
Volume 47E, issue 6, November 2011
ISSN 1366-5545
ELSEVIER
TRANSPORTATION RESEARCH
AN INTERNATIONAL JOURNAL
Part E: Logistics and Transportation Review
SPECIAL SECTION
Green Supply Chain Management: Trends, Challenges, and Solutions
Guest Editors: Jiuh-Biing Sheu and Wayne K. Talley
Editor-in-Chief: W.K. Talley

Transportation Research Part E 47 (2011) 992–1004

Contents lists available at ScienceDirect

Transportation Research Part E

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

Congestion analysis of waterborne, containerized imports from Asia to the United States

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 19 March 2011
Accepted 19 April 2011

Keywords:
Congestion analysis
Intermodal freight transport
Transportation planning
Containerized imports
Queuing

ABSTRACT

A queuing model is introduced for estimating container flow times through port terminals as a function of infrastructure, staffing, and import volume. The model is statistically calibrated on industry data. Flow-time estimates of the model are aggregated with estimates from models previously developed for rail networks to develop estimates of the total container flow times from West Coast ports to inland distribution centers. Integrated with a supply-chain optimization model, the queuing formulas are used to predict import flows by port and landside channel in scenarios of total import growth, varying all-water rates, and a higher import share for nation-wide importers.

1. Introduction

Substantial growth in waterborne, containerized imports from Asia to the Continental United States up through 2007 (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 that triggered major shifts in port and channel allocations of imports. In response to trade growth, there have been major expenditures by public agencies 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.

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 flow times as a function of volume, infrastructure and staffing. While there is much useful queuing literature for operational analysis of individual terminals, to our knowledge there is little research on practical tools for congestion analysis of large import networks. We aim to fill that need in this article.

An important analytical question faced by policymakers concerns how importers would respond to new infrastructure or increased staffing hours, and to new fees or rate increases to pay for construction of the infrastructure or the addition of working hours. Would the importers "stay and pay" or would they re-structure their supply chains to avoid increased charges, shifting import cargoes to other ports and/or other landside channels?

A practical analytical means of estimating container flow times is an important element in addressing that question, i.e., it must be determined whether or not there are sufficient reductions in flow times afforded by the proposed additions to infrastructure or staffing to offset the costs of same. A second purpose of our research is therefore to combine in a practical way the results of queuing analyses of individual transportation links and terminals into estimates of the total container flow times from port of entry to inland distribution centers.

* 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
doi:10.1016/j.tre.2011.05.010

Our specific interest is waterborne 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 ports 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 various queues import containers must negotiate. Next, we review the relevant literature. We then proceed to the development of our proposed queuing models and illustrate their application. In particular, we discuss their integration into elasticity analysis of importers' response to infrastructure or staffing additions, and to fees to use such additions.

2. Overview of supply-chain strategies and supply-chain queues

Waterborne, containerized imports flow through a series of queues. The type and sequence of queues experienced by containers handling goods for a particular importer depend on the supply-chain strategy adopted by the importer. For the purposes of understanding the impact of congestion on import flows under alternative supply-chain strategies, it is therefore convenient to stratify imports by supply-chain strategy.

2.1. Classification of supply-chain strategies

Broadly speaking, in industrial practice there are two basic supply-chain strategies for managing flows of containerized imports from Asia to the Continental United States:

2.1.1. Push supply chains

Importers purchase transportation of marine containers from Asian factories to their regional distribution centers (RDCs). Allocation of containers to RDCs is decided before booking vessel passage. Landside movement to RDC may be via IPI (inland point intermodal service), whereby the marine box is loaded onto a double stack well car on-dock or drayed from the port terminal to an off-dock rail intermodal terminal (AKA a *ramp*), then moved in a double-stack train to a ramp in the general area of the RDC, then re-loaded onto a chassis for final dray to the RDC. Landside movement also may be via dray direct from port terminal to a local RDC or by over-the-road trucking to RDCs in regions not as distant as the regions for which IPI service is utilized. As of 2007, about 70% of total Asia – Continental USA imports were handled in Push supply chains (Leachman 2010).

2.1.2. Push–Pull supply chains

A set of 1–5 ports for handling all imports to the Continental USA is selected by the importer. In the hinterland of each selected port the importer maintains an import warehouse for storing goods that are imported far in advance of demands at its RDCs and for which it desires to delay making the decision to allocate goods to regions until regional demand forecasts become more reliable. Nearby each selected port the importer also contracts a trans-loader/de-consolidator to unload the contents of marine boxes, sort the imported goods by destination, and re-load the goods into domestic rail containers and highway trailers. Under Push–Pull, the decision is made before booking vessel passage as to how to allocate marine containers to the selected ports of entry (if there is more than one), but the decision as to how to allocate port volumes to RDCs is deferred. Just before vessel arrival, an allocation of the marine boxes is made to the trans-loader/de-consolidator in the hinterland of the port, the import warehouse in the hinterland of the port, and the local RDC. Most containers are routed via the trans-loader/de-consolidator; a smaller fraction is routed directly to the import warehouse. In the case of high-volume importers, a fraction of import containers may be routed directly to the local RDC. Drays of the marine boxes from the port terminal to these three destinations are made accordingly. For boxes routed to the trans-loader/de-consolidator, decisions are made just before the time of vessel arrival about how to allocate the contents of each marine box into domestic rail containers and highway trailers destined to various inland RDCs, the local RDC and the import warehouse. The trans-loader/de-consolidator processes the contents of the marine boxes and dispatches domestic rail containers and highway trailers accordingly. The domestic rail containers loaded by the trans-loader/de-consolidator are drayed to a nearby rail terminal, moved by train to a ramp in the general area of the destination RDC, then re-loaded onto chasses for final dray movement to the RDC. The highway trailers loaded by the trans-loader/de-consolidator are drayed to the local RDC, drayed to the import warehouse, or trucked to RDCs in regions not as distant as the regions for which domestic rail service is utilized. For boxes routed to the import warehouse, the goods in those boxes are unloaded and placed in storage. At some future times decisions will be made to allocate those goods to RDCs. For goods allocated to the local RDC, there is local dray movement. For goods allocated to distant regions, domestic rail containers are brought to the import warehouse, loaded and drayed to a nearby rail intermodal ramp. The domestic containers are moved by domestic double stack train to a rail terminal in the same area as the destination RDC, then re-loaded onto chasses for final dray movement to the RDC. For goods allocated to other regions for which rail intermodal service is not available or is not economical, the goods are loaded into highway trailers for truck movement to the RDCs in those regions. As of 2007, about 30% of Asia – Continental USA waterborne containerized imports were handled in Push–Pull supply chains (Leachman, 2010). The share of imports handled this way has been steadily rising for about a decade.

