Transportation Research Part E 99 (2017) 54–76

Contents lists available at ScienceDirect

Transportation Research Part E

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

Managing congestion in supply chains via dynamic freight routing: An application in the biomass supply chain

Mohammad Marufuzzaman [a,*], Sandra Duni Ekşioğlu [b]

[a] Department of Industrial and Systems Engineering, Mississippi State University, Starkville, MS 39759, United States
[b] Department of Industrial Engineering, Clemson University, Clemson, SC 29634, United States

ARTICLE INFO

Article history:
Received 10 August 2016
Received in revised form 5 December 2016
Accepted 12 January 2017
Available online 1 February 2017

Keywords:
Biomass supply chain network design
Dynamic freight routing
Facility congestion
Benders decomposition
Rolling horizon heuristic

ABSTRACT

This paper manages congestion in the supply chain via dynamic freight routing and using multi-modal facilities in different time periods of a year. The proposed mixed integer non-linear program (MINLP) model captures the trade-offs that exists between investment, transportation, and congestion management decisions. A linear approximation of the proposed MINLP model is then solved using a hybrid Benders-based rolling horizon algorithm. The performance of the algorithm is tested on a case study that uses data from the Southeast USA biomass supply chain network. Extensive numerical experiments provide managerial insights to manage congestion from the biomass supply chain network.

1. Introduction

The main objective of this paper is to develop models to help manage congestion in supply chains which rely on using multiple transportation modes and multi-modal facilities for freight delivery. The models proposed capture the trade-offs between congestion and facility location; congestion and transportation mode selection; transportation network design and product seasonality.

A number of applications for this model can be found in the agricultural sector. For example, in the USA grains are typically transported via railways. The amount shipped via railways has indeed increased in the last few years. In 2014, the amount of grains transported via rail increased by 26% as compared to 2013; and 25% as compared to the average during 2011–2013 (Energy Information Administration, 2014). Due to the physical capacity of railyards and railway lines, this increase in the amount of grain (and other bulk products) transported via railways caused congestion and, consequently, traffic delays. These delays impact railway companies and their customers in multiple ways. First, in response to customers' complains, railway companies have to submit weekly reports to federal regulators explanting the nature of the delays. Second, in order to reduce traffic delays, customers may seek other transportation service providers. The goal of this research is to show that transportation planning which relies on reducing congestion via dynamic freight rerouting impacts costs, traffic, and safety.

Multi-modal facilities such as railway yards, sea ports, airports have limited capacities. Capacity issues impact the time it takes to process shipments, and consequently, contribute to traffic delays and congestion. Researchers have investigated this problem and proposed strategies to prevent congestion and improve the performance of multi-modal facilities in a variety of

* Corresponding author.
E-mail address: maruf@ise.msstate.edu (M. Marufuzzaman).

http://dx.doi.org/10.1016/j.tre.2017.01.005
1366-5545/

application areas. For example, Pels and Verhoef (2004) and Raffarin (2004)propose a few pricing strategies to control flight congestion in airports. Ebery et al. (2000) and Sasaki and Fukushima (2003)add an additional capacity constraint in their model formulation to mitigate the impact of congestion in airports. To the best of the authors' knowledge, the impact of capacity limitations of multi-modal facilities on traffic congestion along multiple modes of transportation has not been carefully studied. To address this need we propose a congestion management strategy which dynamically allocates freight to different modes of transportation, and dynamically selects multi-modal facilities to use along the delivery routes. This allocation changes from one period to the next during a given time horizon. The facility allocation decisions are affected by transportation costs, product availability due to production seasonality, and transportation-related seasonality. The aim is to minimize the overall system costs, including transportation costs, shipment delay costs, and facility allocation costs.

Fluctuations of traffic flow at multi-modal facilities, due to either demand seasonality or weather conditions, have historically been managed via adjustments of capacity. For example, corn and other grains are delivered by barge from the Midwest to the Gulf of Mexico all year around, other than a few weeks during winter due to the drought in the northern section of the Mississippi river. During this time period corn and grains are delivered via rail or trucks. This practice minimizes delays at multi-modal facilities, and shifts transportation volumes from one to another transportation network. The model proposed in this study helps decision makers evaluate the (short-term and mid-term) impacts that dynamically allocating multi-modal facilities and transportation modes in the supply chain have on the overall system performance.

The model we propose is an extension of the fixed charge network design problem, which is known to be an $\mathcal{NP}$-hard (Magnanti and Wong, 1981). Therefore, solving large instances of our problem is a challenging task. This challenge motivated the development of solution approaches which solve the problem efficiently. The methods proposed are a rolling horizon heuristic, an accelerated Benders decomposition algorithm, and a combinatorial Benders decomposition algorithm. The extensive numerical analysis indicates that, the hybrid combinatorial Benders decomposition based rolling horizon algorithm provides high-quality solutions within a reasonable amount of time when used to solve small-sized problems. For large-sized problems, the stand-alone rolling horizon heuristics and hybrid Benders-based rolling horizon algorithm provide near optimal solutions in a reasonable amount of time. The performance of the algorithms proposed is evaluated using a case study developed based on real-world data from the Southeast USA. This case study is focused on the design of a biomass supply chain. The outcomes of the case study are a number of managerial insights, such as, a plan for dynamic deployment of multi-modal facilities and a plan about the amount of biomass to transport via different modes of transportation. These plans are created using data about biomass feedstock seasonality, facility congestion, and transportation costs. These plans aid decision makers in better managing supply chains.

The paper is organized as follows: Section 2 provides a comprehensive literature review; Section 3 formulates the mathematical model; Section 4 introduces the solution algorithms; Section 5 presents numerical results and draws managerial insights; and finally Section 6 provides conclusions and future research directions.

