Transportation Research Part C 38 (2014) 73–84

Contents lists available at ScienceDirect

Transportation Research Part C

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

An intermodal freight transport system for optimal supply chain logistics

Arnab Bhattacharya [a], Sai Anjani Kumar [a], M.K Tiwari [a,*], S. Talluri [b,*]

[a] Department of Industrial Engineering and Management, Indian Institute of Technology, Kharagpur 721302, India
[b] Department of Supply Chain Management, Broad School of Business, Michigan State University, 351 N. Business Complex, East Lansing, MI 48824, United States

ARTICLE INFO

Article history:
Received 14 May 2012
Received in revised form 31 October 2013
Accepted 31 October 2013

Keywords:
Spatio-temporal data mining
Intermodal transport
Support vector machines
Mixed integer programming

ABSTRACT

Complexity in transport networks evokes the need for instant response to the changing dynamics and uncertainties in the upstream operations, where multiple modes of transport are often available, but rarely used in conjunction. This paper proposes a model for strategic transport planning involving a network wide intermodal transport system. The system determines the spatio-temporal states of road based freight networks (unimodal) and future traffic flow in definite time intervals. This information is processed to devise efficient scheduling plans by coordinating and connecting existing rail transport schedules to road based freight systems (intermodal). The traffic flow estimation is performed by kernel based support vector mechanisms while mixed integer programming (MIP) is used to optimize schedules for intermodal transport network by considering various costs and additional capacity constraints. The model has been successfully applied to an existing Fast Moving Consumer Goods (FMCG) distribution network in India with encouraging results.

1. Introduction

The growing interest in collaborative logistics is fuelled by increasing pressure on companies to operate more efficiently and enhance the productivity of their supply chains. Transportation and logistics management is fast becoming one of the key components of the entire supply chain valuation for many organizations. Due to increasing globalization in recent decades, especially in emerging economies, the importance of logistics management has been on the rise. Traditionally, shippers and carriers have focused their attention on minimizing their own costs to increase profitability, but more recently focus has shifted towards system wide cost reduction to increase profitability of the entire logistics chain. A key component of a logistics chain is the transportation system network. The costs associated with transportation amount to around one third of the total logistics costs which necessitates effective and cost efficient transport coordination mechanisms for managing complex networks involving shipments from manufacturing plants through intermediate distribution centers to customer retail locations.

In many developing economies, majority of the freight transport is undertaken by road based vehicles while rail and water based services remain either largely unutilized or highly disorganized in their functioning and coordination. Road based freight transport has increased significantly over the past few decades in the distribution channels. This rapid increase has led to massive overuse of the road networks without much improvement in the existing infrastructure resulting in various externalities like traffic congestion, increased energy consumption and negative environmental impact. Road capacities, especially outside urban areas, are still inadequate, and several road segments are in inferior condition in the developing countries. In addition, the port and operational transshipment terminals are few in number with low levels of service due

* Corresponding authors.
E-mail addresses: talluri@msu.edu, mkt09@hotmail.com (S. Talluri).

0968-090X/$ - see front matter
http://dx.doi.org/10.1016/j.trc.2013.10.012

to lack of berths, supporting equipment and maintenance. In order to address these issues, we focus on an efficient organization of intermodal transport system which can alleviate these externalities based on the current status of the distribution channel of the supply chain network. Our approach in this paper is aimed at taking advantage of intermodal transport wherever possible in a network heavily dependent on unimodal transport.

According to Mahoney (1986), "Intermodality" is the movement of freight via two or more dissimilar modes of transportation. Hayuth (1987) defines it as the movement of cargo from shipper to consignee by at least two different modes of transport under a single rate, through-billing, and through liability. In general, research in the area of intermodal transport systems not only assists in developing effective transport networks, but also contributes to reducing negative impact on environment and energy consumption. In developing countries such a system will drastically improve the utilization of transport resources and services, leading to better scheduling and delivery with lower logistics costs and higher levels of efficiency.

In this paper, an integrated intermodal freight transport system is developed for an established FMCG distribution network in India. The system first analyzes the traffic status of the road based freight vehicles in various operational zones (spatial clusters) of the network at different time intervals. The system then computes a congestion index for that spatial cluster which is then utilized for any decision to engage other transport modes in the network, such as rail and water links wherever available. The decision for engaging another mode of transport is based on the results of optimizing the total costs involved in the intermodal strategy. In this model, rail is chosen as the primary alternative to road due to certain advantages that include better connectivity, regulated schedules, diversified distribution channel and faster delivery time within acceptable transport cost limits. A single alternate also reduces the complexity and is a fair representation of the networks in most developing countries. However, our approach can easily be extended to other available transport modes with time-space representations. The intermodal train service routes are determined under specific time slots when the train services are offered.

The rest of the paper is organized as follows. The next section reviews literature relating to similar work conducted in the fields of spatio-temporal mining and intermodal transport optimization. We then describe the general methodology in detail. The following section introduces the modeling approach utilized and discusses the details of problem formulation. The results are presented and discussed next, followed by conclusions and future extensions in this area.

2. Literature review

In this section, related work in spatio-temporal traffic flow prediction and intermodal transport optimization is discussed which sets the stage for the problem addressed in this paper. From a methodology standpoint, majority of the work in these areas focuses on advanced predictive analytics using non-linear, non-parametric regressive models and integer programming for combinatorial optimization.

2.1. Spatio-temporal short-term traffic flow prediction

Since the 1970s, univariate time series models have been widely used for traffic flow prediction, especially Box–Jenkins autoregressive integrated moving average (ARIMA) models (Hamed et al., 1995). Subsequently, ARIMA and exponential smoothing (ES) models, such as Holt's–Winter's approach, have been used for comparison purposes whenever a new forecasting model for short-term traffic is proposed (Park et al., 1998). Over the past decade, Neural Network (NN) models have been extensively used in the field of transportation engineering. In addition to flow, other traffic parameters including speed (Ishak et al., 2003) and occupancy (Zhang, 2000) have been predicted in real-time by NN models. Several other techniques have been applied to predict real-time traffic flow. Some of these include multivariate state space time series (Stathopoulos and Karlaftis, 2003) multivariate non-parametric regression (Clark, 2003; Smith and Demetsky, 1996), dynamic generalized linear models (Lan and Miaou, 1999), Hybrid fuzzy rule based system approach (Dimitriou et al., 2008) and Kalman filtering models (Okutani and Stephanedes, 1984). Lin (2001) proposed a forecasting model based on the Gaussian maximum likelihood (GML) estimation method to perform one step ahead forecasts using 5-min traffic flow data. Lin's methodology used both current and historical data traffic in an integrated manner. Probabilistic Principal Component Analysis has been effectively employed using intra-day trend of traffic flow series. This largely removes the issue of missing data while keeping prediction errors relatively low (Chen et al., 2012). Better results have been observed for advanced genetic algorithm based multilayered optimization strategy for neural networks (Vlahogianni et al., 2005). Historic aggregated data from large databases of neighboring stationary detectors and congestion analysis was used to predict traffic flow instabilities (Treiber and Kesting, 2012). Recently, support vector regression (SVR) is being widely applied to predict traffic parameters such as travel time (Wu et al., 2004). The major advantage of SVR is that it avoids over-fitting and allows for a faster training process than other algorithms for multi-dimensional data.

