pricing, staffing) facing intermodal operators. To the best of our knowledge, there is no published work on queuing models specifically developed for or applied to rail intermodal terminals. While there is a large literature of queuing analyses of rail line-haul networks, our review of this literature finds that nearly all such queuing models focus on sub-systems and are suited to supporting operational decision-making (scheduling, control, and optimization) in such sub-systems.

In the following section we will propose queuing models for rail intermodal terminals, and rail line-haul channel sub-systems as well. However, our models are designed to help industry managers and governmental policymakers make timely and informed decisions at strategic and tactical planning levels. These models are designed to be components of an integrated analysis of total container flow times through landside channels. That is, we target the gap in the literature to support managers making decisions at strategic and tactical levels. We now discuss in more detail the literature treating rail-related queues.

3.1. Rail line haul

There exists a considerable literature of both analytical and simulation-based methods which have been used to study delays and capacity assessment in railroad line haul networks with specific configurations.

Frank (1966) studied delay on a single track rail line with unidirectional and bidirectional traffic. The author estimated the number of trains that could travel on the network by considering only one train on each link between sidings and using single train speeds, and assuming deterministic travel times. This work was later extended by Petersen (1974) to accommodate for two different train speeds, while assuming independent and uniformly distributed departure times, equally spaced sidings and a constant delay for each encounter between two trains. Chen and Harker (1990) extended this model to calculate delays for different types of trains over a specified single track section as a function of the schedules of the trains and the dispatching policies. They assumed that faster trains can overtake slower trains, sidings are equally distributed, meets and overtakes occur only between two trains at a time, and the existence of a constant probability of delay between trains. This modeling technique was extended by Harker and Hong (1990) to a partially double-track rail network which consisted of a single-track section with sidings and double-track sections. Özekici and Şengör (1994) used Markov chain techniques to study the effects of various dispatching patterns and arrival patterns of passengers on delays and passenger waiting times. Given a travel time probability density function for a train on a track link, the authors proposed a departure time transition matrix for the calculation of the expected departure delay. Carey and Kwiecinski (1994) analyze delays to trains moving in the same direction on the same track at different speeds that cannot pass each other. Analyzing a network with multiple unidirectional and bidirectional tracks, crossings and sidings, Higgins and Kozan (1998) presented an analytical model to quantify the delays for individual passenger trains on individual track links and for the schedule as a whole in an urban rail network. Gibson et al. (2002) developed a regression model to define a correlation between capacity utilization and reactionary delay. Burdett and Kozan (2006) developed capacity analysis techniques and methodologies for estimating the absolute (theoretical) traffic carrying ability of facilities over a wide range of defined operational conditions. They address the factors on which the capacity of a network depends on, namely, the proportional mix of trains, direction of travel, length of trains, planned dwell times of trains, the presence of crossing loops and intermediate signals in corridors and networks. Yuan and Hansen (2007) proposed probability models that provide an estimate of delays and the use of track capacity.

Simulation has been extensively used for estimating delay in railroads. Petersen and Taylor (1982) presented a structured model for rail line simulation. They divide the rail line into track segments representing the stretches of track between adjacent switches and develop algebraic relationships to represent the model logic. Dessouky and Leachman (1995) used a simulation modeling methodology to analyze the capacity of tracks and delay to trains in a complex rail network. Their methodology addresses both single and double-track lines. Their model has a distinct advantage of accounting for track speed limits, headways, and actual train lengths, speed-limits acceleration and deceleration rates in order to determine the track configuration that minimizes congestion delay to trains. This work was extended by Lu et al. (2004). Hallowell and Harker (1998) improved upon the work by Harker and Hong (1990) by incorporating dynamic meet/pass priorities in order to approximate an optimal meet/pass planning process. Extensive Monte Carlo simulations are conducted to examine the application of an analytical line model for adjusting real-world schedules to improve on-time performance and reduce delay. Krueger (1999) used simulation to develop a regression model to define the relationship between train delay and traffic volume. The parameters involved are network parameters, traffic parameters and operating parameters. Murali et al. (2010) presented a simulation-based technique to generate delay estimates over track segments as a function of traffic conditions, as well as network topology to facilitate routing and scheduling freight trains.

Queuing theory also has been used for estimating delay in railroads. Greenberg et al. (1988) presented queuing models for predicting dispatching delays on a low speed, single track rail network supplemented with sidings or alternate routes. Train departures are modeled as a Poisson process, and the long travel times between sidings result in alternating busy periods of traffic moving in one direction and then the other. This work assumed sidings to have infinite capacity. Huisman and Boucherie (2001) investigated delays to a fast train caught behind slower ones by capturing both scheduled and unscheduled movements. Wendler (2007) presented an approach for predicting waiting times using a queuing system with a semi-Markovian kernel. The arrival process is determined by the requested train paths. The description of the service process is based on an application of the theory of blocking times and minimum headway times. Although these efforts show promise, to the best of our knowledge, there is no queuing-theoretic approach for estimating transit times through general, large-scale rail freight networks.

3.2. Rail inter-modal terminals

Although there is considerable contemporary attention to intermodal transportation (see e.g., Ishfaq and Sox, 2010), in general, literature on rail intermodal terminals/yards is scarce. There seems to be a lack of publications concerning applications of queuing techniques to intermodal terminals outside of ports. In some parts of the world, intermodal terminals are viewed as an emerging technology in railway systems (see the survey by Bontekoning et al. (2004)). Most research in this area addressed operational challenges, such as dispatching and scheduling of the handling equipment (see e.g., Boysen and Fliedner, 2010). In this domain, researchers have mostly used simulation as a tool for managerial decision making. Kozan (2006) provided a simulation model to investigate delays of trains for different service configurations. The author uses simulation outputs to find an optimum balance of the cost of train delays and variation from the desired level of service. Benna et al. (2008) provided a simulation-based tool for planning and designing railroad container terminals.