2. Literature review

The impacts of congestion on hub-and-spoke network design modeling have been studied in the literature. Grove and OKelly (1986) investigate the relationship that exists between hub-and-spoke networks and congestion by simulating the daily operations of a single assignment, hub-and-spoke network. Marianov and Serra (2003) model the hub-and-spoke network as an $M/D/c$ queuing network and solve the model using a Tabu search heuristic. Elhedhli and Hu (2005) introduce a non-linear cost term in the objective function of an uncapacitated hub location design problem in order to quantify the impacts of congestion on supply chain costs. In Elhedhli and Wu (2010), the authors extend their previous study to a capacitated hub-and-spoke network design problem. In both papers, the authors linearize the non-linear cost function by approximating it via a set of tangent hyperplanes. They propose a Lagrangian relaxation algorithm which solves the linearized model in a reasonable amount of time. Most recently, Camargo et al. (2009) model the congestion using a convex cost function. They develop a mixed integer nonlinear programming model for a multiple allocation hub-and-spoke network design problem. The authors successfully solve the problem on a network with 81 nodes by using a generalized Benders decomposition algorithm. A few studies (Miranda et al., 2011; Vidyarthi and Jayaswal, 2014) focus on the impact of congestion under demand uncertainty. A brief overview of the hub location problems and solution methodologies can be found in a recent study by SteadieSeifi et al. (2014).

A few researchers have already highlighted the importance of incorporating congestion in the modeling of biomass supply chain since it is a factor that greatly impacts the efficiency of transportation systems. Biomass is a bulk product, thus, modes of transportation such as rail and barge are typically used for long-haul and high volume deliveries. Hess et al. suggest that the use of hub-and-spoke transportation networks will greatly reduce high-volume and long-haul transportation costs of biomass (Hess et al., 2009). Bai et al. (2011) analyze the impacts of congestion on the supply chain performance by introducing a traffic congestion factor in the facility location model they use. This model decides the optimal location of refineries and the flow of biomass and ethanol in the transportation network. The authors propose a Lagrangian relaxation algorithm that is nested within a branch-and-bound framework which finds a high quality feasible solution in a reasonable amount of time. Finally, the authors conduct a sensitivity analysis to show the effects of highway congestion on biorefinery location and supply chain costs. Most recently, Hajibabai and Ouyang (2013) propose an integrated mathematical model that aims to minimize the total cost of facility construction, roadway capacity expansion, including highway links and railway segments,

and biomass/biofuel transportation delay due to congestion. The authors develop a hybrid decomposition algorithm that integrates Genetic algorithm with a Lagrangian relaxation algorithm to solve the model efficiently. A real world case study is developed to show the effects of roadway expansion and traveler time value on supply chain design decisions.

A number of recent works focus on optimizing biomass logistics and transportation costs. Studies by Kumar et al. (2005), Mahmudi and Flynn (2006), and Searcy et al. (2007)evaluate the transportation mode selection on the biomass supply chain performance. These studies mainly focus on supply chain decisions at operational level. Works by Eksioglu et al. (2009), Eksioglu et al. (2010), Huang et al. (2010), An et al. (2011), Xie and Ouyang (2013), Marufuzzaman et al. (2014)develop optimization models that evaluate the impact of integrating strategic and tactical decisions in the performance of biomass supply chain. The aim is to deliver biomass at a more competitive price to the end users. To represent more realistic cases, studies conducted by Kim et al. (2011), Chen and Fan (2012), Gebreslassie et al. (2012), and Marufuzzaman et al. (2014)consider biomass supply, demand, and technology uncertainty in the modeling process.

Most of the existing literature considers truck as the only mode of transportation available to ship biomass from the feedstock suppliers to the end users. These studies were inspired by the practice with corn-based ethanol plants (Brower, 2010). Based on this work, the breakpoint for biomass transportation is 50 miles. That means, if biomass was to be delivered to facilities located further away, then, biofuels would not be cost competitive with fossil fuels. However, the increased demand for biomass supply requires the use of high-volume transportation vehicles such as, rail and barge. Shipments from a number of farms are consolidated at multi-modal facilities, and then delivered to biorefineries via rail or barge. Research conducted by Eksioglu et al. (2010) shows the importance of using high-volume transportation modes in the biomass supply chain network. Xie et al. (2014) extend this work by developing a fully integrated multi-modal transportation system for the cellulosic biofuel supply chain network. Roni et al. (2014) propose a hub-and-spoke supply chain network for long-haul delivery of biomass to coal-fired power plants. The authors also present a Benders decomposition algorithm that solves the largest instance of the problem in a reasonable amount of time. Other studies that analyze the impacts of multi-modal transportation hubs in the supply chain network design decisions are Oosterhuis et al. (2005), Melo et al. (2005), and Hinojosa et al. (2008).

While this work is somewhat similar to the existing literature, there are a few differences which make this work unique. First, our model considers the impact of facility (rather than highways/railway) congestion in the supply chain decisions. Our case study indicates that there are applications in which it is relevant capturing facility congestion. Second, unlike other studies presented in the literature, the model we propose uses dynamic facility allocation and dynamic transportation mode selection in order to alleviate the impacts of congestion in the supply chain performance. In particular, the case study captures the impact of feedstock seasonality and multi-modal facility congestion on supply chain network design. We demonstrate numerically that these dynamic changes to multi-period supply chain network decisions reduce total supply chain costs. Third, we propose a number of solution algorithms which rely on using exact methods (Benders decomposition) and heuristics (rolling horizon algorithm) to provide quality solutions in a reasonable amount of time.

3. Problem description and model formulation

This section provides two mixed-integer, nonlinear programming (MINLP) model formulations for a supply chain management problem. The first model formulation, referred to as [**CM**], considers that multiple modes of transportation are available to deliver a single product from a number of suppliers to a number of facilities via multi-modal facilities. The model captures the impacts of congestion on decisions about the modes of transportation and multi-modal facilities to use. The second model, referred to as [**DCM**], is an extension of [**CM**] which dynamically identifies a set of multi-modal facilities to use or discontinue using, in each time period, in order to minimize the overall system-wide costs. These MINLP models are described in Section 3.1. Section 3.2 provides a linear approximation of model [**DCM**]. These linear mixed-integer programs (MILP) are solved using the algorithms proposed in Section 4.