A number of incident detection algorithms have been developed over the past three decades. One of the most popular algorithms is the California Algorithm (Payne and Tignor, 1978). This algorithm is based on the logical assumption that a traffic incident increases the traffic occupancy at the upstream portions of the incident and significantly decreases the traffic occupancy downstream of the incident. However, majority of these algorithms are not reliable in differentiating between recurrent and non-recurrent congestion events. The Minnesota Algorithm (Chassiakos and Stephanedes, 1993) attempts to minimize false alarms and missed incidents by filtering out the effects of high frequency random fluctuations in traffic flow using averaging occupancy measurements over contiguous short-term intervals.

2.2. Optimization of intermodal transport network

Several contributions exist on train routing and scheduling and on applied service network design for train operations. Gorman (1998) focus on service network design models with schedules for CSX transportation and Santa Fe Railways, respectively. Southworth and Peterson (2000) shed light on GIS based intermodal freight network modeling with cost effective network maintenance and traffic route selection for various modes. Yano and Newman (2001) present a dynamic modeling approach to schedule departures of freight and trains to and from a single terminal. Newman and Yano (2000) proposed a train routing model which includes schedules. However, freight demand is modeled to originate and destined to rail terminals, thus drayage moves are not considered. Gorman (1998) developed a linear mixed integer programming (MIP) model for train scheduling with limited terminal operations. The MIP models determine the optimal scheduling on a space-time representation of a network for two types of trains including the train make-ups and empty wagon repositioning problem. Wong et al. (2008) developed a MIP-model to minimize total transfer time for smooth flow and obtained optimal timetable using a heuristic. Yamada et al. (2009) detailed a bi-level programming to obtain multimodal multiuser flow and a benefit-cost ratio (BCR) is considered as objective function to identify best combination for network effectiveness. The related results are obtained using genetic algorithm and Tabu search methods. Caprara et al. (2011) developed a column generation methodology to solve integer linear programming (ILP) model for obtaining efficient routes considering service level of goods. Majority of the models in the literature obtained either optimal travel route using different modes or obtained schedules for usage of different modes. Traffic congestion along different routes is an important issue, which has not been integrated into these approaches working on intermodal freight. This paper develops an integrated modeling structure by which a travel manager will be able to decide whether to consider a unimodal or intermodal strategy based on cost-time tradeoff for a complex distributions network.

3. Methodology

The methodology has two distinct steps which are integrated and implemented as a part of the overall intermodal transport architecture for the distribution channel. We now discuss each of these steps below.

The first step of the model involves finding the degree of traffic congestion of the road based freight services in the logistics network of our supply chain. The main objective of this part of the study is to perform short-term freeway traffic flow predictions under normal or abnormal events. As stated earlier, in developing economies road based services are the main mode of transport for shipment of goods from the manufacturing or distribution centers to the customer locations. For a network that is heavily dependent on road based services, it becomes critical to assess the traffic congestion status at regular intervals, which becomes critical in sectors where the road infrastructure is inadequate to handle the large flow of goods especially into metropolitan and industrial centers. This vehicle tracking and congestion analysis measures are further used for determining the need for intermodal transport that includes rail or water in case of high congestion in the road network, which results in cost savings and better delivery turnaround. The vehicle data have both space (spatial) and time (temporal) dimensions, which are ideal for dynamic response of the system. The analysis is important in developing a strategic decision making mechanism as part of an optimal logistics chain with defined scheduling and delivery timelines.

Traffic congestion is evaluated based on the spatio-temporal data mining process to detect incidents for definite time stamps (e.g. each day of the week, each hour of the day etc.) based on quantitative metrics such as velocity, volume of occupancy, and identifying outliers in each spatial cluster by comparing the traffic data at a particular time interval with the historical models at the same time slot. The outliers, with congestion index above a threshold value, serve as alarms. They indicate high congestion level at that time interval(s) for the given spatial cluster, which is the district or locality observed at that instant of time. The data is acquired through GIS (Geographical Information Systems) or freeway detectors positioned on the road network at regular intervals or mileposts from different spatial clusters in the network.

The short-term traffic flow analysis is followed by determination of the future predicted pattern of the above short-term traffic movement. Accurate prediction of future movements in short time frames gives the schedule manager important insights regarding the congestion in the network and the ability to make decisions for the future based on the patterns appearing on a daily basis. Work in online prediction (Zeng et al., 2008) for short-term flow has shown that results from SVR are better than competing algorithms such as NN and other non-parametric models. So, in our model, we use SVR for short-term traffic forecasting in each cluster. The prediction gives a detailed network occupancy forecast for the definite timestamps associated with the system. The second step deals with the problem of optimization of an intermodal network consisting of road and rail as the two primary means of transport. The decision to employ rail for the distribution channel is based on the traffic spatio-temporal congestion analysis and traffic congestion forecasts in the road network clusters. High congestion with high degree of movement and occupancy in the forecast evokes the need to shift to rail. The model considers the time scheduling and synchronization of the train network and the terminal operations to minimize the transit times associated. The improvement in operations is evaluated based on intermodal trip parameters and comparing it to long-haul trucking.

Fig. 1 depicts a small cluster of the logistics network of a consumer goods company consisting of four containers with different customer origins and destinations that can share the railway freight services. There are three railway terminals. The total transit time for the intermodal chain will be the sum of the drayage transportation times, loading and unloading times at transshipment terminals, rail connection delays, rail transportation time and transportation delays in the drayage movements. This sum of the operations times has to be competitive and more efficient than direct long haul road based services for the intermodal transport to be employed in favor of road. The focus is to reduce the costs associated with these time delays along with the operation costs. We assume that trains always run with the same number of wagons with an upper bound for maximum freight capacity. Transshipment operations at the terminals are considered. Also, the train terminals can hold inventory for short periods of time. There is a limitation on the number of tracks available at the railway terminals, and hence the number of trains available for operations at any given time. As a result, schedule synchronization becomes very important especially as the decision for switching over to intermodal transport is a completely random event based on the congestion analysis of the road network. There is a tradeoff between high frequency of trains resulting in reduced transit times and high operational and handling costs. A MIP approach is utilized to model the problem to capture this tradeoff between operation and time related costs.