4. Adapting queuing theory for container flow time analysis

The theory of waiting lines is based on probabilistic analysis of service systems. In a service system, customers arrive according to some random process. If a server is available, a customer proceeds immediately into service. Service commences and requires a random amount of time, after which the customer departs the system and the server is released. If on the other hand all servers are busy, the customer waits for the next available server. The expected waiting time (i.e., the probabilistic average waiting time) is a function of the probability distributions for customer inter-arrival times and service times in the service system. An important and widely used formula from queuing theory (see, for example, Hopp and Spearman, 2001) is as follows:

$$WT = \left(\frac{ca^2 + ce^2}{2}\right)\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)}\right)\left(\frac{PT}{A}\right), \tag{1}$$

where WT is waiting time, ca is the normalized variance in customer inter-arrival times, ce is the normalized variance in service time (including allowance for equipment break-downs), u is the fraction of time a server is engaged in serving customers, m is the number of parallel servers, PT is the average service time ("process time"), and A is the average fraction of time the server is available to provide service (i.e., the equipment is not in break-down mode nor under maintenance and the crew is not on break). This formula, originally developed by Sakasegawa (1977), is general enough to accept two-moment data for general arrival and service distributions in a system with multiple servers.

The expected (statistical average) total time a customer spends in the system, known as the *flow time*, is expressed as

$$FT = WT + SFT, \tag{2}$$

where WT is the waiting time as in (1) and SFT is the *standard flow time*, i.e., the expected time the customer will be in the system once service begins. SFT expresses how long it takes the customer to transit the system when there is no waiting for a server, while PT expresses how long the server is consumed serving one customer. In many applications, SFT and PT are identical, but in some situations they are not. For example, a system may consist of a single bottleneck step that may entail considerable waiting time plus other preceding and following steps with generous capacity involving little or no waiting.

In the study of containerized imports we are concerned about the impacts on container flow times resulting from changes in utilization (arising from changes in traffic level, changes in available facilities, and/or changes in hours of operation). As an approximation, we can assume that, without technological change, the terms in (1) concerning inter-arrival variability, server availability, process time and standard flow time are constant when we make modest changes to traffic volume, operating hours or facility counts. We also assume that technology is very similar across alternative facilities at the same stage of the supply chain, i.e., that values for A, PT, SFT and variability parameters are very similar across different facilities performing the same function. This suggests that container flow time through alternative facilities at any particular stage of the logistics chain satisfies (approximately) the following equation:

$$FT = a\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)}\right) + b, \tag{3}$$

where a and b are constants reflecting inter-arrival variability, server availability, process time and standard flow time at that stage, and the middle term includes parameters concerning utilization and number of servers as defined for application of (1) above.

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002

Table 1
Productivity data for rail intermodal terminals.

Terminal[a]	Est. 2006 lifts per crew shift per acre	Est. 2006 utilization
T1	0.94	0.704
T2	0.79	0.596
T3	0.81	0.605
T4	0.81	0.609
T5	0.93	0.699
T6	0.86	0.642
T7	0.82	0.615
T8	1.00	0.752
T9	0.66	0.492
T10	0.86	0.648
T11	0.94	0.702
T12	0.67	0.502
T13	0.98	0.738
T14	0.49	0.369

[a] Terminal names are coded in order to protect confidentiality of data furnished by the railroads. Terminal acreages and operating schedules are confidential.

The analytical strategy taken in this study is to statistically fit Eq. (3) to industry data, i.e., to estimate the values of a and b for the various logistics-chain queues described above, specifically, for container flow times through rail intermodal terminals and rail line-haul segments. The development of models for these applications is described in following sections. In particular, we will elaborate on the interpretation of servers and metrics of their utilization for various types of facilities.

5. Rail terminal congestion modeling

The 2006 statistics were provided to us in confidence by BNSF and Union Pacific concerning (1) average time from completion of loading stack trains ("release") until train departure from on-dock terminals at selected West Coast ports, and (2) available acreage and staffing schedules at 14 off-dock West Coast intermodal terminals, and the average time from in-gating of container-on-chassis until train departure in each month as a function of the total monthly terminal lift volume.

For on-dock trains, actual times from release to departure varied from 2.1 h to 8.4 h. Given the small overall flow time, it did not seem worth the trouble to model this in great detail. A simple weighted average of times from release by the port terminal to the railroad until train departure is 7.1 h.

Container flow times through off-dock terminals were much more significant. A standard industry metric of inter-modal terminal productivity is lifts per acre per unit time. But such a metric is inadequate from the point of view of applying queuing theory, because it does not account for the number of loading crews nor the hours such crews are on duty. For application of the queuing model to off-dock rail intermodal terminals, the number of servers m is taken as the number of crews working in parallel to load railroad well cars. Each crew mans either an overhead gantry crane to lift containers off chassis or the ground and place them in railroad well cars. Utilization of a rail terminal loading crew is more problematic to define. There needs to be a definition of the maximum capacity of the loading crew. For terminals manned three shifts per day by one crew lifting containers onto rail well cars, industry-reported import lifts per acre per working day (where a full working day includes three shifts of operation) generally are in the range of 5–10 lifts per acre. To establish a utilization figure, we assumed 12 lifts per acre per working day is the capacity of one loading crew working 24 h, i.e., 4 lifts per acre per 8-h crew shift, or 0.5 lifts per acre per working hour is equivalent to 100% utilization of one loading crew. Utilization is then computed as follows:

Lifts per acre per crew hour = (Actual lifts per acre in the month)/[(No. of crew hours worked in the month)].
Utilization u = No. of import lifts per acre per crew-hour/0.5.

For example, suppose a terminal handled 22,371 containers in a month over 22 working days. Each working day had three shifts, with one loading crew on duty each shift. The terminal has 170 acres. Then $u = \{[22{,}371/170]/24 * 22 * 0.5\} = 49.85\%$.