3.1. A nonlinear model formulation

Consider a supply chain network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of arcs (see Fig. 1). Set $\mathcal{N}$ consists of the set of suppliers $\mathcal{I}$, the set of candidate multi-modal facility locations $\mathcal{J}$, and the set of plant locations $\mathcal{K}$, thus, $\mathcal{N} = \mathcal{I} \cup \mathcal{J} \cup \mathcal{K}$. Each supplier $i \in \mathcal{I}$ produces s_{it} units in period $t \in \mathcal{T}$. Each plant $k \in \mathcal{K}$ demands b_{kt} units in period t. We assume that a substitute product exists in the market, and can as well be used if this supply chain is unable to meet demand. Let U_{kt} denote the corresponding demand shortage. The cost of the substitute product is denoted by π_{kt}. This cost plays the role of a penalty in the objective function and it is paid for every unit of unmet demand. This penalty cost is indeed a threshold to the total unit cost of the final product. That means, if the total unit cost exceeds this threshold, then, satisfying demand via this supply chain network is not economical. Let Ψ_{ljt} denote fixed cost of using the multi-modal facility of capacity level $l \in \mathcal{L}$, located at $j \in \mathcal{J}$, in period $t \in \mathcal{T}$. These costs include salaries and maintenance costs; and do not include facility location costs since we assume that a facility already exist at $j \in \mathcal{J}$.

The set of arcs $\mathcal{A}$ consists of three disjoint subsets, $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$. Set $\mathcal{A}_1$ consists of arcs (i,j), each arc joining a supplier i $(\in \mathcal{I})$ with a multi-modal facility j $(\in \mathcal{J})$. $\mathcal{A}_2$ consists of arcs (i,j), each arc joining a multi-modal facility i $(\in \mathcal{J})$ with a plant j $(\in \mathcal{K})$. $\mathcal{A}_3$ consists of arcs (i,j), each arc joining a supplier i $(\in \mathcal{I})$ with a plant j $(\in \mathcal{K})$ (without using a multi-modal facility).

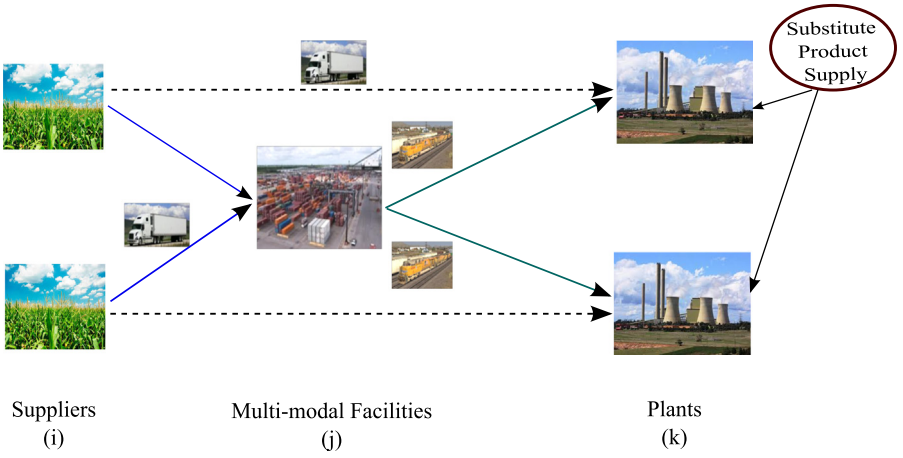


Fig. 1. Supply chain network for co-firing biomass with coal.

The amount of products flowing along arcs in set $\mathcal{A}_1$ is typically low, and the length of the arcs is relatively short as compared to arcs in $\mathcal{A}_2$. For these reasons, the mode of transportation used along arcs in $\mathcal{A}_1$ is truck. Multi-modal facilities serve as shipment consolidation points where small shipments from a number of suppliers are consolidated into large shipments to be delivered to production plants. Thus, the volume of shipments along arcs in $\mathcal{A}_2$ is high. Distances traveled along arcs in $\mathcal{A}_2$ are long. For these reasons, transportation modes such as, rail or barge are typically used to ship products along these arcs. The model also allows products to be shipped directly from a supplier to a plant via arcs $(i,k) \in \mathcal{A}_3$ without traversing a multi-modal facility. Due to the low volume of these shipments, truck transportation is used. A variable (per unit) cost c_{ij} is associated with all arcs $(i,j) \in \mathcal{A}$. When products flow on arcs in $\mathcal{A}_2$, an additional fixed cost per cargo container occurs which we denote by ξ_{jkt}. This represents the additional cost of managing each cargo (rail car) in a yard. Table 1 summarizes the set definitions and input parameters used in the model formulation presented in this section.

The primary decision variables $\mathbb{Y} := \{Y_{ljt}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ determine the size and location of a multi-modal facility, that is,

$$Y_{ljt} = \begin{cases} 1 & \text{if a multi-modal facility of capacity } l \text{ is used at location } j \text{ in time period } t \\ 0 & \text{otherwise.} \end{cases}$$

The second set of decision variables $\mathbb{Z} := \{Z_{jkt}\}_{j\in\mathcal{J},k\in\mathcal{K},t\in\mathcal{T}}$ determines the number of containers flowing between each (facility, plant) pair in period t. The remaining decisions determine the routing of products from suppliers to plants, and identify the size of unmet demand at each plant. The decision variables X_{ijkt} denote the amount transported from supplier i to plant k using multi-modal facility j in period t; decision variables X_{ikt} denote the amount shipped from supplier i directly to plant j in period t (without using a multi-modal facility); and U_{kt} denote the amount of unsatisfied demand at plant k in period t.