Figs. 1 and 2 depict these strategies in terms of the stages of transit and inventory and the types of transportation vehicles employed (marine container, line-haul domestic container or trailer, captive-to-region domestic trailer).

2.2. Taxonomy of supply-chain queues

The series of queues experienced by waterborne containerized imports under the alternative strategies are summarized as follows (queues experienced before vessel transit are common to all strategies and are ignored for our purposes):

Port of entry queues:

– Vessel berthing (all supply chains).
– Vessel unloading (all supply chains).
– Loading railroad well cars on-dock (on-dock rail for IPI movement in Push supply chains).
– Loading outbound drays at port terminals (dray to off-dock rail terminal for IPI movement in Push supply chains; dray to local RDC and truck to other-region RDC in both Push and Push–Pull supply chains; dray to import warehouse and dray to trans-loader/de-consolidator in Push–Pull supply chains).
– Removal of strings of loaded well cars from on-dock loading tracks, and assembly of well car strings into longer trains (on-dock IPI in Push supply chains).

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

– Processing containers at de-consolidation/trans-loading cross-docks (Push–Pull supply chains).
– Loading outbound trailers and domestic containers at import warehouses (Push–Pull supply chains).
– Highway traffic congestion impeding dray movements (dray to off-dock rail terminal for IPI movement in Push supply chains; dray to local RDC and truck to other-region RDC in both Push and Push–Pull supply chains; dray to trans-loader/de-consolidator and drays to import warehouses in Push–Pull supply chains; drays from import warehouse or trans-loader/de-consolidator to domestic container rail terminals in Push–Pull supply chains).
– Loading marine containers or domestic containers at off-dock rail terminals (Push–Pull supply chains and off-dock IPI in Push supply chains).

Queues in line-haul movement and in destination handling:

– Delays to line haul movement of double-stack container trains for meets with opposing rail traffic on single-track lines, and for following and overtaking slower rail traffic moving in the same direction (all supply chains).
– Highway traffic congestion impeding truck movement to near-region RDCs (all supply chains).
– Unloading containers from stack trains onto chasses for destination dray movement (all supply chains).
– Unloading containers at destination RDCs (all supply chains).

There is considerable disparity in the relative impacts of these various 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 impacts in aggregate measured in days. We approximate the delays stemming from other queues we perceive to be relatively insensitive to import volume with fixed distributions. We have already treated rail-related queues (Leachman and Jula, submitted for publication). More detailed explanation of not-rail-related queues is provided as follows.

2.3. Queues at port terminals

In general, each import container experiences the following phases as it passes through a port terminal: unloading from vessel, storage, and transfer to land-based transportation services (i.e., loading on a truck chassis or loading into a railroad

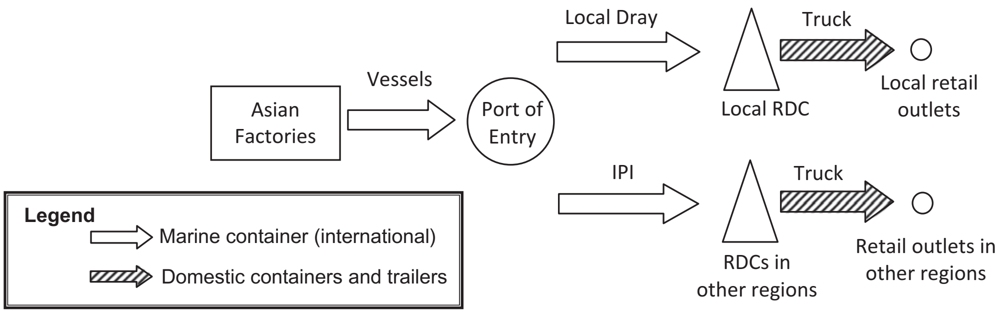

Fig. 1. Push supply chain.

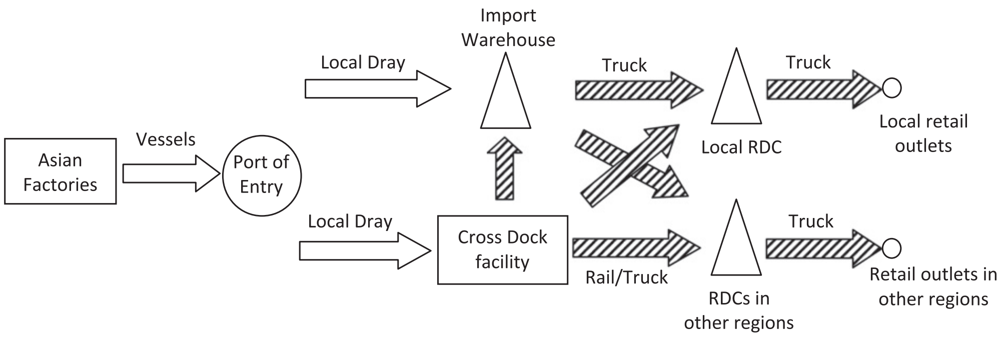

Fig. 2. Push–Pull supply chain.

double-stack well car). Some ports also serve to trans-ship containers between vessels, but such terminals are not within our scope.

In the unloading phase, containers are unloaded and transported from the vessel to the storage yard or to a railcar-loading staging area, both typically close at hand. The equipment involved generally includes dockside cranes at berths, terminal trucks (drays) to move the box from dockside to storage areas or to on-dock rail loading areas, and rubber tired gantry (RTG) cranes or top lifters storing and retrieving containers at storage yards and rail loading areas. The dockside cranes are in charge of lifting containers from the vessels and dropping the boxes on the intra-yard drays. The RTGs work the container storage area and are responsible for storing the containers and for retrieving the containers once the outbound dray is positioned and ready for loading. Top-pickers load railroad double-stack well cars. Top-pickers also can put away containers in storage or staging areas if they can be stored on the immediate top edge of the storage stacks, and if outbound drays from storage or staging areas can be dispatched in any order (so that the box on the top edge of the cube can be loaded first).

Typically at USA ports, an import container experiences two lift cycles in a port terminal, performed by separate crews. The first cycle is a lift out of the ship and placement on an in-yard dray for movement to a temporary storage or staging area within the terminal. The second cycle is a lift out of a position in the staging or storage area onto a truck chassis or into a railroad double-stack well car (in the case of on-dock rail) for movement out of the terminal.

2.3.1. 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 outbound drays or in railroad well cars (if the terminal has on-dock rail service). Typically, separate cubes are maintained for containers to be loaded in rail well cars, for containers to be drayed off the dock, and for export containers (to be loaded on vessels). Each cube has a separate workforce and in effect is a separate queuing system. Generally, the terminal has little or no control over, or information about, the timing and sequence in which import boxes will be picked up by draymen. As the vessel size grows or as the total import volume pushed through a given terminal acreage is increased, the average height of the stacks comprising the dray cube grows, and it becomes more laborious to retrieve outbound boxes for drays. The 2004 melt-down at the San Pedro Bay ports (i.e., the Ports of Los Angeles and Long Beach) was most strongly manifested in this queue.