Table 1 displays data for eleven West Coast rail intermodal terminals used to calibrate the model.

The result of fitting Eq. (3) to these data and statistics on actual container dwell times using regression is a = 8.76 h (0.365 days) and b = 8.01 h (0.334 days). A comparison of actual data to predictions of the queuing model is presented in Fig. 1. As may be seen, the agreement between actual container flow times and predictions of the queuing model is very good, and the model seems suitable for predicting changes in dwell time as a function of terminal infrastructure and staffing.[2]

[2] The cluster of actual and predicted delay points on the lower right of the graph are for two terminals operated with two loading crews working in parallel. Other data points are for terminals staffed with a single loading crew per shift. Thus these points do not line up on a utilization vs. flow time curve passing through the data points for the terminals staffed with a single loading crew.

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002

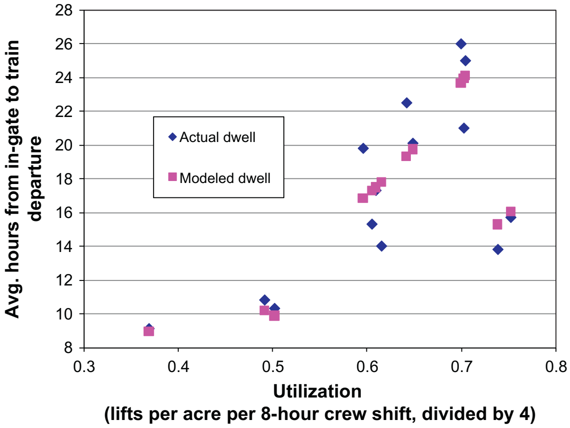

Fig. 1. Actual vs. modeled dwell times at rail intermodal terminals (West Coast terminals handling domestic and/or marine containers).

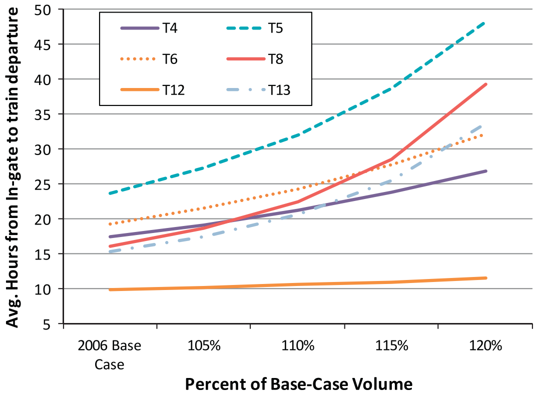

Fig. 2. Predicted container dwell times at selected rail intermodal terminals (2006 acreage, staffing and operating hours).

To illustrate analysis using the model, we took 2006 terminal volumes (hereafter called the ‘‘2006 Base Case’’), scaled the volumes upwards uniformly for all terminals up 20%, and applied the queuing model to predict the increase in container dwell times assuming acreage, staffing and operating hours fixed at 2006 levels. The results are displayed in Fig. 2 for six selected terminals. As may be seen, terminals T5, T8 and T13 are the most sensitive out of this group of terminals, with dwell times growing by factors of two or more for a 20% increase in volume.

6. Rail line-haul congestion modeling

Rail line hauls from West Coast ports to Midwestern gateways such as Chicago or Memphis comprise runs of more than two thousand miles through a rail network consisting of segments with varying traffic levels and varying track configurations. To estimate container transit time across the rail line-haul network, separate analyses were developed for each segment. In particular, different queuing models were developed for single-track and multiple-track segments of the network, discussed as follows.

6.1. Single-track main line segments

When there are opposing train movements on single-track rail lines, one train must pull off the main track and stop in a passing siding. Generally, single-track rail lines are engineered with passing sidings spaced at roughly equal intervals of running time, e.g., spaced about 10 min apart for trains achieving the main-track speed limit. At locations where trains are liable to make conditional stops, e.g., at crew terminals or mountain summits, multiple sidings may be provided. The time the single-track segment is allocated to serve a train movement is

$$PT = (D + TL)/V + 5/60, \quad (4)$$

where PT is the queuing-theoretic process time expressed in hours, D is the distance across the segment in miles, TL is the train length in miles, and V is the speed in MPH over the track segment. To the total time the segment is occupied by a train movement we add a minimum lag time, assumed to be 5 min. This minimum lag time accounts for (1) the time required for the train dispatcher to recognize that a train has cleared the limits of the segment, and to re-set switches and signals for a following movement, and (2) the lost time for a second train to proceed with caution from an approach signal until it recognizes a clear signal authorizing movement into the segment (the distance between the approach signal and the signal at the start of a single-track segment is typically 5000 ft. or more), and (3) the consequent time required to accelerate up to track speed.

Apart from time spent waiting in a siding for one or more opposing movements, there is extra delay from the loss of speed slowing down to enter the siding and accelerating back to track speed after leaving the siding, as well as a delay to release the train brakes once clearance to proceed is secured. In effect, the process time over a single-track segment is longer for a stopped train. We notate this extra time requirement when a train takes siding, i.e., the amount of extra time compared to (4), the process time for a through train movement, by L. We estimate L as follows:

$$L = \frac{2}{60} + \frac{V/C}{3600}, \tag{5}$$

where C is the assumed acceleration rate, expressed in miles per hour per second (MPHPS). A value of 0.25 MPHPS was assumed for C. L includes an allowance of 2 min to release the brakes after clearance is secured (first term), plus the time required to accelerate the train back up to track speed (second term). During the acceleration, we approximate the train's average speed as $V/2$, so only half this time is lost. It is assumed that there is an equivalent amount of loss to slow the train down when entering the siding.

The values of the fixed parameters assumed in Eqs. (4) and (5) were confirmed from review of centralized traffic control (CTC) graphical records of the Southern Pacific Railroad in PRC Voorhees (1981).