4. Modeling approach

4.1. Traffic data analysis

Spatio-Temporal databases are maintained in the network chain to store current traffic information, color coded traffic maps and information of each individual transport vehicle (of any mode) in the distribution network. GIS and detectors on tracking stations are primarily employed to store these data. The route planning problem will be dependent on traffic information in the path ahead and the predicted congestion for a series of time intervals. A methodology is devised to detect the recurrent and non-recurrent congestions in the road network. Both the spatial and temporal information is used for the traffic analysis (unlike traditional methods which use either spatial or temporal information). One of the first and most popular algorithms is the California algorithm (Payne and Tignor, 1978) which relates traffic occupancy to incident detection using a spatio-temporal dataset and decision tree modeling. Our analysis extends the original methodology while exploring a support vector mechanism for congestion prediction. It involves the following steps:

- *Data preparation and categorization:* The database is scanned to identify noisy data, which are removed. Records of the special incidents (eg. Holidays, Blockades) are labeled and data is categorized for different time intervals (for example, hour, day, week, month, etc.)
- *Model generation:* Temporary traffic models are developed based on the mean value of a variable that represents or acts as a proxy to congestion, such as speed or volume occupancy. In our model speed is used as the metric for congestion analysis. A specific lower limit threshold distance *d_low* is determined. This is compared to the vector distance between historical data and the observed data with distances greater than *d_low* eliminated from the model.
- *Congestion analysis:* The congestion evaluation process for the road traffic data is implemented in this step. Traffic data is updated for distinct timestamps and cleaned in runtime. Congestion detection is performed by calculating the Mahalanobis distance between real time data and the corresponding time slot in the traffic model. If the distance calculated is greater than a specified threshold *d_high*, it signals a certain incident causing congestion in the area of analysis. Mahalanobis distance between vector $\vec{x}$ (real time vehicle data) from a group of values with mean $\vec{\mu}$ (calculated from historical data) is calculated using expression (1) below:

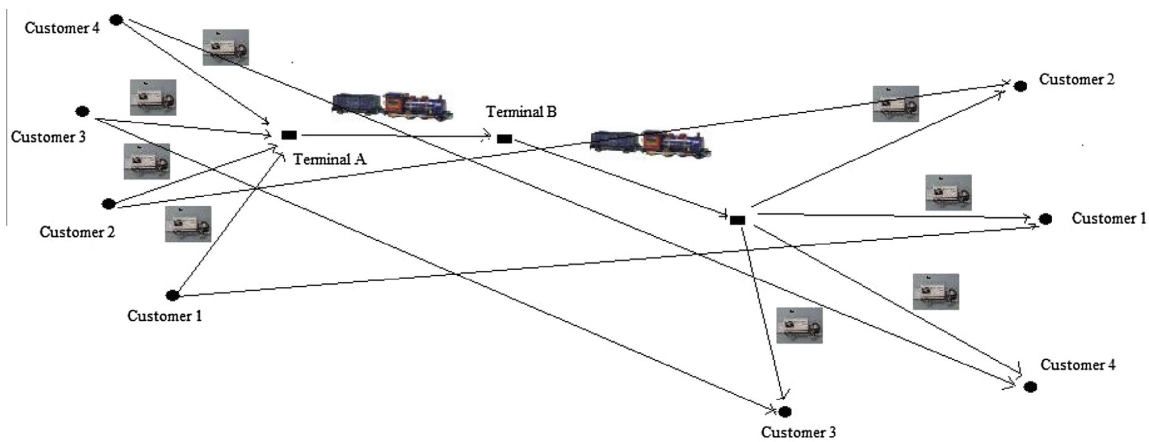

Fig. 1. Intermodal transport chain.

$$d(\vec{X},\vec{\mu}) = \sqrt{(\vec{X}-\vec{\mu})S^{-1}(\vec{X}-\vec{\mu})}$$
$$d(\vec{X},\vec{\mu}) \sim \chi^2_{n-1} \tag{1}$$

It follows the chi-square distribution, where *d_low* and *d_high* can be assigned using probability distribution for a certain level of confidence (we use 95% and 98% levels).

- *Detection level for consecutive timeslots: Det_Lev* (similar to a signal level) is time dependent and based on number of incidents recorded for consecutive time slots (which in turn determines congestion). For the fixed number of time slots considered, the detection level will increase with each new incident (indicator of congestion) in that particular timestamp and decreases when no incident is detected. The detection level function is formulated as given below, which has an incident detection rate $\mu(t)$.

$$Det\_Lev(t) = Det\_Lev(t-1) + \mu(t) \quad \text{when congestion is } det\text{ ected}$$
$$Det\_Lev(t) = De\_Lev(t-1) - \mu(t) \quad \text{when congestion is not } det\text{ ected} \tag{2}$$

Once the signal level increases over a certain level (*Det_Lev_limit)*, short-term forecasting of traffic movement for the particular cluster is initiated.

- *Support vector regression forecasting model*: For the given problem, the set of data points (in the specific cluster) where the signal function exceeds the upper threshold value (*Det_Lev(t)* > *Det_Lev_limit*), is analyzed for congestion prediction. The mathematical analysis is provided below, which is calculated for each spatial cluster (*Clust_i*, $i = 1, \ldots, n$) in the network route.
  A linear regression function is defined in accordance with SVR theory to find the predicted values.
  Total number of points in cluster: m
  Set of points in the network: $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$

$$h(x) = w^T\varphi(x_i) + b \tag{3}$$

We define the above function taking into account: ***F***, a feature space; **w**, a vector in ***F***; $\phi(x_i)$, a mapping function which maps each x_i to a vector in *F*; ε, an insensitive loss function as proposed by Vapnik (1995); it is the maximum error allowed during training of the dataset; *C*, penalty cost for deviation during training process; $\xi_i^+ \xi_i^-$, slack variables corresponding to the size of the excess deviations; w^Tw, regularized term controlling the function capacity; $\sum(\xi_i^+ + \xi_i^-)$, empirical error calculated from the insensitive loss function.Based on SVR literature, the given optimization problem is solved:

$$\textbf{Minimize} \quad \frac{1}{2}w^Tw + C\sum(\xi_i^+ + \xi_i^-) \tag{4}$$

$$\textbf{Subject to} \quad \begin{array}{l} y_i - w^T\Phi(x_i) - b \leqslant \varepsilon + \xi_i^+ \\ w^T\Phi(x_i) + b - y_i \leqslant \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geqslant 0 \end{array} \tag{5}$$

The above problem is solved using the Kuhn Tucker Conditions (KKT) as described in Ma et al. (2003).