2.3.2. Vessel berthing and vessel unloading

Vessel arrivals at some US ports (especially the San Pedro Bay ports) are quite peaked across the days of the week. While the port terminal crews handling operations to dispatch import containers out of the terminal and to receive export containers into the terminal work fixed weekly shift schedules, gangs handling vessel loading and unloading work on an on-call basis timed to the actual vessel calls. Considering the flexible staffing and considering their investment in berths and terminal space, to date the ports have been able to unload and re-load the vessels so as to maintain this peaked pattern. Were they not able to do so, vessel arrivals could be spread more uniformly across the days of the week. Berthing and unloading queues, while important to steamship lines for managing vessel utilization, do not seem to be as serious (in terms of their impact on container flow times) as the outbound queues at the port terminal. Because of weather variability and other factors, vessel transit times across the Pacific are somewhat variable. As a modeling strategy, we choose to aggregate vessel berthing and unloading time with vessel transit time, and we assume this aggregate time is independent of volume. Mean and standard deviation statistics on vessel transit time (including the berthing and unloading time) are used in the computation of safety stock requirements in supply chains (Jula and Leachman, 2011a).

2.4. Queues outside the port terminals

2.4.1. Trans-loaders/de-consolidators

Vessel arrivals at certain West Coast ports are very peaked by day of week. For example, at the Ports of Los Angeles and Long Beach, about 70% of vessel calls occur on Saturday and Sunday. Trans-loaders/de-consolidators work steady 5-day-per-week operations. Arrivals of imports at their facilities are heavy on Mondays and Tuesdays, more moderate on Wednesdays, lighter on Thursdays and even lighter on Fridays. But their processing rates are fairly steady across the 5-day week. There is thus some production-smoothing delay for imports handled through their facilities. The trans-loaders/de-consolidators control the overall volume they handle contractually, so this delay tends to be fairly predictable and stable. There are minimal barriers to entry for more cross-dock service providers. Virtually any sheltered freight dock or warehouse with sufficient door spots and parking space can be adapted for cross-dock operations, and indeed the vast majority of facilities in such service are hand-me-down structures that previously performed other types of distribution activity. We choose to model the time to process imports through trans-loaders/de-consolidators as fixed and independent of volume. Mean and standard deviation statistics on trans-loader time (including the production smoothing time) are used in the computation of safety stock requirements in supply chains (Jula and Leachman, 2011a).

2.4.2. Drays from port terminals

The weekend peaking of vessel arrivals has a similar impact on drays. The dray fleet is generally sized to handle the weekly volume. The weekly dray volume tends to get spread out over five working days. At the Ports of Los Angeles and Long Beach, the queues of draymen waiting to enter port terminals to pick up import boxes at the start of their shifts on Mondays and Tuesdays are very severe. Worse, the box population at the port terminals is at the peak of the weekly cycle on these days, so time to retrieve a box inside the terminal is at its peak (as discussed above). Thus draymen may be unable to make as many trips on Mondays or Tuesdays as on other days. Fortunately, the queues for draymen to get in the gate at port terminals are generally abated by Thursday and Fridays, and the entire weekly volume is completed by the end of the week. If the entire weekly dray volume was not done by end of shift Friday, and that condition was repeated in consecutive weeks, the system would experience a melt-down. Otherwise, this queue is primarily a production-smoothing phenomenon, similar to the workload smoothing at the trans-loaders/de-consolidators, and it tends to be stable and predictable.

Peak-hour road traffic can stretch out dray trip times by an hour or so, but this is a relatively small effect compared to the queues and melt-down effects described above, which are measured in days. We choose to model dray transit times as fixed and independent of volume. Mean and standard deviation statistics on dray transit times (including the production smoothing time) are used in the computation of safety stock requirements in supply chains (Jula and Leachman, 2011a).

In the next section, we will develop and calibrate practical queuing models for estimating container flow times through the most serious type of queues: Loading drays and on-dock rail cars at port terminals. The intent of this model is to estimate the impacts on container flow times from (1) changes in container volumes or (2) changes in levels of staffing or available infrastructure.

3. Literature review

There are many published works providing an overview of port terminal operations and the equipment employed – see, e.g., Vis and de Koster (2003), Steenken et al. (2004), and Murty et al. (2005). Stahlbock and Voß (2008) provide a survey of recent literature in this area and indicate only a few studies providing integrative views of container terminal logistics have been published to date.

Currently, much research is focused on dockside problems, such as dockside retrieval, and the closely related stowage planning for export containers. For an early work on queuing theory in berth assignment and berth investment decisions see Edmond and Maggs (1978). Legato and Mazza (2001) present a queuing network model and a simulation analysis of the logistic processes of arrival, berthing and departure of vessels at a container terminal. A simplified version of the proposed model was later used by Laganá et al. (2006), who focus on the power of grid computing for solving simulation optimization problems of a stochastic nature such as the assignment of berth slots and cranes to shipping services. Canonaco et al. (2008) present a queuing network model for the management of container discharge and loading at any given berthing point. Due to its complexity, the authors use a discrete-event simulation to propose solutions and evaluate the outcomes of different policies regarding crane assignment and scheduling.

The literature also presents applications of queuing models for analyzing container loading and unloading operations. Garrido and Allendes (2002) applied a cyclic queue model to a case study of the Port of San Antonio, Chile. Kang et al. (2008) use a cyclic queuing model for fleet optimization of unloading operations at container terminals. The cyclic queuing model assumes exponentially distributed service times and steady state operations.

In the domain of interfacing with landside transport at port terminals, Powell and Carvalho (1998) propose a dynamic model for real-time optimization of the flow of flatcars considering constraints for assignment of trailers and containers to flatcars. A reduced flatcar fleet is made possible due to useful information for decision makers provided by the developed global logistics queuing network model.

Improved terminal performance cannot always be obtained by solving isolated sub area problems, but may require better integration of the various operations connected to each other. The limited research in this domain mostly focuses on using simulation to develop an analysis (see Stahlbock and Voß, 2008).

Analytical approaches that use modern queuing techniques instead of discrete event simulation in order to evaluate terminal allocation and layout planning problems can be found in, for example, Kozan (1997), and Van Hee and Wijbrands (1988). Alessandri et al. (2008) uses systems of storage and handshake queues to propose a dynamic discrete-time model of container flows in maritime terminals. The authors provide feed–back control algorithms to assign the resources in order to optimize the system.