Inevitably, train movements on busy single-track railroads tend to get fleeted, i.e., a stretch of single track tends to experience a busy period featuring a series of trains in the westbound direction, then a busy period featuring a series of trains in the eastbound direction, and so on. This practice results from efforts to minimize total train delays in light of the above fact that a train entering a single-track segment at track speed has a shorter process time over the segment than a train starting into the segment from a dead stop in a siding. This requires a modified queuing analysis detailed in Greenberg et al. (1988). We summarize their queuing formula for single-track segments as follows.

We assume traffic on the rail line is symmetric, i.e., there are an equal number of trains run in each direction. For Western USA rail main lines, this is a good assumption. Let N denote the average or expected number of trains per day over the segment, including movements in both directions. The utilization u of the segment is expressed as

$$u = N \cdot PT/24, \tag{6}$$

where u expresses the fraction of time the line segment is hosting train movements (including minimum lag time), PT is defined by (4), and it is assumed that $N < 24/PT$. The probability that the track segment is experiencing a busy period in the opposite direction when a train arrives at the siding located at the start of a single-track segment is given by

$$P^{delay} = \frac{1}{2}\left[\frac{(e^{u/2}-1)(e^{u}+1)}{1+(e^{u/2}-1)(e^{u}+1)}\right]. \tag{7}$$

The expected delay (in hours) at the track segment is the probability of delay multiplied by the expected remaining length of the busy period when the train arrives at the siding. This is given by

$$E^{delay} = P^{delay}\left[\frac{N}{48}e^{u/2} - \frac{(PT)(e^{u/2})}{e^{u/2}-1} + L\right], \tag{8}$$

where N is the number of trains per day using the segment, P^{delay} is given by (7), u is given by (6), L is given by (5) and PT is given by (4).

A busy period of trains moving in the opposite direction is not the only source of delay on a single-track railroad. A fast-moving train may overtake a slower train moving in the same direction, and then lose time until the slower-moving train can be shunted off into a siding. The slow-moving train also experiences delay when overtaken by a faster train. Such delays are not accounted for by (8) but will be taken up in the next section.

6.2. Delays for following and overtaking slower traffic

In multiple-track segments of the rail network, a current-of-traffic can be established on pairs of main tracks whereby each track normally hosts train movements in only a single direction. If all trains in the same direction moved at the same speed, there should be little or no delays. But different types of trains move at different speeds. To allow fast trains to overtake slower trains moving in the same direction, some trains must be routed against the current of traffic, generating potential delays to traffic in one direction or the other. We would expect such delays to rise as utilization rises, but to be ameliorated as more main tracks are made available. Thus we expect the overtake delay function to behave like the generic queuing formula (3). This formula is applied to both single-track and multiple-track segments, where m is taken as the number of main tracks and u is computed as

$$u = (N \cdot PT)/(24 \cdot m). \tag{9}$$

A couple of modifications to (3) are required. Because we calibrate the unknown constants a and b over data from multiple track segments with varying process times PT, we need to include process time in the formula. Moreover, we will be calibrating to total train transit time, not just the waiting time, so we must include the standard flow time (SFT). The total expected (i.e., average) transit time for a train movement across the segment is the expected delay for opposing-movement busy periods (if single track), plus the delays for overtakes (following the form of Eq. (3)), plus the standard flow time, plus time allowances for conditional stops (re-crewing and refueling). SFT is computed as

$$SFT = D/V. \tag{10}$$

The overall formula used for the estimated flow time, i.e., the total transit time over a series of single- or multiple-track segments comprising an intermodal route, is

$$FT = \sum_{i \in ST} E_i^{delay} + a \sum_i \left(\frac{u_i^{\sqrt{2(m_i+1)}-1}}{m_i(1-u_i)} \right)(PT_i) + b + \sum_i SFT_i + ncc \cdot scc + nrf \cdot srf, \tag{11}$$

where the subscript i has been added to refer to line segment i within the overall origin–destination route. The first term in the expression for route flow time expresses the estimated waiting time (dispatching delays) for opposing movement busy periods on single track, computed using (8). The notation "$i \in ST$" in the first term limits the summation to the segments belonging to the set ST of single-track segments within the route, whereas the sums in the second and fourth terms are over all line segments. The second term expresses the delays due to overtakes. For each segment i, PT_i is computed using (4) and u_i is computed using (9). Here, a may be thought as the multiplicative parameter accounting for track non-availability (when possessed by maintenance staff or weather-related disruptions), as well as for the variability in train inter-arrival times and segment process times. The parameter b comprising the next term is a fixed factor to account for train delays independent of track capacity. The fourth term, the sum of the segment standard flow times, expresses the theoretical running time over the route. SFT_i for each segment i is computed as in (10). The last two terms express allowances for crew changes and locomotive refueling, where ncc denotes the number of crew changes on the route and nrf denotes the number of locomotive refueling stops on the route. The parameter scc denotes the assumed standard time, in hours, for a crew change and is computed as

$$scc = 0.25 + TL/12.5, \tag{12}$$

i.e., scc is taken as 15 min plus the time to move one train length at 12.5 miles per hour. The parameter srf denotes the assumed standard time, in hours, for performing locomotive refueling. It is set to be 1.5 h in this study.

6.3. Rail network database

A database of track configurations was compiled for the major rail intermodal corridors from West Coast ports to Midwest cities that are major destinations for intermodal trains from the West Coast port cities. Specifically, track configuration data was developed from the ports of Seattle, Tacoma, Oakland, and Los Angeles–Long Beach to the destinations Minneapolis, Chicago, Kansas City, Memphis, Dallas and Houston. Each rail corridor from a port to a destination was broken down into a series of line segments hosting approximately uniform numbers of through train movements. These segments were further broken down into sub-segments with constant numbers of main tracks.