The prediction can be based on timestamps such as 5 min, 10 min and 15 min. The forecasting involves the following steps:

Step 1: Use a prediction horizon of one time stamp for a given time series $d(t)$, $t = 1, 2, 3, \ldots$ and prediction origin *Or*, time at which the prediction process starts.
Step 2: Construct a set of training samples *train(Or, B)* from *d(t)*, where $train(Or, B) = \{D(t), y(t), t = B, \ldots Or\}$ and $D(t) = [\{d(t), \ldots d(t-B+1)\}]'$, $y(t) = d(t+1)$ and B is dimension of the training set.
Step 3: Train a predictor *Pr*(*train* (*Or, B), D*) from the training set *train*(*Or*, *B*) and predict $d(O+1)$ using $d(O+1) = Pr(train(Or, B), D(O))'$.
Step 4: Prediction origin is updated and the process is repeated. Results are processed using different kernel functions such as Polynomial kernels.
Step 5: Prediction Performance is calculated by mean absolute percent error (MAPE) and Sensitivity (S).

$$\text{MAPE}(\%) = \frac{1}{n}\sum(|\, y_i - \hat{y}_i \,| / y_i) * 100$$
$$S = (\text{number of correct prediction/total prediction}) * 100 \tag{6}$$

The predicted values are then subjected to the detection analysis for the consecutive timeslots of the prediction horizon by comparing the Mahalanobis distance between the predicted values and the historical traffic model, and followed by finding the corresponding detection level. We compile the results of all the spatial clusters and find the congestion index for each route in the network.

4.2. Intermodal transport optimization

The traffic flow prediction analysis forecasts the congestion in future timeslots in the network. This becomes the basis for devising intermodal strategy in the particular routes where congestion is likely. The intermodal transport module will then be activated to consider the possibility of using other transport modes if available for delivery schedule starting from those particular sectors. The objective of the model is to capture the trade-off between operational costs and the value of time cost incurred in the freight transit times and minimize the sum of the two. The characteristics of the model are given below.

4.2.1. Network structure and modal corridor

Modal corridors are defined to make the scheduling process easily comprehendible. These corridors can be defined for any mode with limited infrastructure in terms of deployment and having structured scheduling pattern. Each rail corridor has an associated fixed cost which includes the crew operating cost, maintenance and depreciation cost or leasing cost of locomotives and wagons, and fuel costs. In our model, train corridors are considered, with predefined time-dependent paths on the physical infrastructure. A corridor allows only one train (or any mode for which the corridor is defined) to run at a time. The network is composed of arcs between the terminals which represent the routing connections between the terminals, and the arcs between the terminals and the customer nodes as the possible drayage movements (as shown in Fig. 2). The network is modeled as a spatio-temporal network which takes into account the scheduling aspect of the service.

4.2.2. Customer dependent freight demand

Freight demand on the service network has a certain level of uncertainty because of lack of detailed knowledge of customer demand. The network also has correlations with demand levels, as the train frequency increases with demand levels. The demand is assumed to arrive from large customer zones with some quantitative estimate of demand potential.

4.2.3. Spatial clusters of customer zones, terminal operations, and drayage movements

Customers are grouped into different customer zones (spatial clusters) associated with one or more railway terminals. Drayage movements require at least one truck for each container. The model assumes that transportation costs are equal for all customers in a customer zone and is commodity indifferent. Transshipment becomes critical in terminal operations. Terminals are also capable of holding inventory, until they are transported to the customer nodes via drayage movements. Variable vehicle transfer/transshipment costs are attached to each departing train corridor arc and departing drayage arc with additional cost of loading and unloading.

4.2.4. Mathematical formulation

In this section, we present the mathematical formulation for the MIP optimization model. It is an arc flow based model. Following notion is used for our problem:

$\wp$: set of all customer zones.
$\hbar_z$: set of customer zone nodes for each customer zone $z \in \wp$.

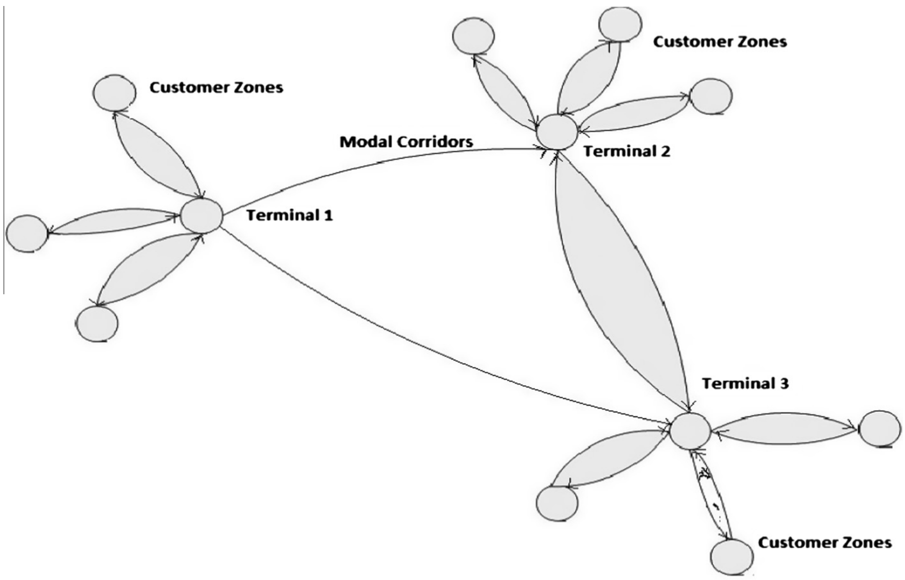

Fig. 2. Intermodal transport network.