Caris et al. (2008) provide an overview of planning decisions and solution methods proposed in the literature in the domain of intermodal freight transport systems, and find a lack of research on the strategic- and tactical-level issues facing intermodal operators. While there is a large literature of queuing analyses of port terminals, our review of this literature finds that nearly all such queuing models focus on sub-systems and seem designed to support operational decision-making (scheduling, control, and optimization) in such sub-systems.

In the following section we will propose queuing models for port terminals. 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.

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 or providing service (i.e., the equipment is not in break-down 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). To first approximation, we can assume that, without technological change, the terms in (1) concerning 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 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 (1) above.

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 container flow times through port terminals. The development of this model is described in the next section.

5. Port terminal congestion modeling

A common productivity metric reviewed by managers of intermodal terminals is containers handled per acre per year. The basis for this metric is that, as more space is made available, it is easier to make required container movements. In space-constrained port terminals loading outbound drays, the boxes must be stacked in the staging area. This results in the need to maneuver around other boxes or to lift and move other boxes out of way when the desired box is buried. Thus, utilization of more space improves productivity (i.e., it reduces service time and waiting time in a queuing-theoretic sense).

Figs. 3 and 4 illustrate the impact of acreage on port terminal productivity. Data points from two West Coast terminals are displayed. Both terminals are staffed by a single loading crew per shift, and both work around the clock five days per week. Fig. 3 provides a plot of average import container dwell time vs. number of import containers per working day. Each point is a monthly statistic. Based on these data, it might seem that Terminal B is more efficient than Terminal A in the sense that lower and more consistent dwell times are achieved while handling higher import volumes. Fig. 4 re-plots the waiting times vs. the number of import containers per working day per acre (i.e., vs. a revised volume metric accounting for the available acreage). It is now clear that Terminal A is much more congested, attempting to handle much more volume per acre. A pattern emerges: as volume per acre per working day increases, dwell times increase and become more volatile.

For the purposes of this study, we expand the industry-standard lifts-per-acre productivity metric to account for the hours of operation of the terminal. We express utilization in terms of lifts per acre per crew per hour. The idea here is that, with more hours worked per day or more crews working in parallel, throughput per day should increase in a terminal with a given acreage.

For application of the queuing model, the number of servers m is taken as the number of crews working in parallel to load truck chasses or railroad well cars. Utilization of a port terminal crew is more problematic to define. There needs to be a definition of the maximum capacity of a loading crew. For terminals manned three shifts per day by one crew lifting containers onto truck chasses or 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. (Including export lifts, total lifts are roughly double these amounts.) To establish a utilization figure, we posited 12 import lifts per acre per working day as equivalent to 100% utilization of a terminal staffed with one loading crew on duty every shift. Utilization is then computed as follows:

$$\text{Lifts per acre per crew hour} = (\text{Actual lifts per acre in the month})/[(\text{No. of crew hours worked in the month})]$$

$$\text{Utilization } u = \text{import lifts per acre per crew-hour}/0.5$$

For example, suppose a terminal handled 22,371 import 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 $u = \{[22{,}371/170]/24 * 22 * 0.5\} = 49.85\%$.

We secured monthly data for calendar 2006 from five container terminals at West Coast ports. These data include monthly import and export container volumes, number of shifts the gate was open during the month, number of loading crews on duty per shift, and average container dwell times (for imports, measured from ship docked and ready for unloading until container trucked out the gate or until on-dock rail train released to railroad).

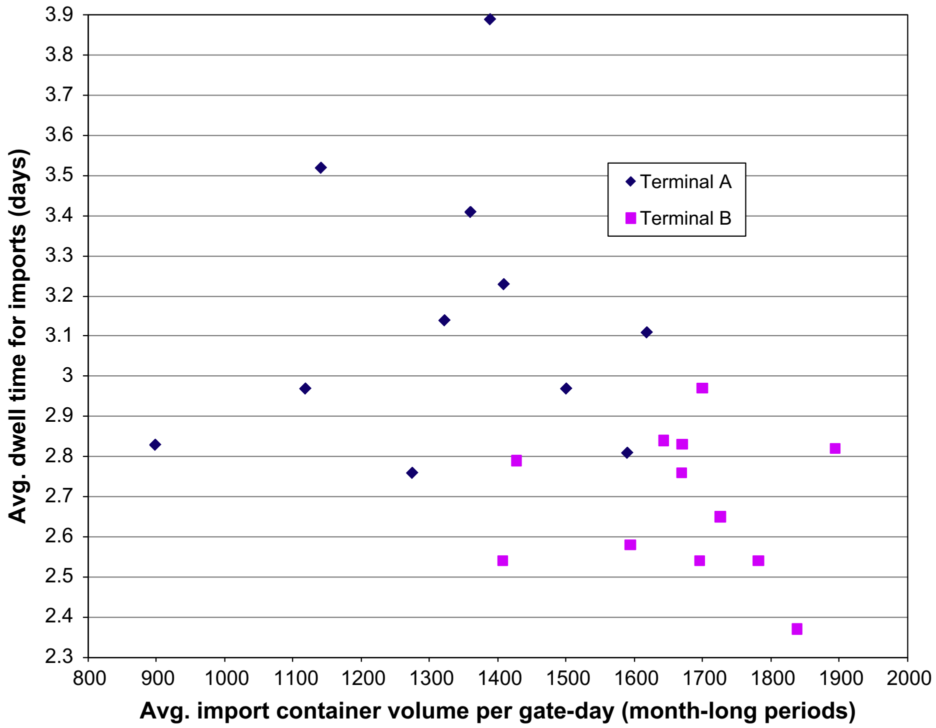

Fig. 3. Import container dwell times vs. import volume at selected terminals (one crew per shift, three shifts per day).

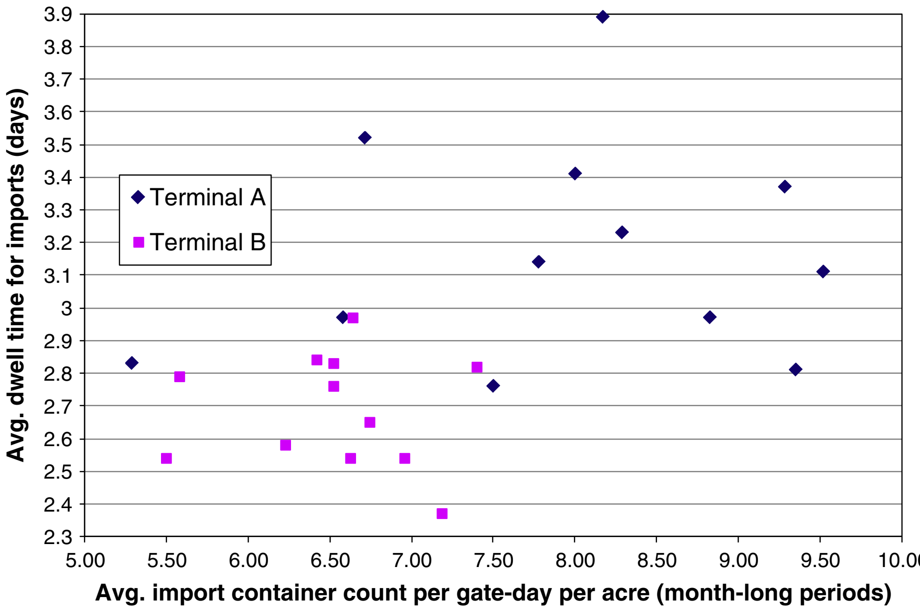

Fig. 4. Import container dwell times vs. import volume at selected terminals, accounting for acreage (one crew per shift, three shifts per day).