The database, entirely developed from publically available information (Altamont Press, 2007), specifies the segment length, estimated average train speed, number of main tracks. The number of sidings (single-track segments) or the number of main-track crossover locations (multiple-track segments) also is specified.

6.4. Calibration of the rail line-haul model

BNSF and UP provided us with confidential information concerning 2006 peak and off-peak train counts by segment and concerning peak and off-peak average times from train departure at origin terminal to arrival at destination terminal. BNSF selected a single week from the peak time of year and a single week from the off-peak time of year to provide average train counts and average transit times. UP provided the same data types, but their data reflected separate averages over all peak-time-of-year weeks and over all off-peak-time-of-year weeks during 2006. Separate counts were provided for marine stack trains, domestic-box intermodal trains, and other trains. Separate transit time statistics were provided for marine and domestic stack trains.

These data were used to calibrate the queuing model. For a particular service in a particular corridor, say, domestic intermodal on the Seattle–Chicago run, the total train counts on each segment of the run were used to compute the segment utilization, which in turn was plugged into the queuing model (11) to provide an expression for the total transit time in the corridor as a function of the unknown parameters a and b. Such expressions for all intermodal corridors in peak and off-peak periods were then treated as a data set for the statistical calibration of the parameters a and b. Separate calibrations were

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002

Table 2
Statistical parameters of the rail line-haul transit time model.

Service type	*a* (h)	*b* (h)
Domestic intermodal	3.506	5.53
International intermodal	1.224	51.41

made for trains hauling domestic containers (domestic intermodal service) and trains hauling marine containers (international intermodal service).

In some cases, there are alternate routes comprising a particular corridor. As an example, UP can route its Los Angeles–Chicago intermodal trains via El Paso or via Salt Lake City. As another example, BNSF can route its Tacoma–Chicago intermodal trains via Wenatchee, WA or via Vancouver, WA. We made an allocation of trains to routes, consistent with total train counts provided by the railroads, in order to calibrate the model. The results of calibration using regression are shown in Table 2.

Generally, the railroads achieve faster transit times for domestic intermodal service than for international intermodal service. The average value of goods shipped on domestic intermodal trains is higher, as these trains carry the higher end of the value spectrum for imports and they carry time-sensitive domestic freight such as wine, canned goods and package express. The domestic intermodal trains are generally provided with more locomotive horsepower per ton and are given dispatching preference (where practical). In light of this practice, the differences in the figures in Table 2 can be understood. As may be seen from the values for the fixed factor b, there is about 5.5 h of transit time for domestic intermodal trains and 51.4 h of transit time for international intermodal trains not explained by the parameters in (11), i.e., not explained by track capacity, ordinary running time, re-crewing and refueling. These extra times may reflect phenomena such as stops to pick up or set out at intermediate terminals, mechanical problems en route or inadequate horsepower to make track speed, time held out of terminals because of terminal congestion, and reduced dispatching priority. The larger value of the fixed factor b for international intermodal service is expected.

The coefficient a, concerning sensitivity to the tightness of track capacity, is much larger for domestic intermodal service, about 3.5, compared to a factor of about 1.2 for international intermodal. The achieved transit times for domestic intermodal trains depend to some extent on their high dispatching priority; as a line gets loaded up with trains, or if the line has fewer tracks, it becomes increasingly difficult to provide priority preference to certain trains. A rail line is most fluid and experiences the least total dispatching delays when all trains are afforded the same priority and move at comparable speeds. Hence the larger value of the queue-factor coefficient in the formula for domestic intermodal service is not surprising.

Fig. 3 displays a comparison of actual average transit times provided by the railroads to predictions of the model. The "total queue factor" on the horizontal axis of the graph is the value of the utilization-server term of the queuing model, i.e., the value of the coefficient on a in Eq. (11). Agreement is far from perfect; the actual data shows disparity in average transit times ranging from 70 to 170 h, but the model can only predict disparity in the range 80–130 h. Some of this disparity likely reflects the limited amount of data on which the model is calibrated.[3] Nonetheless, we submit that the model does explain much of the difference in transit times among the corridors, traffic levels and service types and is therefore useful for analytical studies of the entire Asia–USA import trade flow.

To illustrate the use of the model in predicting transit time reductions afforded by potential infrastructure improvements, we re-applied (11) to the 2006 data after modifying the database to assume the completion of double-tracking UP's Sunset Route from Southern California to El Paso and completion of double-tracking of the remaining gaps in BNSF's "Transcon" line from Southern California to Kansas City. Results of this analysis are portrayed in Fig. 4. Shown are the model-predicted average transit times (averaged across the two railroads) for the 2006 peak traffic period for the 2006 infrastructure and for a hypothetical case where the current railroad double-tracking projects would have been completed in time for the 2006 peak season. As may be seen, the double-tracking would have reduced 2006 peak-season average intermodal transit times 2–8 h, depending on the corridor.

As another example of use the rail line-haul queuing model, the late 2006 infrastructure was subjected to increasing rail intermodal traffic levels, and the model was used to predict the increases in rail transit times for the various corridors. Fig. 5 portrays the results for peak-traffic-period domestic intermodal service in selected corridors. The "2006 Base Case" refers to peak-period traffic levels during 2006. Averages across the two railroads of model-predicted transit times are shown.[4] In this scenario, all non-intermodal rail traffic is assumed to experience zero growth; while all intermodal train movements are assumed to grow at the rates shown.

As may be seen, without infrastructure improvements, transit times in the Southern California–Memphis and Southern California–Dallas corridors are growing more quickly than transit times in the other selected corridors. The growth rate

[3] During 2006 both railroads carried out track capacity expansion projects and/or major track renewal projects in corridors serving the Southern California ports. These projects at times may have been disruptive of train service, resulting in extraordinary delays. This might account for the cluster of actual transit times around 150–170 h, which seem to be outliers from the rest of the data. (The model predicts transit times of 130–140 h for this cluster.) Another concern is that averages over only one week of operation of BNSF trains during peak and off-peak periods was provided to calibrate the model. The relatively infrequent but large disruptions characteristic of contemporary railroading are probably not present in these data.