$\bigcup_{z\in\wp}\hbar_z$: union of all customer zone nodes for each customer zone z in the network.
S: set of all terminals in the network.
$\hbar_s$: set of terminal nodes for each terminal $s \in S$.
$\bigcup_{s\in\delta}\hbar_s$: union of all terminal nodes in network.
ς: set of commodities that represents demand in the network.
$o(p)$: origin (starting) customer zone node of commodity $p \in \varsigma$.
$d(p)$: destination (ending) customer zone node for commodity $p \in \varsigma$.
t_k^γ: time of availability of commodity p for each customer zone node k.
$\mathfrak{I}$: set of modes representing different trains with different number of containers.
ς_m: $\{(i,j) \mid (i,j) \in \bigcup \hbar_s\}$: set of train corridors (arcs) connecting terminal i to j.
$\hbar_{s,m}^-(i), \hbar_{s,m}^+(i)$: inward and outward neighbors of node i from terminal s.
σ^-: $\{(k,i) \mid k \in \bigcup \hbar_z, i \in \bigcup \hbar_s\}$: set of drayage arcs from customer zone z to terminal node s.
σ^+: $\{(i,k) \mid i \in \bigcup \hbar_s, k \in \bigcup \hbar_z\}$: set of drayage arcs from terminal node s to customer node z.
$\hbar_z^-(i), \hbar_z^+(i)$: set of inward and outward customer zone nodes in zone z from terminal node i.
$\hbar_s^-(z), \hbar_s^+(z)$: set of outward and inward terminal nodes from terminal s for customer zone z.

4.2.4.1. Notation.
$x_{i,p}^\psi$: holding amount of commodity p at customer zone node k.
$x_{i,p}^\mu$: amount of commodity p unloaded at terminal node i.
$x_{i,p}^\nu$: amount of commodity p loaded at terminal node i.
$x_{i,p}^\omega$: inventory of commodity p at terminal node i to customer node k.
$x_{i,h,p}^{\delta+}$: flow of commodity p from customer zone k to terminal node i.
$x_{i,j,m,p}^\lambda$: flow of commodity p on train carrier m running between terminals i, j.
$\mathrm{y}_{\mathrm{i,j,m}}^\lambda = \begin{cases} 1 \\ 0 \end{cases}$ if train carrier m uses terminals (i,j).

4.2.4.2. Decision variables.
$t_{i,s}$: start time associated with terminal at node i.
$t_{i,e}$: end time associated with terminal at node i.
t_k^γ: transit time of inventory arc from terminal node k.
$t_{ki}^{\delta+}(t_{ik}^{\delta-})$: difference between terminal's and customer zone start (end) times shows the time associated with dryage movement.
t_{ij}^λ: transit time of train canal arc(i,j)
c_i^ω: unit cost of inventory arc from terminal node i
$c_{ij}^\lambda = c_i^l + b.t_{ij}^\lambda$: unit cost train canal arc (i,j), here b is a parameter that accounts for cost associated with the time. The value of b indicates the additional cost associated with a unit time increment.
$c_{ki}^{\delta+} = c_{ki}^{d+} + b.t_{ki}^{\delta+}$: unit cost of using drayage arcs from customer zone k to terminal node i.
$c_{ik}^{\delta-} = c_{ik}^{d-} + b.t_{ik}^{\delta-}$: unit cost of using drayage arcs from terminal node i to customer zone k.

The range of the unit costs (given in Table 1) for usage of drayage arcs and the unit train canal costs arise due to the linear relationship of these costs to $t_{ki}^{\delta+}, t_{ik}^{\delta-}$ and t_{ij}^λ respectively.

w_i: inventory capacity limit of terminal node i.
u_i: handling capacity of terminal node i.
v_i: train capacity of terminal node i.
c_i^μ: unit cost for unloading arc in terminal node i.
c_i^ν: unit cost for loading arc in terminal node i.
$f_{i,j,m}^\lambda$: fixed cost of routing train carrier m on train corridor arc(i,j).

The objective function involves the minimization of fixed train corridor costs and container transit costs as shown in (7) below:

$$\min z = \sum_{(i,j)}\sum_{m\in\hbar} f_{ijm}^\lambda y_{ijm}^\lambda + \sum\sum\sum c_{ij}^\lambda x_{ijpm}^\lambda + \sum\sum c_i^\omega x_{ip}^\omega + \sum\sum c_i^\mu x_{ip}^\mu + \sum\sum c_i^\nu x_{ip}^\nu + \sum\sum c_{ki}^{\delta+} x_{kip}^{\delta+} + \sum\sum c_{ik}^{\delta-} x_{ikp}^{\delta-} \quad (7)$$

The following are the constraints that are utilized for flow conservation:

Table 1
Instance attributes.

Nodes						
Total nodes	350					
Terminals	140	Load/unloading cost				
Terminal nodes	210					
	Vehicle transfer cost ($)	Handling cost ($)	Storage cost ($)	Handling capacity (units)	Storage capacity (units)	Train capacity (units)
Ranges	68–88	70–100	15–35	60–110	100	2–4
Arcs						
No. of arcs	787					
Train canal arcs	Mode 1	Mode 2	Mode 3			
	225	109	75			
	203	106	69			
Train transfer arcs						
Train canal cost ($)	Train transfer cost ($)	Train canal cost ($)	Train transfer cost ($)	Train canal cost ($)	Train transfer cost ($)	
9000–14,000	6500	19,800–27,000	130,000	38,600–50,460	27,000	
No. of drayage arcs	450					
Unit transportation cost ($)						
Range	78–150					
Commodities (units)						
No. of commodity	100					
	Low	Medium	High			
Commodity amount range (units)	3–6	5–9	6–12			
Total	399	618	798			
Time horizon	200 h					

Constraint 1: Commodity flow balance in customer zones. The outflow, holding and the demand levels at each customer zone needs to be balanced.

$$x^{\psi}_{n_z-(k),p} + a_{kp} = x^{\psi}_{kp} + \sum_{\bigcup \hbar_s+(k)} x^{\delta+}_{kjp} \quad \forall k \in \cup \hbar_z \tag{8}$$

Constraint 2: The flow balance constraint for commodities arriving at their destination customer zone. The inflow must be such that it satisfies the demand.

$$\sum_{\bigcup \hbar_s-(z)} \sum_{k \in \hbar_z} x^{\delta-}_{ihp} = a_{zp} \quad \forall z \in \wp \tag{9}$$