Formula (3) was statistically fit to these data using a least squares calculation. Under the least squares approach, coefficients a and b were selected such that the sum of squared errors of all data points is minimized, i.e., $Min \quad SSE = \sum_{i=1}^{n}(D_i - F_i)^2$, where D_i is the actual value of observation i, and F_i is its predicted value using Eq. (3). After removing a few outlier data points, we used the Excel Solver to find the best values for a and b and then confirmed the fit by visual observation. The result is a = 7.36 h (0.307 days) and b = 55.4 h (2.31 days). That is, the model predicts that import container dwell time is 2.3 days plus 0.31 times the queuing formula's utilization factor as in Eq. (3).

Fig. 5 displays a comparison of the queuing model's predictions to actual data for the two terminals whose data appears in Figs. 3 and 4. Actual container dwell times depend on a host of factors besides the productivity of the terminal crew: how quickly dray customers come to retrieve their box once they are notified (which in turn affects how high the stacks of containers grow), availability of chasses and railroad well cars, issues with unloading the vessel, boxes held up by US. Customs or Homeland Security, etc. So the real data is quite "noisy" in the sense that waiting time based on utilization plus a standard flow time does not fully explain dwell time. Nonetheless, it is asserted that the queuing model quantifies the effects of terminal congestion on average import container flow time.

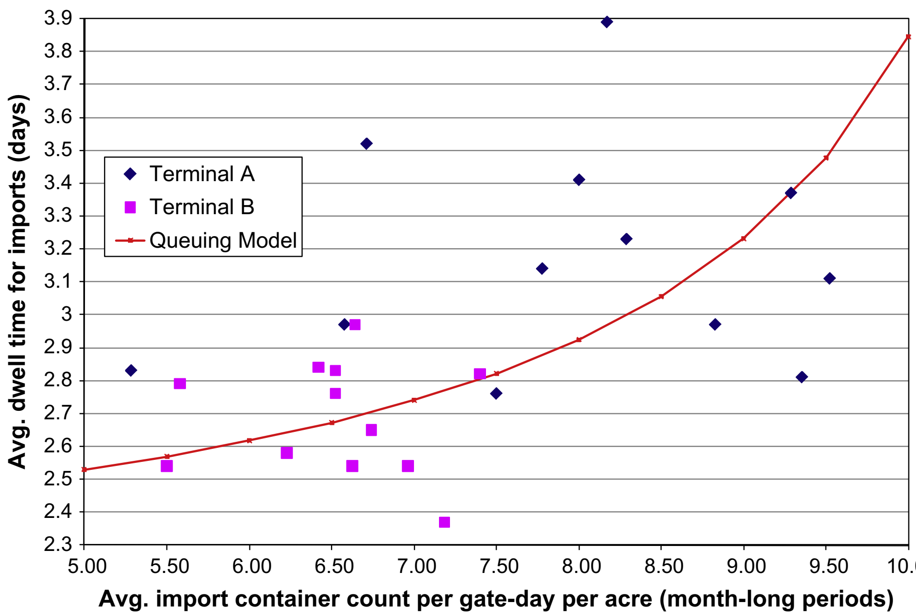

Fig. 5. Modeled and actual import container dwell times vs. import volume at selected terminals (one crew per shift, three shifts per day).

Table 1
Port terminal data.

Port	Assumed acreage available for Asia–US imports[a]	Assumed share of continental US import volume	Assumed avg. no. of crews per terminal per shift	Assumed avg. no. of gate shifts per day	2006 import containers per gate-day per acre (for the assumed market shares)	Estimated utilization
Vancouver–Prince Rupert	431	0.0277	1	3	7.70	0.721
Seattle–Tacoma	1034	0.0804	1	3	9.29	0.774
Oakland	759	0.0556	1	3	8.75	0.729
Los Angeles–Long Beach	2968	0.4589	2	3	18.48	0.770
Lazaro Cardenas	210	0.0152	1	3	8.65	0.721
Houston	345	0.0356	1.33	3	12.32	0.770
Savannah	966	0.0838	1.33	3	10.36	0.648
Charleston–Wilmington	396	0.0252	1	3	7.61	0.634
Hampton Roads	994	0.0737	1	3	8.87	0.739
NY–NJ	1002	0.1439	2	3	17.18	0.781

[a] 80% of available acreage at US East Coast ports assumed available for handling Asian imports. 20% of Vancouver acreage and 85% of Prince Rupert acreage assumed available for handling Asia–USA imports. 70% of available acreage at Lazaro Cardenas assumed available for handling Asia–US imports. Other space at these ports is assumed to be reserved for other trades. At all other ports, all acreage is assumed available for handling Asia–US imports.

To illustrate the potential use of the queuing model, consider the data in Table 1 concerning assumed available acreage, assumed staffing at various ports, and an assumed scenario of port shares of total Asia–USA waterborne containerized imports. The authors secured customs data on 2006 volumes by port of Asia–United States imports (Leachman, 2010). The 2006 import volume from Asia to US of 7,706,000 containers (12,430,000 TEUs) gives rise to terminal utilizations in the 45–72% range. Now suppose we scale the 2006 import volume by 105%, 110%, 115% and 120% without adding acreage, crews or operating shifts at any port. The predictions of the queuing model applied to entire ports are plotted in Fig. 6. For this scenario, it would seem that the San Pedro Bay Ports, Houston and New York–New Jersey would have the most urgent need to expand crewing, operating hours and/or acreage to avoid unfavorable impacts on container flow times.

The reader is cautioned that aggregating all terminals in a port into one queuing system to be presented to queuing formula (1) will underestimate container flow times if there is significant variation among the terminals in utilization and/or in numbers of loading crews working in parallel. The hyperbolic functions portrayed in Fig. 6 are not symmetric, i.e., the average waiting time across two terminals both working at 50% utilization is significantly less than the waiting time averaged across one terminal at 25% and the other at 75% utilization. For accurate results, the queuing model should be separately applied to individual terminals.