[4] The 2006 Base Case transit times are those predicted by the model, not actual times.

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002

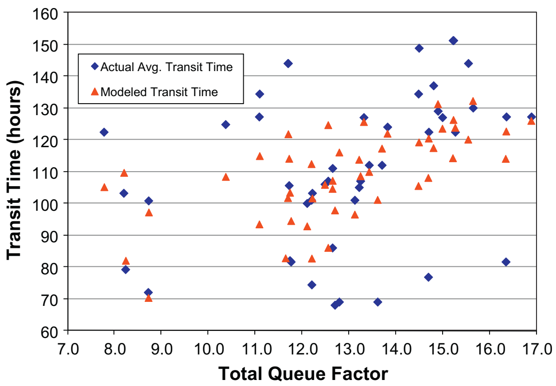

Fig. 3. Comparison of actual and modeled rail intermodal train transit times.

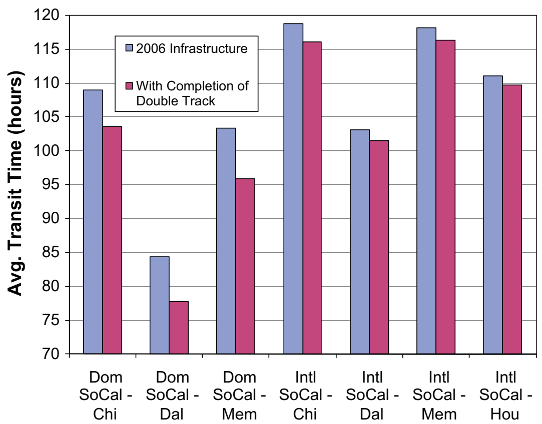

Fig. 4. Predicted transit time reductions from double tracking (2006 peak period).

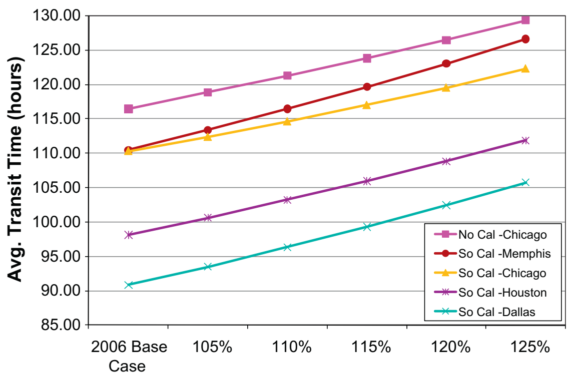

Fig. 5. Predicted increase in peak-period transit times for domestic-box. Intermodal trains, as a function of intermodal traffic growth. Note: Growth rates apply only to intermodal traffic. Zero traffic growth is assumed for non-intermodal traffic. All transit times, including for the Base Case, are model-predicted values, not actual data.

of transit time in the Southern California–Houston corridor is next largest. Growth rates of transit times in the Northern California–Chicago and Southern California–Chicago corridors are slowest and are very similar. This is just an example of the type of analysis that may be performed using the model. The story could be quite different once growth rates in non-intermodal traffic are included (especially export coal).

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002

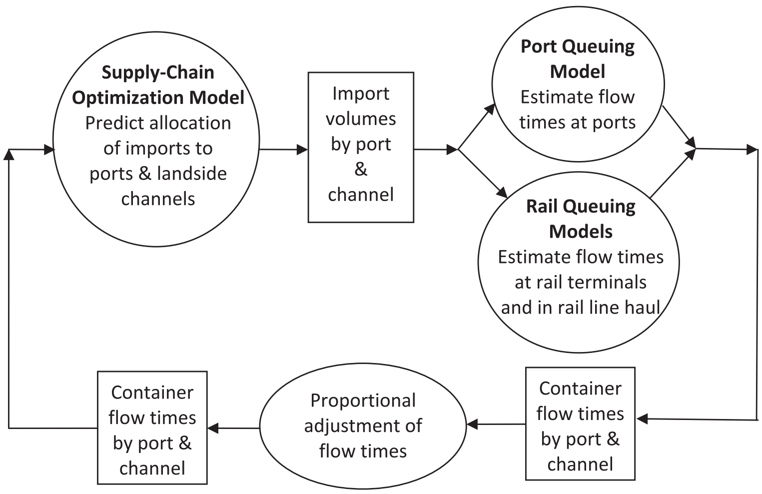

Fig. 6. The structure of the Short-Run Elasticity Model and its components.

7. Application in national-level analysis of import flows

In research sponsored by the Southern California Association of Governments, the authors have developed so-called Long-Run and Short-Run Elasticity Models to help answer the questions of whether or not importers will use new infrastructure or increased terminal staffing in exchange for higher fees (Leachman, 2010; Jula and Leachman, 2011a). The Long-Run Model assumes the mean and standard deviation of container flow times by port and landside channel are fixed, implicitly making the assumption that investments in infrastructure and staffing levels would be made as necessary to maintain flow times in the face of increased volume or share of total imports. The Short-Run Model takes given infrastructure levels for ports and railroads as input and estimates the allocation of import flows to ports and landside channels, optimizing the approximate supply-chain costs for all importers. Costs considered include all transportation and handling costs borne by the importers, plus inventory holding costs for pipeline inventories and for safety stocks maintained at the RDCs. This model involves iterative calculations of a supply-chain optimization model and queuing models, as depicted in Fig. 6. The supply-chain optimization model minimizes total transportation, handling and inventory costs for importers, taking as given the container flow times by channel (Jula and Leachman, 2011a, 2011b). The Queuing Models include a Port Queuing Model and Rail Queuing Models. The Port Queuing Model is calibrated to estimate the time to process import containers through port terminals from vessel arrival until completion of loading of double-stack trains or departure on highway dray (Leachman and Jula, in press), while the Rail Queuing Models incorporate the queuing formulas developed in Sections 5 and 6. Collectively, the results of application of these models plus fixed factors are used to develop estimates of total flow times by channel for the alternative supply-chain strategies.