Constraint 3: The sum of the flow of commodities on arriving train arcs, arriving drayage arcs, and the loading arcs must be equal to the sum of the flow on departing train arcs, departing drayage arcs, and the unloading arcs:

$$\sum_{m} \sum_{\cup \hbar_{sm}+(i)} x^{\lambda}_{hip} + \sum_{k \in \bigcup \hbar_z+(i)} x^{\delta+}_{kip} + x^{\mu}_{ip} = \sum_{m} \sum_{\cup \hbar_{sm}-(i)} x^{\lambda}_{ijpm} + \sum_{l \in \bigcup \hbar_z-(i)} x^{\delta-}_{ilp} + x^{v}_{ip} \tag{10}$$

Constraint 4: For each terminal node i, the sum of the inventory flow from the predecessor (inventory level before the start of the time period) and what is unloaded must be equal to the sum of what is loaded plus what is left in inventory:

$$x^{\omega}_{n_s+(i)ip} + x^{\mu}_{ip} = x^{\omega}_{ip} + x^{v}_{ip} \tag{11}$$

Constraint 5: Inventory level must be lower than the terminal node storage capacity:

$$\sum_{h \in \hbar_{sm}+(i)} y^{\lambda}_{him} = \sum_{j \in \hbar_{sm}-(i)} y^{\lambda}_{ijm} \tag{12}$$

Constraint 6: The maximum number of commodities that can flow on a corridor is constrained by the capacity of the train arc's module.

$$\sum_{p} x^{\lambda}_{ijpm} = q_m y^{\lambda}_{ijm} \tag{13}$$

where $x^{\Psi}_{i,p}, x^{\mu}_{i,p}, x^{v}_{i,p}, x^{\omega}_{i,p}, x^{\delta+}_{i,h,p}, x^{\lambda}_{i,j,m,p}, y^{\lambda}_{i,j,m} \geq 0$.

5. Computational results and analysis

A number of computational experiments were conducted in Matlab R2010b to test the model. The behavior of the network is examined by simulating different scenarios with various time and commodity amounts. The simulation was performed in Simulink 7.6 corresponding to Matlab R2010b.

For the simulation, initially a test network is generated for analysis of the traffic flow and intermodal optimization process consisting of 30 rail terminals and 20 customer zones. The rail terminals are part of a rail infrastructure which connects the customer zones. There are three types of train carriers with capacities of 50, 100, and 200, respectively. There are a total of 140 customer zone nodes and 210 terminal nodes. The number of available train corridors for each of the three modes is 229, 105 and 75. In addition to that, 203,106, and 69 train canal arcs have been added to each of the three types of rail carriers. The sum of all the variables above amounts to 804 binary decision variables and 450 drayage arcs are added between the customer and terminal nodes. The model behavior is examined under varying commodity amount values grouped under Low, Medium and High. The planning horizon has been outlined for 200 h. The terminal nodes representing a terminal are assumed to have the same capacity of trains, transfer operations and inventory. Table 1 outlines the list of all the parameters used in our study which includes ranges for the various costs, capacity arguments and the commodity amount we used for this network.

We use the vehicle GIS data of an established FMCG network stretching across the Northern and Western parts of India. The database used covers the vehicle movements (part of the transport logistics wing) in the western corridor connecting Delhi to Jaipur, Ahmedabad and Mumbai. The total distance of the main route connecting Delhi to Mumbai is around 1600 km and the data is collected in different timestamps of 5, 15, 30 and 60 min for a period of 12 months. The entire geography was divided into distinct district/zone identification (id's) that served to identify the location and the route of the vehicle. The results are given below.

Fig. 3 depicts the movement of the transport vehicles in different timestamps of the day observed for a week. The different colors represent the vehicles in the corresponding districts (district id's) observed in a particular segment of the network between Jaipur and Mumbai (on average the busiest route). The concentration of the vehicles indicates the level of congestion on the route and also the region where the concentration is the highest. A month wide observation also shows the concentration of vehicles (starting from a particular origin) in different routes (identified by district id's). It helps to identify the routes and the spatial locations in the network where there is congestion (the detection alarm is implemented in each route of the network) based on the timestamp under observation. Fig. 4 shows the forecast of the traffic condition in each district based on the support vector prediction and the distribution of vehicles expected in each district which is necessary for the decision to shift to another mode of transport in the network. Districts where the congestion is expected to be high can shift towards rail transport wherever available in the network, within the predefined schedule of the rail service. The decision to shift to intermodal transport lies with the cost and time efficiency that such a collaboration will present to the manager and the organization in general, especially for the FMCG and heavy goods transport companies.

Tables 1–4 provide the SVR parameters and prediction results used for the network. The network involved distribution of the 100 most important commodities in each of the 140 customer zones as their origin and one of the 20 customer zones as their destinations. Given the problem size, it is evident that it is not trivially solved. MATLAB 2010b MIP solver is used for the given optimization problem. Similarly based on the capacity constraints there is a range for storage and train capacity while commodity range values reflect the degree of fragility of the product (high value reflects high fragility) which will attribute

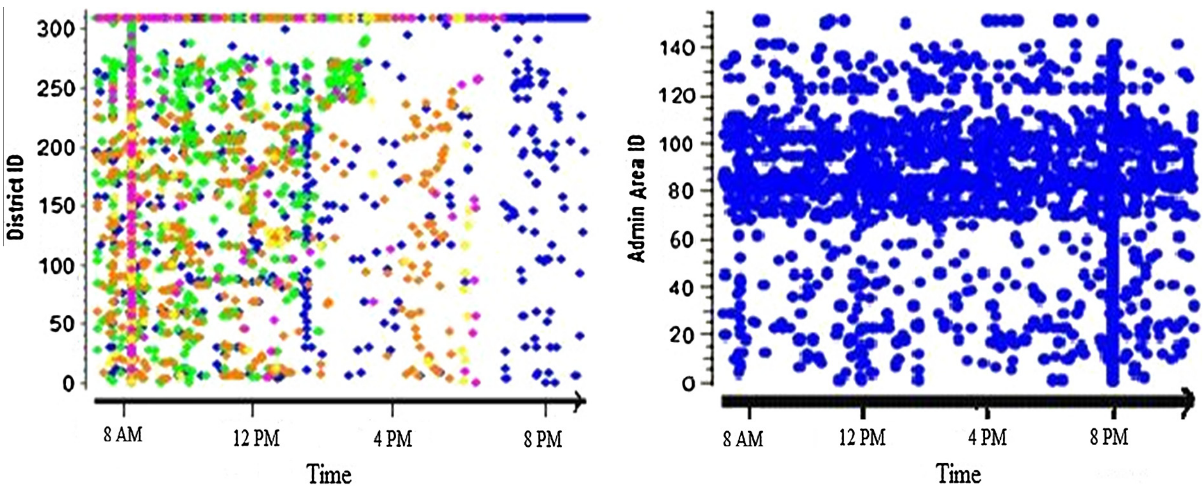

Fig. 3. Time stamp analysis of the traffic movement. Greater the concentration of points more is the congestion for that particular timestamp and location. The different colors represent the vehicles in the corresponding district (locality) (district id's).