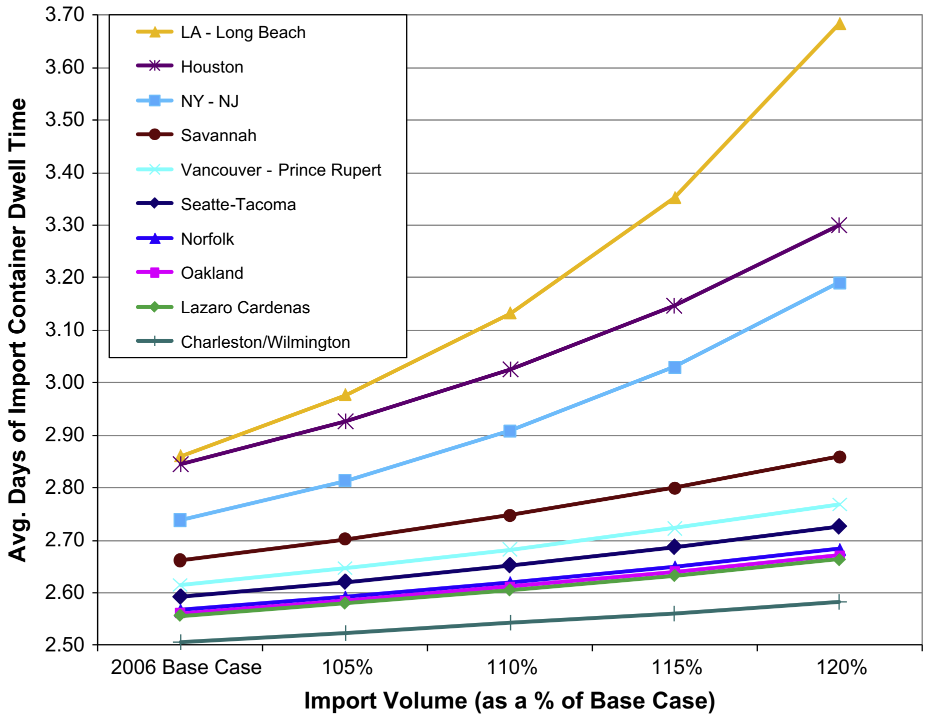

Fig. 6. Predicted port to gate flow times (2006 acreage, staffing and operating hours).

6. Application of the queuing models in elasticity analysis

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 (Jula and Leachman, 2011b). 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 a Queuing Model, as depicted in Fig. 7. The Supply-Chain Optimization Model minimizes total transportation, handling and inventory costs for importers, taking as given the container flow times by channel. The Queuing Model incorporates the queuing formulas developed in Section 4 plus fixed factors to estimate total flow times by channel for the alternative supply-chain strategies. This model is 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 (2011b) 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 Jula and Leachman (2011b) 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 port is depicted in Fig. 8. As may be seen, freezing all infrastructure and staffing at 2006 levels, the San Pedro Bay ports are anticipated to handle only about 10% more growth before congestion induces major diversion of imports to other ports. The greatest growth would occur at the Ports of Norfolk (i.e., Hampton Roads) and Oakland, with lesser growth at other North American container ports. Infrastructure for containerized imports at Norfolk and Oakland is relatively underutilized, and the model predicts more than 80% growth at Norfolk and 50% growth at Oakland for a 20% growth in total imports.

Several scenarios of future import flows overlaid with potential fees assessed on imports passing through the San Pedro Bay ports were formulated and analyzed. These include a Near-Term Likely Scenario, in which infrastructure was updated to 2009 levels and rail and steamship rates were updated to 2009 levels; a Pessimistic Scenario, in which all-water steamship line rates (for service via the Panama Canal) are dropped 10% relative to steamship rates for service to West Coast ports; an Optimistic I Scenario, in which all-water steamship line rates (for service via the Panama Canal) are raised 10% relative to steamship rates for service to West Coast ports; and an Optimistic II Scenario, in which the share of total imports accounted for by large, nation-wide retailers rises from 40% to 50%. All these scenarios assume the total Asia–Continental USA import volume is fixed at the 2006 level. The "Pessimistic" and "Optimistic" labels on these scenarios reflect the consequences on volumes routed through the San Pedro Bay ports, i.e., optimistic scenarios result in much increased volumes routed through those ports while pessimistic scenarios result in much reduced volumes. See Leachman (2010) for details.

Both Long-Run and Short-Run model calculations were made for these scenarios, overlaid with potential fees ranging from $0 to $500 per FEU (forty-foot equivalent unit) assessed on all San Pedro Bay import containers. Results are depicted in Figs. 9 and 10. The solid-line curves in the top half of the figures show the trends in total import volume routed via San Pedro Bay (expressed as a percentage of the 2006 "Base Case" San Pedro Bay import volume); the dotted-line curves in the bottom half of the figure show the trends in trans-loaded import volumes via San Pedro Bay.

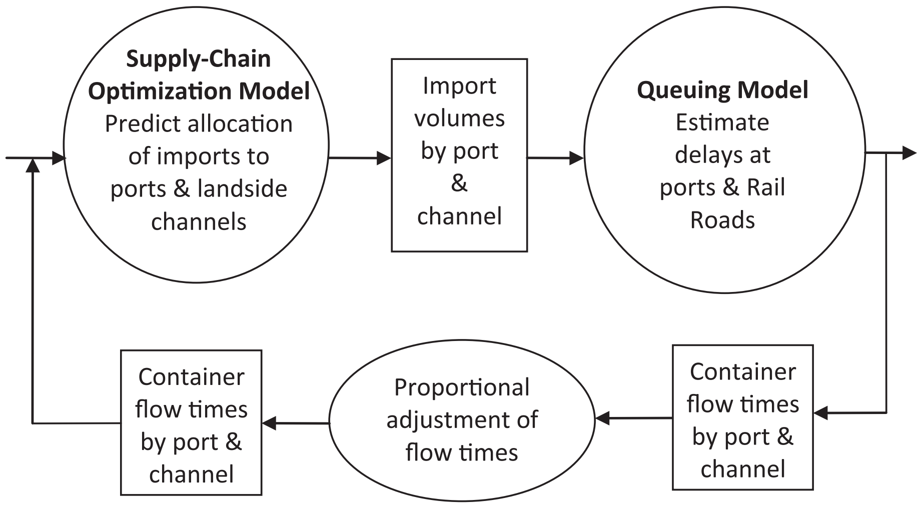

Fig. 7. The structure of the Short-Run Elasticity Model and its components.