Consider the case when the availability of the raw material is seasonal. During the peak season the volume of product flowing through a multi-modal facility is high. This increase of product flow causes congestion at multi-modal facilities.

Table 1
Description of the sets and parameters.

Symbol	Description
Sets	
$\mathcal{I}$	Set of suppliers
$\mathcal{J}$	Set of multi-modal facilities
$\mathcal{K}$	Set of plants
$\mathcal{L}$	Set of plant capacities
$\mathcal{T}$	Set of time periods
Parameters	
ψ_{ljt}	The fixed cost of using a multi-modal facility of capacity $l \in \mathcal{L}$ at location $j \in \mathcal{J}$ in period $t \in \mathcal{T}$
η_{ljt}	The recovery gain associated with discontinuing to use a facility j of capacity l in period t
ξ_{jkt}	The fixed cost of a cargo container for transporting products along arc $(j,k) \in \mathcal{A}_2$ in period t
c_{lkt}	The unit flow cost along arc $(l,k) \in \mathcal{A}$ in period t
c_0	The congestion factor
s_{it}	The amount of product available at site $i \in \mathcal{I}$ in period t
b_{kt}	Product requirements at plant k in period t
v^{cap}	Cargo container capacity
C_{lj}	The maximum capacity of a multi-modal facility of size $l \in \mathcal{L}$ at location j
π_{kt}	Te unit penalty cost of not satisfying the needs of plant k in period t

Congestion impacts the shipment delivery time, and as a consequence, transportation costs. Thus, the model will try to overutilize cheaper facilities while leaving more expensive facilities rarely exploited. To remedy this, Elhedhli and Wu (2010) add a congestion term in the objective function that grows exponentially when the total flow of biomass in the multi-modal facility approaches its capacity (C_{lj}). The authors further show that the closer the total flow to the capacity, the more severe the congestion is. Thus, the congestion rate at a facility $j \in \mathcal{J}$ can be expressed as:

$$\sum_{t\in\mathcal{T}}\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}{\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt}-\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}\right).$$

Based on this equation, an increase on the existing flow at facility $j \in \mathcal{J}$ impacts the average waiting time. This increase has indeed a great impact on the average waiting time when the total flow at facility $j \in \mathcal{J}$ approaches very close to its capacity C_{lj}. Therefore, using this equation we can realistically address the impact of product flow on facility congestion. Note that most of the facilities are designed to utilize for higher capacities. Thus, if the total flow approaches the facility capacity C_{lj}, this should not result in higher system level cost. However, in this study we made an assumption that avoids over utilizing cheaper facilities in the peak biomass production season. The assumption will force the model to use multiple facilities and thus will increase system reliability. Let c_0 be the congestion factor which represent the impact of congestion on transportation costs. Now, the system-wide congestion costs are equal to

$$\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}c_0\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}{\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt}-\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}\right).$$

We can now propose the following MINLP model which minimizes the total costs in the transportation network. We referred to this as formulation [**CM**].

$$[\mathbf{CM}]\quad Minimize\sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt}+\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}c_{ikt}X_{ikt}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}c_{ijkt}X_{ijkt}\right.$$
$$\left.+\sum_{j\in\mathcal{J}}c_0\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}{\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt}-\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}\right)+\sum_{k\in\mathcal{K}}\pi_{kt}U_{kt}\right)$$

Subject to

$$\sum_{k\in\mathcal{K}}X_{ikt}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}X_{ijkt}\leqslant s_{it}\quad\forall i\in\mathcal{I},\ t\in\mathcal{T} \tag{1}$$
$$\sum_{i\in\mathcal{I}}X_{ikt}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}X_{ijkt}+U_{kt}=b_{kt}\quad\forall k\in\mathcal{K},\ t\in\mathcal{T} \tag{2}$$
$$\sum_{i\in\mathcal{I}}X_{ijkt}\leqslant v^{cap}Z_{jkt}\quad\forall(j,k)\in\mathcal{A}_2,\ t\in\mathcal{T} \tag{3}$$
$$\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}\leqslant\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt}\quad\forall j\in\mathcal{J},\ t\in\mathcal{T} \tag{4}$$
$$\sum_{l\in\mathcal{L}}Y_{ljt}\leqslant 1\quad\forall j\in\mathcal{J},\ t\in\mathcal{T} \tag{5}$$
$$Y_{ljt}\in\{0,1\}\quad\forall l\in\mathcal{L},j\in\mathcal{J},\ t\in\mathcal{T} \tag{6}$$
$$Z_{jkt}\in Z^{+}\quad\forall j\in\mathcal{J},k\in\mathcal{K},\ t\in\mathcal{T} \tag{7}$$
$$X_{ijkt},X_{ikt},U_{kt}\geqslant 0\quad\forall i\in\mathcal{I},\ j\in\mathcal{J},\ k\in\mathcal{K},\ t\in\mathcal{T} \tag{8}$$

where $c_{ijk}=c_{ij}+c_{jk}$ denote the unit transportation cost along arc $\{(i,j),(j,k)\}$.

The objective function minimizes the total of transportation costs. More specifically, the first and second term represent the fixed cost of using multi-modal facilities and the fixed cost of transporting cargo containers between two facilities. The third and fourth terms represent the variable transportation costs. The fifth term represents the facility congestion cost. The sixth term is the penalty cost of unmet demand.

Constraints (1) show that the amount delivered from supplier $i \in \mathcal{I}$ in period $t \in \mathcal{T}$ is limited by its capacity. Constraints (2) indicate that the total demand at plant $k \in \mathcal{K}$ will be fulfilled either through the supply chain, or, substitute products available in the market. Constraints (3) indicate that the amount shipped between two facilities is limited by the number of available containers and container capacity. Constraints (4) indicate that the amount shipped to a multi-modal facility is limited by its capacity. Constraints (5) ensure that, at most one multi-modal facility of size $l \in \mathcal{L}$ is operating at a particular location $j \in \mathcal{J}$ in period $t \in \mathcal{T}$. Constraints (6) are the binary constraints, and (7) are the integer constraints. Constraints (8) are the non-negativity constraints.