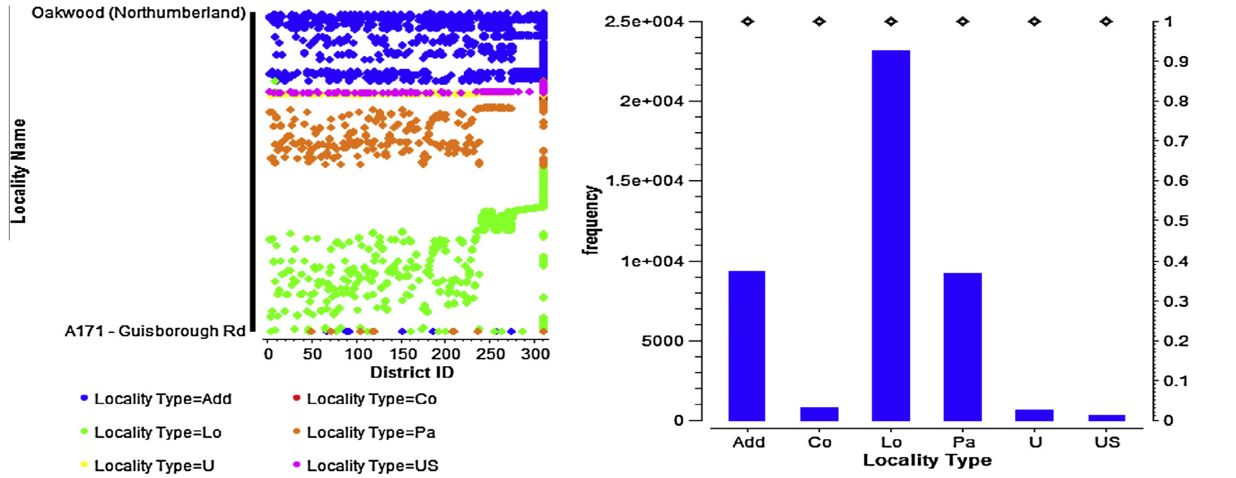

Fig. 4. Locality (spatial coordinates) wise distribution of vehicles in the network – different localities are indicated by different color codes. A histogram is shown with locality type and its associated frequency.

Table 2
SVR Parameter Ranges for regression.

Parameters	Start range	End range
C	0	1,000,000
Epsilon	0	1
d(Polynomial kernel parameter)	0	5
Gamma (Polynomial kernel parameter)	0	5
r(Polynomial kernel parameter)	0	5
nu (Radial basis function parameter)	−5	5

Table 3
SVR values for timestamp 8AM -8PM in region 1.

MAPE(L)	DS(L)	Epsilon	*d*	Gamma	*r*	Win size
1.4089	58.54	0.0213	2	1.247	4.5355	10
1.488	57.14	0.0106	3	2.504	3.9518	20
1.7269	56.67	0.0184	3	2.2115	4.4671	30
1.7327	45.04	0.0149	1	0.9338	0.0541	40
2.8427	51.25	0.0119	3	0.5706	2.4815	50
2.9433	40.58	0.0099	2	2.8285	4.3628	60
5.8682	40.57	0.0075	4	3.1718	3.1718	70

to different handling costs associated with each. The simulation was performed in Simulink, while MATLAB optimization toolbox was utilized to solve the MIP. Table 5 depicts the results of the optimization problem for different commodity levels used for the simulation.

Our analysis points out some very important and useful managerial implications. First, we provide specific details on when intermodal transport is favored to road network, i.e., the time periods in which considerable congestion exists in the road network and suggest optimal intermodal strategy. Our analysis shows that the concentration was very high

Table 4
SVR values for timestamp 8AM-8PM in region 2.

MAPE(L)	DS(L)	Epsilon	*d*	Gamma	*r*	Win size
1.469	55.232	0.0099	2	0.4234	1.9312	10
1.487	52.442	0.0134	2	2.121	3.3123	20
1.526	54.901	0.0124	2	0.2328	2.7086	30
1.624	48.54	0.0914	1	1.775	2.0356	40
3.368	51.123	0.0117	1	1.1306	3.3248	50
3.7558	43.358	0.0091	3	3.9946	2.4384	60
3.804	44.126	0.0155	3	2.6472	4.4231	70

Table 5
Computational scenario results.

Flow	Low	Medium	High
Number of feasible solutions	15	10	4
Best feasible solution value ($)	1,068,860	1,444,330	1,746,030
Lower bound at time limit	617,764	905,356	1,136,908
Lower bound/best feasible solution gap	42.20%	37.31%	34.84.5%
LP solution time (sec.)	250	275	326
LP solution	978,007	1,281,120	1,438,729
LP/Best feasible solution gap (%)	8.5	11.3	17.6

Table 6
Time/cost tradeoff results of the network.

Time value	Low	Medium	High
Operating costs ($)	2,856.030	3,245.810	3,951.580
Number of train services utilized	22	26	35
Train services available	51	43	67
Service capacity index	234	322	466
Total transit time	217.500	179.832	173.700
Transit time decrease (low time → med. time; med. time → high time) (%)	17.32	3.41	3.41
Service utilization (%)	65.27	55.34	41.19
Total handling operations	3560	3122	3345

between 8AM to 12PM in most of the district id's. The congestion detection level was high in this period for all range of timestamps considered, and the support vector forecast also supports the trend for high congestion in the intervals considered. This necessitates the use of intermodal transport in this period while it has to satisfy the overall cost and time limits.

The computation results for the intermodal optimization show that it is difficult problem to solve. Within a simulation time limit of 5 h, we identified feasible solutions for the different scenarios within acceptable limits from the best feasible solutions. The gap between the corresponding linear program (LP) and best solution is 8.5%, 11.3% and 17.6% respectively for different cargo flow scenarios. The overall analysis shows that the operational costs increase with the number of commodities, the fragility quotient of the commodities involved, and with transshipment time.