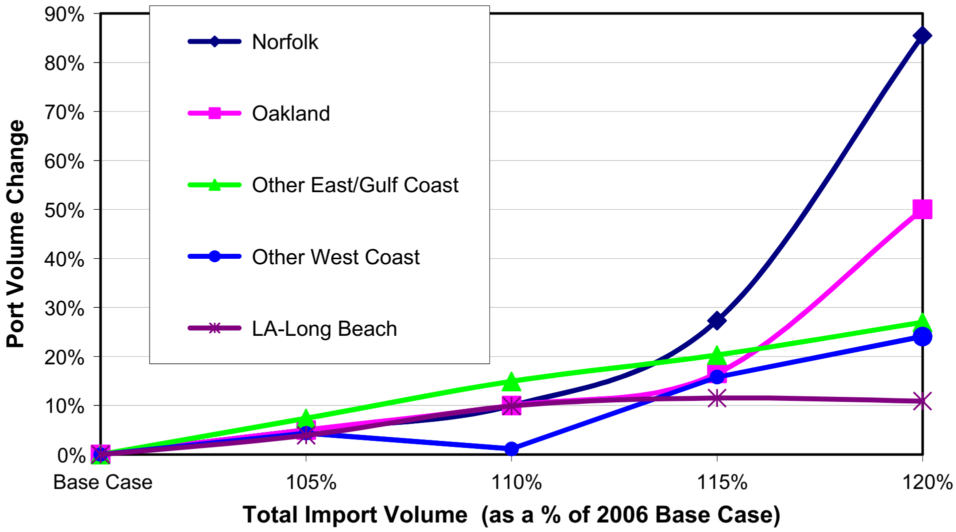

Fig. 8. Predicted growth of Asia–USA imports at North American ports if infrastructure and staffing are fixed at 2006 levels.

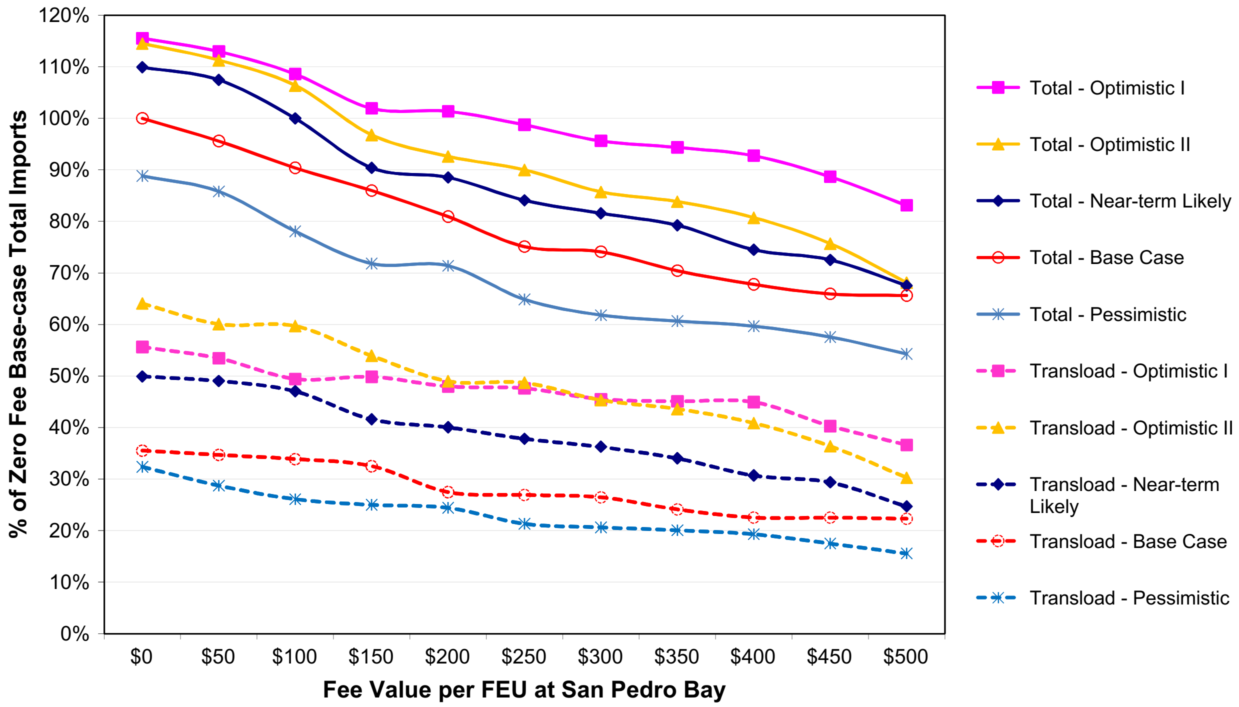

Fig. 9. Short-Run elasticities of imports via the San Pedro Bay ports in future scenarios.

Results for the Near-Term Likely Scenario indicate that changes in rail and steamship rates since 2007 have been favorable for San Pedro Bay. In the case of no new container fees, this scenario results in a 10% gain in market share in the short run and a potential 20% gain in the long run if additional infrastructure and staffing investments are made to enable 2006 container flow times to be realized at the higher volume levels. In the long run, fees up to about $75 per FEU could be assessed while still realizing the 2006 market share. Considering just the trans-loaded import volumes, fees up to about $125 per FEU could be assessed.

The Optimistic scenarios are even more favorable to the Southern California ports. Without new fees, these scenarios result in a 15% gain in market share in the short run and potentially a 30–40% gain in the long run. If new fees are instituted, these gains are abated; in the long run, a fee in the range of $125–$175 per FEU brings total market share back to the 2006 level. Trans-loaded imports in these scenarios are much less elastic to fees, exceeding 2006 volume levels until new fees rise to $250 per FEU.

On the other hand, the Pessimistic scenario exhibits serious drops in market share for the San Pedro Bay ports. Even if no new fees are introduced, a 10% drop in all-water rates results in a about a 12% drop in total San Pedro Bay imports in the