In order to reduce the impact of congestion on supply chain costs, the routing of the product could dynamically change using different modes of transportation and/or multi-modal facilities. We propose an extension of model [**CM**] to captures the impacts of dynamic product routing on supply chain decisions. The aim is to identify the modes of transportation and the multi-modal facilities to use (or discontinue use) at the beginning of each time period, so that, the overall system costs are

minimized. The model considers operating costs, transportation costs, and congestion costs. Let β_{ljt} and Δ_{ljt} be the closing cost and benefits (e.g., renting the equipment's used by the contactors to other parties) associated with discontinue use of a multi-modal facility $j \in \mathcal{J}$ of size $l \in \mathcal{L}$ in time period $t \in \mathcal{T}$, respectively and typically $\Delta_{ljt} \gg \beta_{ljt}$. Thus, one can simply obtain the recovery gain η_{ljt} due to discontinue use of a multi-modal facility by subtracting β_{ljt} from Δ_{ljt} i.e., $\eta_{ljt} = \Delta_{ljt} - \beta_{ljt}$. The following is the model formulation we propose. We refer to this as model [**DCM**].

$$[\textbf{DCM}] \quad \textit{Minimize} \quad \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}Y_{ljt}(1-Y_{lj,t-1}) - \eta_{ljt}Y_{lj,t-1}(1-Y_{ljt})\right) + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt} + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}c_{ikt}X_{ikt}\right.$$
$$\left. + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}c_{ijkt}X_{ijkt} + \sum_{j\in\mathcal{J}}c_0\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}{\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}\right) + \sum_{k\in\mathcal{K}}\pi_{kt}U_{kt}\right)$$

Subject to: (1)–(8).

The first and second terms of the objective function represent the cost of using and discontinue using of a facilities. The third term represents the fixed cost of transporting cargo containers among multi-modal facilities. The fourth and fifth terms calculate the variable transportation cost along the arcs in the transportation network. The sixth term represents facility congestions costs. The seventh term is the penalty cost for not meeting demand.

3.2. A linear approximation model for [DCM]

Model [**DCM**] contains two nonlinear terms in the objective function. One of the nonlinear terms contains the product of two decision variables $Y_{ljt}Y_{lj,t-1}$. Since both variables are binary, the following technique can be used to transform [**DCM**] into a linear program (Ghaderi, 2012).

Let $F_{ljt} = Y_{lj,t-1}Y_{ljt}$. F_{ljt} is a binary variable which takes the value 1 when both, $Y_{lj,t-1}$ and Y_{ljt} are equal to 1; and takes the value 0 otherwise. Let $\overline{R}_{ljt}$ and $\widehat{R}_{ljt}$ be two decision variables defined as follows:

$$\overline{R}_{ljt} = Y_{ljt}(1-Y_{lj,t-1}) = Y_{ljt} - F_{ljt} \quad \forall l \in \mathcal{L},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{9}$$
$$\widehat{R}_{ljt} = Y_{lj,t-1}(1-Y_{ljt}) = Y_{lj,t-1} - F_{ljt} \quad \forall l \in \mathcal{L},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{10}$$

Constraints (9) and (10) can as well be rewritten as follows:

$$Y_{lj,t-1} + \overline{R}_{ljt} = Y_{ljt} + \widehat{R}_{ljt} \quad \forall l \in \mathcal{L},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{11}$$
$$\overline{R}_{ljt}, \widehat{R}_{ljt} \in \{0,1\} \quad \forall l \in \mathcal{L},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{12}$$

Constraints (11) resemble the flow conservation constraints in network flow models. Therefore, the solution to the linear relaxation of constraints (11) (letting $0 \leqslant \overline{R}_{ljt} \leqslant 1$ and $0 \leqslant \widehat{R}_{ljt} \leqslant 1$) results in an integer solution due to the total unimodularity of node-arc incidence matrix of flow conservation constraints in network flow models (Ahuja et al., 1993).

This partial linear approximation of model [**DCM**] is denoted by [**PDCM**].

$$[\textbf{PDCM}] \quad \textit{Minimize} \quad \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}\overline{R}_{ljt} - \eta_{ljt}\widehat{R}_{ljt}\right) + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt} + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}c_{ikt}X_{ikt} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}c_{ijkt}X_{ijkt}\right.$$
$$\left. + \sum_{j\in\mathcal{J}}c_0\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}{\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}\right) + \sum_{k\in\mathcal{K}}\pi_{kt}U_{kt}\right)$$

Subject to: 1, 2, 3, 4, 5, 6, 7, 8 and (11), (12).

Alternatively, without introducing the additional variables i.e., $\overline{R}_{ljt}$ and $\widehat{R}_{ljt}$, one can still linearize the first two terms of the objective function in model [**DCM**] by adding the following two constraints:

$$1 + F_{ljt} - Y_{ljt} - Y_{lj,t-1} \geqslant 0 \quad t > 1,\ l \in \mathcal{L},\ j \in \mathcal{J} \tag{13}$$
$$Y_{ljt} + Y_{lj,t-1} - 2F_{ljt} \geqslant 0 \quad \forall t > 1,\ l \in \mathcal{L},\ j \in \mathcal{J} \tag{14}$$

Thus, the partial linear approximation of model [**DCM**], denoted by [**PDCM**(**A**)], is given below.

$$[\textbf{PDCM}(\textbf{A})] \quad \textit{Minimize} \quad \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}Y_{ljt} - \eta_{ljt}Y_{lj,t-1} - (\psi_{ljt} - \eta_{ljt})F_{ljt}\right) + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt} + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}c_{ikt}X_{ikt}\right.$$
$$\left. + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}c_{ijkt}X_{ijkt} + \sum_{j\in\mathcal{J}}c_0\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}{\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}\right) + \sum_{k\in\mathcal{K}}\pi_{kt}U_{kt}\right)$$

Subject to: 1, 2, 3, 4, 5, 6, 7, 8 and (13), (14).