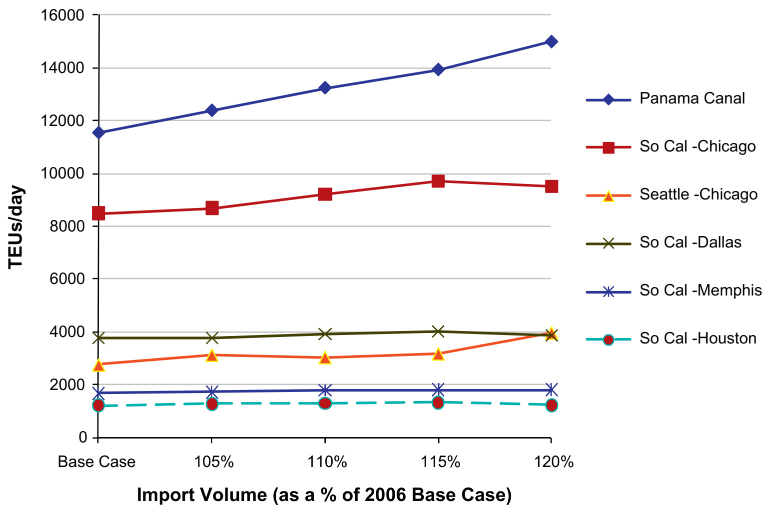

Fig. 7. Predicted growth by channel of Asia–USA imports if infrastructure and staffing are fixed at 2006 levels.

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002

The Queuing Models are used to estimate changes in container flow times by channel as a function of changes in channel volumes calculated by the supply-chain optimization model. A proportional control factor is used to gradually adjust flow times in the iterations in order to secure convergence of the overall model. See Jula and Leachman (2011a) for details.

To assess the evolution of import flows as total import volume grows, the total Asia–USA import volume in the 2006 Base-Case scenario described in Leachman and Jula (in press) was scaled upwards in increments of 5% up to 120% of the Base-Case volume, and fed to the Short-Run Model for calculation of volumes by port and landside channel. A summary of results by channel is depicted in Fig. 7. Depicted are channel volumes (expressed in terms of twenty-foot equivalent units of imports per day) for the all-water route via the Panama Canal to Gulf and East Coast ports and for selected rail corridors from the West Coast (e.g., Southern California–Chicago). As may be seen, freezing all infrastructure and staffing at 2006 levels, total importer supply chain costs are minimized if imports are proportionally routed increasingly via the Panama Canal. This is especially true for imports to Southeastern points, with rail channels from Southern California to Texas and Memphis exhibiting little or no growth. Rail volumes from Southern California to Chicago rise at a slower rate than the rate of growth in Panama Canal traffic until 115% of the Base Case Volume, beyond which the Southern California to Chicago rail corridor stops growing and volume in the Seattle to Chicago rail corridor begins to grow significantly.

8. Conclusion

The contributions of this article to the literature concern the introduction of practical queuing models for estimating container flow times through rail intermodal terminals as a function of volume, infrastructure, staffing levels and operating schedules and practical queuing models for container flow times in long-distance rail networks as a function of track configurations and traffic volumes. To the best of our knowledge, the queuing model to estimate container flow times through rail intermodal terminals is unprecedented. The standard industry metric of lifts per acre has been enhanced to account for crews and crew-hours, thereby leading to an effective metric of utilization. The rail line-haul model is significant because it is practical for tandem application to generate estimates of total container flow times in long-distance channels and is therefore useful in national-level analysis of containerized import flows and integration with supply-chain optimization analysis. It represents the first time a queuing-theoretic approach has been developed for estimating transit times through large, general freight rail networks. The models have been calibrated on industry data and compared to actual flow times. The models show promise for informing strategy and policy formation efforts at a national level. In this regard, the models have been incorporated in a large planning system to predict flows of Asia–USA imports by port and landside channel as a function of available infrastructure and service schedules, prevailing rates and fees, and potential changes to same (Jula and Leachman, 2011a).

A number of avenues are available for continued research. First, the database of rail line configurations could be expanded to include the Canadian railroad main lines and American main lines east of Chicago and the Mississippi River gateways. At present, about 5% of Asian imports to the United States are routed via the Western Canada ports, but this could grow in the future. While rail movement of Asian imports entering the USA via East Coast ports was negligible in 2006, projects to enhance clearances for double stack trains enabling them to follow more direct routes from East Coast ports to Midwestern cities have recently been completed or are under planning, and such enhancements to the rail network introduce the possibility of so-called "reverse intermodal" service for Asian imports routed via East Coast ports. Second, if data were available, the scope of application of the models could be extended to model flow times for exports; in particular, queuing analyses could be developed of both containerized and bulk exports. Third, as richer data sets become available, the precision and granularity of the queuing formulas could be improved. In particular, the rail line haul queuing models could be disaggregated to incorporate explicit analysis of back-up delays arising at the lengthy and variable-duration stops made at locomotive refueling stations, train inspection points and crew-change points. Finally, given contemporary concerns for emissions and energy efficiency, it would be desirable to assess those aspects of alternative supply-chain channels with the support of the queuing analyses.

Acknowledgements

The authors are grateful to the BNSF and Union Pacific railroads for the provision of 2006 data sets concerning their intermodal terminals and line haul networks. This research was sponsored by the Southern California Association of Governments.