Table 6 provides the results for the various factors of our optimization study. The higher the value of time, higher is the operational cost and the number of feasible solutions decrease. This implies that when the time value increases, the system tries to incorporate more services in a specific time interval for faster transit times, which results in higher operational and handling costs, which is the classic tradeoff between time and cost. Higher time values ensures shorter transit times and more capacity handling. More train canals are thus chosen when there is higher capacity handling, leading to faster services at higher prices which might be acceptable depending on the company's strategy and urgency of transporting the goods. The analysis further shows that high capacity train canals are chosen for high to medium time values, thus effectively increasing available capacity and leading to corresponding decrease in service utilization. Higher number of services implies fewer transfers and shorter inventory holding time during transshipment. Thus, a transport manager of any logistics department can effectively decide on using the intermodal transport strategy based on the congestion analysis and cost-time tradeoff involving an alternate mode (in our case railways).

The model developed here may very well be adapted to other networks with shorter route distances. Shorter route distances would mean lesser number of district-id's, where traffic congestion would need to be monitored. It is prudent to assume that for short distances the inter-modal transport system should be invoked only if traffic congestion levels are unbearable. The reason being that the dramatic increase in cost that would arise if an inter-modal network was used for this scenario. However, this can easily be considered in the model development by defining the *Det_Lev_limit* (Section 4.1) appropriately, in accordance with the given conditions.

6. Conclusions and extensions

In this paper a basic architecture for traffic flow analysis and decision support for intermodal transport mechanism is presented. Mathematical models for predicting future traffic congestion in roads and intermodal transport cost optimization is formulated and analyzed. Insightful managerial implications are presented based on the simulation results of the intermodal transport strategy. The model application received favorable support from the FMCG company executives. The model is implemented in a generic fashion which can be customized according to the area of application and the need of the organization using it. Future research should investigate the use of multivariate time series models that incorporate spatial and temporal correlations among adjacent Vehicle Detection Stations (VDS) to improve prediction accuracy, especially when multi-step look-ahead forecasts are desired. In addition, future work may evaluate the performance of SVR for various look-back intervals, forecasting horizons, and data resolutions. For the intermodal strategy, heuristic methods can prove to be a quick and efficient way of finding feasible solutions. We believe the next step is to research how to handle the new set of constraints for network design models before proceeding to applying such models to real instances.

References

Caprara, A., Malaguti, E., Toth, P., 2011. A freight service design problem for a railway corridor. Transportation Science 45 (2), 147–162.
Chassiakos, A.P., Stephanedes, Y.J., 1993. Smoothing algorithms for incident detection. Transportation Research Record 1394, 8–16.
Chen, C., Wang, Y., Li, L., Hu, J., Zhang, Z., 2012. The retrieval of intra-day trend and its influence on traffic prediction. Transportation Research Part C 22, 103–118.
Clark, S., 2003. Traffic prediction using multivariate nonparametric regression. Journal of Transportation Engineering 129 (2), 161–168.
Dimitriou, L., Tsekeris, T., Stathopoulos, A., 2008. Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. Transportation Research Part C 16, 554–573.
Gorman, M.F., 1998. An application of genetic and Tabu searches to the freight railroad operating plan problem. Annals of Operations Research 78, 51–69.
Hamed, M.M., Al-Masaeid, H.R., Said, Z.M., 1995. Short-term prediction of traffic volume in urban arterials. Journal of Transportation Engineering 121 (3), 249–254.
Hayuth, Y., 1987. Intermodality: Concept and Practice, Structural Changes in the Ocean Freight Transport Industry. Lloyd's University Press, London.
Ishak, S., Kotha, P., Alecsandru, C., 2003. Optimization of dynamic neural network performance for short-term traffic prediction. Transportation Research Record 1836, 45–56.
Lan, C.J., Miaou, S.P., 1999. Real-time prediction of traffic flows using dynamic generalized linear models. Transportation *Research Record* 1678, 168–178.
Lin, W.H., 2001. A Gaussian maximum likelihood formulation for short-term forecasting of traffic flow. In: Proceedings of the IEEE Intelligent Transportation Systems Conference, pp. 150–155.
Mahoney, J.H., 1986. Intermodal freight transportation. Transportation Quarterly 40 (1), 85–88.
Ma, J., James, T., Simon, P., 2003. Accurate online support vector regression. Neural Computation 15, 2683–2703.
Newman, A.M., Yano, C.A., 2000. Scheduling direct and indirect trains and containers in an intermodal setting. Transportation Science 34 (3), 256–270.
Okutani, I., Stephanedes, Y., 1984. Dynamic prediction of traffic volume through Kalman filtering theory. Transportation Research Part B 18 (1), 1–11.
Park, B., Messer, C.J., Urbanik, T., 1998. Short-term traffic volume forecasting using radial basis function neural network. Transportation Research Record 1651, 39–47.
Payne, H.J., Tignor, S.C., 1978. Freeway incident detection algorithms based on decision tree with states. Transportation Research Record 682, 378–382.
Smith, B.L., Demetsky, M.J., 1996. Multiple-interval freeway traffic flow forecasting. Transportation Research Record 1554, 136–141.
Southworth, F., Peterson, B., 2000. Intermodal and international freight network modeling. Transportation Research Part C 8, 147–166.
Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. Transportation Research Part C 11, 121–135.
Treiber, M., Kesting, A., 2012. Validation of traffic flow models with respect to the spatio temporal evolution of congested traffic patterns. Transportation Research Part C 21, 31–41.
Vapnik, V.N., 1995. The nature of statistical learning theory. Springer, New York.
Vlahogianni, E., Karlaftis, G., Golias, J., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transportation Research Part C 13, 211–234.
Wong, R.C.W., Yuen, T.W.Y., Fung, K.W., Leung, J.M.Y., 2008. Optimizing timetable synchronization for rail mass transit. Transportation Science 42 (1), 57–69.
Wu, C.H., Ho, J.M., Lee, D.T., 2004. Travel-time prediction with support vector regression. IEEE Transactions on Intelligent Transportation Systems 5 (4), 276–281.
Yamada, T., Russ, B.F., Castro, J., Taniguchi, E., 2009. Designing multimodal freight transport networks: a heuristic approach and applications. Transportation Science 43 (2), 129–143.
Yano, C.A., Newman, A.M., 2001. Scheduling trains and containers with due dates and dynamic arrivals. Transportation Science 35 (2), 181–191.
Zeng, D., Xu, J., Gu, J., Liu, L., 2008. Short term traffic flow prediction based on online learning. SVR Workshop on Power Electronics and Intelligent Transportation System (PEITS '08).
Zhang, H.M., 2000. Recursive prediction of traffic conditions with neural network models. Journal of Transportation Engineering 126 (6), 472–481.