Model [**PDCM**] or [**PDCM**(**A**)] is still nonlinear due to the presence of congestion cost function (1). We use the following technique proposed by Elhedhli and Wu (2010) to linearize [**PDCM**]. Let introduce a new decision variable:

$$P_{jt} = \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijkt}}{\sum_{l\in\mathcal{L}} C_{lj} Y_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijkt}} \tag{15}$$

We now rewrite $\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijkt}$ in terms of P_{jt} via simple mathematical manipulations of (15) as follow:

$$\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijkt} = \left(\frac{P_{jt}}{1+P_{jt}}\right)\sum_{l\in\mathcal{L}} C_{lj} Y_{ljt} \quad \forall j\in\mathcal{J},\ t\in\mathcal{T}$$

$$\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijkt} = \sum_{l\in\mathcal{L}} C_{lj}\left(\frac{P_{jt}}{1+P_{jt}}\right) Y_{ljt} \quad \forall j\in\mathcal{J},\ t\in\mathcal{T}$$

Let introduce another continuous variable S_{ljt} which is equal to

$$S_{ljt} = \left(\frac{P_{jt}}{1+P_{jt}}\right) Y_{ljt} \quad \forall l\in\mathcal{L},\ j\in\mathcal{J},\ t\in\mathcal{T} \tag{16}$$

If we sum both sides of this equation $\forall l\in\mathcal{L}$, we get

$$\sum_{l\in\mathcal{L}} S_{ljt} = \left(\frac{P_{jt}}{1+P_{jt}}\right)\sum_{l\in\mathcal{L}} Y_{ljt} \quad \forall j\in\mathcal{J},\ t\in\mathcal{T} \tag{17}$$

By the definition of Y_{ljt} and (5) we have $\sum_{l\in\mathcal{L}} Y_{ljt} \leqslant 1$. For this reason, Eq. (17) becomes:

$$\sum_{l\in\mathcal{L}} S_{ljt} \leqslant \frac{P_{jt}}{1+P_{jt}} \quad \forall j\in\mathcal{J},\ t\in\mathcal{T} \tag{18}$$

Also note that, when $Y_{ljt}=0$, constraints (16) force $S_{ljt}=0$. To ensure that this relationship is maintained we add to the model constraints $0 \leqslant S_{ljt} \leqslant Y_{ljt}; \forall l\in\mathcal{L}, j\in\mathcal{J}, t\in\mathcal{T}$.

Lemma 1. *The function* $S_{ljt}(P_{jt}) = \frac{P_{jt}}{1+P_{jt}}$ *is concave in* $P_{jt}\in(0,\infty)$.

Proof. While differentiating the function $S_{ljt}(P_{jt})$ w.r.t. P_{jt}, we get the first derivative, $\frac{\delta}{\delta P_{jt}}(S_{ljt}) = 1/(1+P_{jt})^2 > 0$, and the second derivative, $\frac{\delta^2}{\delta P_{jt}^2}(S_{ljt}) = -2/(1+P_{jt})^3 < 0$. The first derivative is positive while the second derivative is negative; thus, it proves that the function $\left(\frac{P_{jt}}{1+P_{jt}}\right)$ is concave in P_{jt}. □

Lemma 1 implies that function $\left(\frac{P_{jt}}{1+P_{jt}}\right)$ is concave in P_{jt} and can be approximated via a set of tangent cutting planes as shown below (Elhedhli and Wu, 2010):

$$\frac{P_{jt}}{1+P_{jt}} = Min_{h\in\mathcal{H}}\left[\frac{P_{jt}}{(1+P_{jt}^h)^2} + \left(\frac{P_{jt}^h}{1+P_{jt}^h}\right)^2\right],$$

which implies that

$$\frac{P_{jt}}{1+P_{jt}} \leqslant \frac{P_{jt}}{(1+P_{jt}^h)^2} + \left(\frac{P_{jt}^h}{1+P_{jt}^h}\right)^2 \quad \forall j\in\mathcal{J},\ t\in\mathcal{T},\ h\in\mathcal{H}. \tag{19}$$

In (19), the set $\{P_{jt}^h | j\in\mathcal{J}, t\in\mathcal{T}, h\in\mathcal{H}\}$ consists of all the points which are used to approximate $\left(\frac{P_{jt}}{1+P_{jt}}\right)$.

The binary decision variables $\{Y_{ljt}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ take a finite number of values. Therefore, due to relationship established between Y_{ljt} and P_{jt} via Eq. (16), and for a given set of values of X_{ijkt}, variables $\{P_{jt}\}_{j\in\mathcal{J},t\in\mathcal{T}}$ also take a finite number of values. This implies that the set $\mathcal{H}$ should be finite as well. We now derive Eq. (20) using (18) and (19):

$$\sum_{l\in\mathcal{L}} S_{ljt} \leqslant \frac{P_{jt}}{(1+P_{jt}^h)^2} + \left(\frac{P_{jt}^h}{1+P_{jt}^h}\right)^2 \quad \forall j\in\mathcal{J},\ t\in\mathcal{T},\ h\in\mathcal{H} \tag{20}$$

The linear approximation of [**DCM**] is the following. We refer to this as formulation [**LDCM**].

$$[\textbf{LDCM}]\ \ \textit{Minimize}\ \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}\overline{R}_{ljt} - \eta_{ljt}\widehat{R}_{ljt}\right) + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt} + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} c_{ikt}X_{ikt} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} c_{ijkt}X_{ijkt} + \sum_{j\in\mathcal{J}} c_0 P_{jt} + \sum_{k\in\mathcal{K}}\pi_{kt}U_{kt}\right)$$