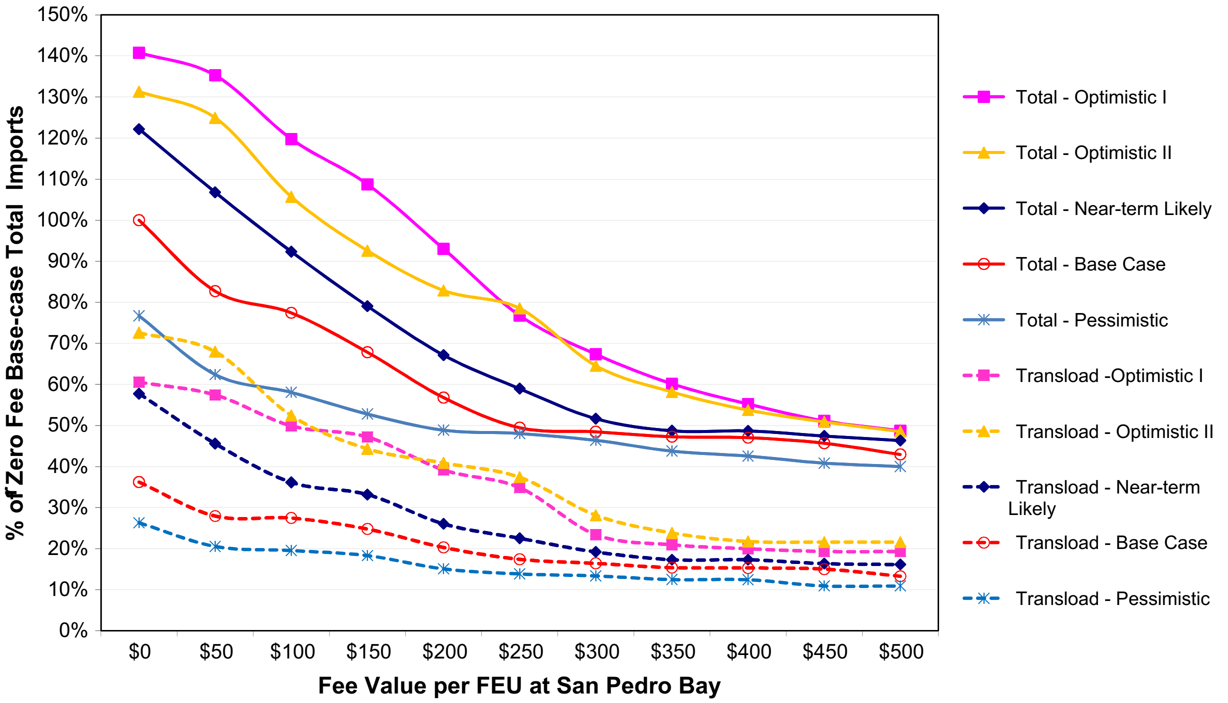

Fig. 10. Long-Run elasticities of imports via the San Pedro Bay ports in future scenarios.

short run, and a drop of almost double that in the long run. If a $200 per FEU fee were introduced in the Pessimistic scenario, total volume routed via San Pedro Bay is predicted to drop almost 30% in the short run and more than 50% in the long run.

It is clear from this analysis that total volume routed via San Pedro Bay is quite elastic to potential container fees. The trans-loaded imports are more inelastic but still decline with fees. The result of potential fees depends heavily on the future scenario of rates charged by the steamship lines and the railroads, and on the share of total imports accounted for by large, nationwide importers. For further details of this analysis, see Leachman (2010).

7. Conclusion

The contributions of this article to the literature are (1) the introduction of a practical queuing model for estimating container flow times through ports as a function of volume, infrastructure, staffing levels and operating schedules, and (2) the integration of queuing models calibrated for port terminals, rail terminals and rail line haul links into a larger model used to estimate total container flow times from port of entry to inland distribution centers. The model has been calibrated on industry data and compared to actual flow times, and its application in policy analysis has been illustrated. While far from perfect, the model shows promise for informing strategy and policy formation efforts.

A number of avenues are available for continued research. First, the scope of application of the models could be extended to embrace all rail channels and all imports to North America, given data describing same. Second, similar models could be developed to model flow times for flows of exports; in particular, queuing analyses could be developed of both containerized and bulk exports. Third, as more detailed and richer data sets are made available, the precision and granularity of the queuing formulas could be improved. For example, separate queuing formulas for port-to-dray dwell time and port-to-on-dock-rail dwell time could be developed from data sets of statistics on dwell times that distinguish the import channel. 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

This research was sponsored by the Southern California Association of Governments. The authors are grateful to many port terminal operators, importers, intermodal marketing companies, third party logistics companies, steamship lines and railroads for providing data for our studies.

References

Alessandri, A., Cervellera, C., Cuneo, M., Gaggero, M., Soncin, G., 2008. Modeling and feedback control for resource allocation and performance analysis in container terminals. IEEE Transportation on Intelligent Transportation Systems 9 (4), 601–614.

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.

Canonaco, P., Legato, P., Mazza, R.M., Musmanno, R., 2008. A queuing network model for the management of berth crane operations. Computer and Operations Research 35 (8), 2432–2446.

Edmond, E.D., Maggs, R.P., 1978. How useful are queue models in port investment decisions for container berths? Journal of the Operational Research Society 29, 741–750.

Garrido, R.A., Allendes, F., 2002. Modeling the internal transport system in a container port. Transportation Research Record: Journal of the Transportation Research Board 1782, 84–91.

Hopp, W., Spearman, M., 2001. Factory Physics, second ed. McGraw-Hill, New York, p. 273.

Jula, P., Leachman, R.C., 2011a. A supply-chain optimization model of the allocation of containerized imports from Asia to the United States. Transportation Research Part E 47, 609–622.

Jula, P., Leachman, R.C., 2011b. 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, 593–608.

Kang, S., Medina, J.C., Ouyang, Y., 2008. Optimal operations of transportation fleet for unloading activities at container ports. Transportation Research Part B 42, 970–984.

Kozan, E., 1997. Comparison of analytical and simulation planning models of seaport container terminals. Transportation Planning and Technology 20 (3), 235–248.

Laganá, D., Legato, P., Pisacane, O., Vocaturo, F., 2006. Solving simulation optimization problems on grid computing systems. Parallel Computing 32 (9), 688–700.

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., submitted for publication. Estimating flow times for containerized imports from Asia to the United States through the western rail network, Working Paper. Transportation Research Part E.

Legato, P., Mazza, R.M., 2001. Berth planning and resources optimisation at a container terminal via discrete event simulation. European Journal of Operational Research 133 (3), 537–547.

Murty, K.G., Liu, J., Wan, Y., Linn, R., 2005. A decision support system for operations in a container terminal. Decision Support Systems 39 (3), 309–332.

Powell, W.B., Carvalho, T.A., 1998. Real-time optimization of containers and flatcars for intermodal operations. Transportation Science 32 (2), 110–126.

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.

Stahlbock, R., Voß, S., 2008. Operations research at container terminals: a literature update. OR Spectrum 30 (1), 1–52.

Steenken, D., Voß, S., Stahlbock, R., 2004. Container terminal operations and operations research—a classification and literature review. OR Spectrum 26, 3–49.

Van Hee, K.M., Wijbrands, R.J., 1988. Decision support system for container terminal planning. European Journal of Operational Research 34 (3), 262–272.

Vis, I.F.A., de Koster, R., 2003. Transshipment of containers at a container terminal: an overview. European Journal of Operational Research 147 (1), 1–16.