References

Altamont Press Publishing Company, 2007. California Region Timetable, Northwest Region Timetable, Rocky Mountain Region Timetable, Southwest Region Timetable, Mountain and Plains Region Timetable, periodically published by Altamont Press Publishing Company LLC, P.O. Box 5264, Fullerton, CA 92838.

Benna, T., Gronalt, M., 2008. Generic simulation for rail-road container terminals. In: Proceedings of 2008 Winter Simulation Conference, vols. 1–5, 2656–2660.

Bontekoning, Y.M., Macharis, C., Trip, J.J., 2004. Is a new applied transportation research field emerging? – a review of intermodal rail-truck freight transport literature. Transportation Research part A – Policy and Practice 38 (1), 1–34.

Boysen, N., Fliedner, M., 2010. Determining crane areas in intermodal transshipment yards: The yard partition problem. European Journal of Operational Research 204 (2), 336–342.

Burdett, R.L., Kozan, E., 2006. Techniques for absolute capacity determination in railways. Transportation Research Part B 40, 616–632.
Carey, M., Kwiecinski, A., 1994. Stochastic approximation to the effects of headways on knock-on delays of trains. Transportation Research Part B 28 (4), 251–267.
Caris, A., Macharis, C., Janssens, G.K., 2008. Planning problems in intermodal freight transport: Accomplishments and prospects. Transportation Planning and Technology 31 (3), 277–302.
Chen, B., Harker, P.T., 1990. Two-moment estimation of the delay on a single-track rail line with scheduled traffic. Transportation Science 24 (4), 261–275.
Dessouky, M.M., Leachman, R.C., 1995. A simulation modeling methodology for analyzing large complex rail networks. Simulation 65 (2), 131–142.
Frank, O., 1966. Two-way traffic in a single line of railway. Operations Research 14, 801–811.
Gibson, S., Cooper, G., Ball, B., 2002. Developments in transport policy: the evolution of capacity charges on the UK rail network. Journal of Transport Economics and Policy 36 (2), 341–354.
Greenberg, B.S., Leachman, R.C., Wolff, R.W., 1988. Predicting dispatching delays on a low speed single tack railroad. Transportation Science 22 (1), 31–38.
Hallowell, S.F., Harker, P.T., 1998. Predicting on-time performance in scheduled railroad operations: methodology and application to train scheduling. Transportation Research Part A: Policy and Practice 32 (4), 279–295.
Harker, P.T., Hong, S., 1990. Two moments estimation of the delay on a partially double-track rail line with scheduled traffic. Journal of Transportation Research Forum 31 (1), 38–49.
Higgins, A., Kozan, E., 1998. Modeling train delays in urban networks. Transportation Science 32 (4), 251–356.
Hopp, W., Spearman, M., 2001. Factory Physics, 2nd ed. McGraw-Hill, New York, pp. 273.
Huisman, T., Boucherie, R.J., 2001. Running times on railway sections with heterogeneous train traffic. Transportation Research Part B 35 (3), 271–292.
Ishfaq, R., Sox, C.R., 2010. Intermodal logistics: the interplay of financial, operational and service issues. Transportation Research Part E 46 (6), 926–949.
Jula, P., Leachman, R.C., 2011a. Long- and short-run supply-chain optimization models for the allocation and congestion management of containerized imports from Asia to the United States. Transportation Research Part E 47 (5), 593–608.
Jula, P., Leachman, R.C., 2011b. A supply-chain optimization model of the allocation of containerized imports from Asia to the United States. Transportation Research Part E 47 (5), 609–622.
Kozan, E., 2006. Optimum capacity for intermodal container terminals. Transportation Planning and Technology 29 (6), 471–482.
Krueger, H., 1999. Parametric modeling in rail capacity planning. In: Proceedings of the 1999 Winter Simulation Conference, pp. 1194–1200.
Leachman, R.C., 2010. Final Report – Port and Modal Elasticity Study, Phase II. Consulting Report Prepared for the Southern California Association of Governments. <http://www.scag.ca.gov/goodsmove/elasticitystudyphase2.htm>.
Leachman, R.C., Jula, P., in press. Congestion analysis of waterborne, containerized imports from Asia to the United States. Transportation Research Part E. doi:10.1016/j.tre.2011.05.010.
Lu, Q., Dessouky, M.M., Leachman, R.C., 2004. Modeling of train movements through complex networks. ACM Transactions on Modeling and Computer Simulation 14, 48–75.
Murali, P., Dessouky, M., Ordóñez, F., Palmer, K., 2010. A delay estimation technique for single and double-track railroads. Transportation Research Part E 46 (4), 483–495.
Özekici, S., Şengör, S., 1994. On a rail transportation model with scheduled services. Transportation Science 28 (3), 246–255.
Petersen, E.R., 1974. Over the road transit time for a single track railway. Transportation Science 8, 65–74.
Petersen, E.R., Taylor, A.J., 1982. A structured model for rail line simulation and optimization. Transportation Science 16, 192–206.
Sakasegawa, H., 1977. An approximation formula $L_q \cong \alpha \cdot \rho^{\beta}/(1 - \rho)$. Annals of the Institute of Statistical Mathematics 29 (1A), 67–75.
PRC Voorhees, 1981. Willamette Valley Freight Interference Study – Technical Final Report, Oregon Department of Transportation, Salem, Ore., June, 1981.
Wendler, E., 2007. The scheduled waiting time on railway lines. Transportation Research Part B 41 (2), 148–158.
Yuan, J., Hansen, I.A., 2007. Optimizing capacity utilization of stations by estimating knock-on train delays. Transportation Research Part B 41 (2), 202–217.

Please cite this article in press as: Leachman, R.C., Jula, P. Estimating flow times for containerized imports from Asia to the United States through the Western rail network. Transport. Res. Part E (2011), doi:10.1016/j.tre.2011.07.002