Subject to

$$Y_{lj,t-1} + \overline{R}_{ljt} = Y_{ljt} + \widehat{R}_{ljt} \quad \forall l \in \mathcal{L},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{21}$$

$$\sum_{k\in\mathcal{K}} X_{ikt} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} X_{ijkt} \leqslant s_{it} \quad \forall i \in \mathcal{I},\ t \in \mathcal{T} \tag{22}$$

$$\sum_{i\in\mathcal{I}} X_{ikt} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} X_{ijkt} + U_{kt} = b_{kt} \quad \forall k \in \mathcal{K},\ t \in \mathcal{T} \tag{23}$$

$$\sum_{i\in\mathcal{I}} X_{ijkt} \leqslant v^{cap} Z_{jkt} \quad \forall (j,k) \in \mathcal{A}_2,\ t \in \mathcal{T} \tag{24}$$

$$\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijkt} = \sum_{l\in\mathcal{L}} C_{lj} S_{ljt} \quad \forall j \in \mathcal{J},\ t \in \mathcal{T} \tag{25}$$

$$\sum_{l\in\mathcal{L}} S_{ljt} - \frac{P_{jt}}{\left(1+P^h_{jt}\right)^2} - \left(\frac{P^h_{jt}}{1+P^h_{jt}}\right)^2 \leqslant 0 \quad \forall j \in \mathcal{J},\ t \in \mathcal{T},\ h \in \mathcal{H} \tag{26}$$

$$S_{ljt} \leqslant Y_{ljt} \quad \forall l \in \mathcal{L},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{27}$$

$$\sum_{l\in\mathcal{L}} Y_{ljt} \leqslant 1 \quad \forall j \in \mathcal{J},\ t \in \mathcal{T} \tag{28}$$

$$Y_{ljt}, \overline{R}_{ljt}, \widehat{R}_{ljt} \in \{0,1\} \quad \forall l \in \mathcal{L},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{29}$$

$$Z_{jkt} \in Z^+ \quad \forall j \in \mathcal{J},\ k \in \mathcal{K},\ t \in \mathcal{T} \tag{30}$$

$$X_{ijkt}, X_{ikt}, S_{ljt}, P_{jt}, U_{gt} \geqslant 0 \quad \forall i \in \mathcal{I},\ j \in \mathcal{J},\ k \in \mathcal{K},\ l \in \mathcal{L},\ t \in \mathcal{T} \tag{31}$$

4. Solution approach

Model [**LDCM**] is an extension of the fixed charge, uncapacitated facility location problem which is known to be an $\mathcal{NP}$-hard problem (Magnanti and Wong, 1981). Due to the computational challenges one could face when solving large instances of this problem, we propose a few solution techniques which provide near optimal solution for model [**LDCM**] in a reasonable amount of time. The solution approaches proposed are a rolling horizon algorithm, a Benders decomposition algorithm, a combinatorial Benders decomposition algorithm, and an integrated rolling horizon algorithm within the Benders decomposition algorithm.

4.1. Rolling horizon heuristics

The algorithm we propose is outlined in Figure Algorithm 1. This algorithm solves the linear approximation model [**LDCM**], therefore, its role is two fold. First, the algorithm should identify a good approximation for $\{P_{jt}\}_{j\in\mathcal{J},t\in\mathcal{T}}$. Second, for fixed values of P^h_{jt}, the algorithm should solve the corresponding approximation efficiently. The algorithm achieves this via a rolling horizon heuristic.

Algorithm 1. A rolling horizon heuristics

$UB^n \leftarrow +\infty, n \leftarrow 1, t^s_0 = 0, M, \epsilon$, *terminate* $\leftarrow$ **false**

Choose an initial set of points: $\{P^{hn}_{jt}\}_{j\in\mathcal{J},t\in\mathcal{T},h\in\mathcal{H}}$
while (*terminate* = **false**) **do**
  stop $\leftarrow$ **false**, $s \leftarrow 1$
  while (*stop* = **false**) **do**
    let
      $Y^s_{ljt} = Y^{s-1}_{ljt}$ and $Z^s_{jkt} = Z^{s-1}_{jkt}$ for $t < t^s_0$
      $Y^s_{ljt} \in \{0,1\}$and $Z^s_{jkt} \in Z^+$ for $t^s_0 \leqslant t \leqslant t^s_0 + M$
      $0 \leqslant Y^s_{ljt} \leqslant 1$ and $Z^s_{jkt} \in R^+$ for $t > t^s_0 + M$
    Solve the approximate sub-problem [**LDCM**(**s**)] using CPLEX
    if($t^s_0 > |\mathcal{T}|$) **then**
      stop $\leftarrow$ **true**
    end if
    $s \leftarrow s + 1$
  end while
  $UB^n \leftarrow v$[**LDCM**]
  let $P^{h_{new}}_{jt} = \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X^n_{ijkt}}{\sum_{l\in\mathcal{L}} C_{lj} Y^n_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X^n_{ijkt}}$
  if $((UB^{n-1} - UB^n)/UB^n \leqslant \epsilon)$ **then**

(*continued on next page*)

Algorithm 1 (*continued*)

$UB^n \leftarrow +\infty, n \leftarrow 1, t_0^s = 0, M, \epsilon$, *terminate* $\leftarrow$ **false**

terminate $\leftarrow$ **true**
else
$P_{jt}^{h,n+1} = P_{jt}^{h,n} \cup \{P_{jt}^{h_{new}}\}$
end if
$n \leftarrow n + 1$
end while

The algorithm starts by initializing $\{P_{jt}^h\}_{j\in\mathcal{J},t\in\mathcal{T},h\in\mathcal{H}}$. Once these values are fixed, the problem is solved using the rolling horizon algorithm. The quality of values assigned to P_{jt}^h impacts the quality of the solutions we find when solving [**LDCM**]. Therefore, the values of P_{jt}^h are updated iteratively and the problem is resolved. The quality of the approximation is tested by comparing successive solutions obtained. The algorithm terminates when improvements from one iteration to the next are smaller than a threshold value ϵ.

The rolling horizon heuristic decomposes problem [**LDCM**] into a series of subproblems which are solved sequentially (Balasubramanian and Grossmann, 2004;Kostina et al., 2011). These subproblems are designed to be easier to solve as compared to the overall problem. To make the subproblems easy, we relax a few of the binary and integer constraints, while maintaining other variables as integer and binary. We also assign values to a few decision variables based on information received from solving other subproblems. The set of binary and integer variables which are not relaxed is small, and differs from one problem to the next. This set of variables corresponds to a subset of periods within the time horizon. Let $s \in \mathcal{S}$ index the subproblems created. With each subproblem s we associate two parameters, t_0^s and M^s. Parameter t_0^s represents a time period and M^s represents a duration. For each subproblem $s \in \mathcal{S}$ we enforce the following: (i) $Y_{ljt} \in \{0,1\}$and $Z_{jkt} \in Z^+$ for $t_0^s \leqslant t \leqslant t_0^s + M$; (ii) $0 \leqslant Y_{ljt} \leqslant 1$ and $Z_{jkt} \in R^+$ for $t > t_0^s + M$; (iii) $Y_{ljt} = Y_{ljt}^{s-1}$ and $Z_{jkt} = Z_{jkt}^{s-1}$ for $t < t_0^s$. Where, $Y_{ljt}^{s-1}, Z_{jkt}^{s-1}$ are solutions from solving subproblem $s-1$.

The objective function value (v[**LDCM**]) obtained when solving [**LDCM**] via the algorithm proposed here is indeed an upper bound (UB) for [**DCM**]. This is because, Eq. (19) is an outer approximation of $\frac{P_{jt}}{1+P_{jt}}$. Additionally, the rolling horizon algorithm is a greedy procedure which provides us with an integer solution, and therefore, an upper bound for [**DCM**].

4.2. Benders decomposition

We propose a modified version of the Benders decomposition algorithm (Benders, 1962) to solve [**LDCM**]. This algorithm takes advantage of the special structure of the problem. The algorithm separates the problem into two subproblems: an *integer master problem* and a *linear subproblem*. Problem [**LDCM**(**SUB**)] can be further decomposed by time period, into $|\mathcal{T}|$ subproblems. The underlying Benders reformulation for [**LDCM**] is the following:

$$\textit{Minimize} \quad \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}\overline{R}_{ljt} - \eta_{ljt}\widehat{R}_{ljt}\right) + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt}\right) + [\textbf{LDCM}(\textbf{SUB})](X,P,U|\widehat{Y},\widehat{Z})$$

Subject to: (21)–(31).

Problem [**LDCM**(**SUB**)]$(X,P,U|\widehat{Y},\widehat{Z})$ represents the Benders subproblem. Specifically, this problem is:

$$[\textbf{LDCM}(\textbf{SUB})] \quad \textit{Minimize} \quad \sum_{t\in\mathcal{T}}\left(\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}c_{ikt}X_{ikt} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}c_{ijkt}X_{ijkt} + \sum_{j\in\mathcal{J}}c_0P_{jt} + \sum_{k\in\mathcal{K}}\pi_{kt}U_{kt}\right)$$

Subject to

$$\sum_{k\in\mathcal{K}}X_{ikt} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}X_{ijkt} \leqslant s_{it} \quad \forall i \in \mathcal{I},\ t \in \mathcal{T} \tag{32}$$

$$\sum_{i\in\mathcal{I}}X_{ikt} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}X_{ijkt} + U_{kt} = b_{kt} \quad \forall k \in \mathcal{K},\ t \in \mathcal{T} \tag{33}$$

$$\sum_{i\in\mathcal{I}}X_{ijkt} \leqslant v^{cap}\widehat{Z}_{jkt} \quad \forall (j,k) \in \mathcal{A}_2,\ t \in \mathcal{T} \tag{34}$$

$$\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt} = \sum_{l\in\mathcal{L}}C_{lj}S_{ljt} \quad \forall j \in \mathcal{J},\ t \in \mathcal{T} \tag{35}$$

$$\sum_{l\in\mathcal{L}}S_{ljt} - \frac{P_{jt}}{(1+P_{jt}^h)^2} - \left(\frac{P_{jt}^h}{1+P_{jt}^h}\right)^2 \leqslant 0 \quad \forall j \in \mathcal{J},\ t \in \mathcal{T},\ h \in \mathcal{H} \tag{36}$$

$$S_{ljt} \leqslant \widehat{Y}_{ljt} \quad \forall l \in \mathcal{L},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{37}$$

$$X_{ijkt}, X_{ikt}, S_{ljt}, P_{jt}, U_{gt} \geqslant 0 \quad \forall i \in \mathcal{I},\ j \in \mathcal{J},\ k \in \mathcal{K},\ l \in \mathcal{L},\ t \in \mathcal{T} \tag{38}$$

Problem [**LDCM**(**SUB**)] is feasible for any fixed values $\{\widehat{Y}_{ljt}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ and $\{\widehat{Z}_{jkt}\}_{j\in\mathcal{J},k\in\mathcal{K},t\in\mathcal{T}}$. The feasibility of [**LDCM**(**SUB**)] is ensured via constraints (33).

Next, we provide an equivalent formulation of the Benders reformulation of [**LDCM**(**M**)], which is the Benders *master problem*.

$$[\mathbf{LDCM}(\mathbf{M})] \quad \textit{Minimize} \quad \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}\overline{R}_{ljt}-\eta_{ljt}\widehat{R}_{ljt}\right)+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt}\right)+\theta$$

Subject to

$$\theta^n+\sum_{j\in\mathcal{J}}\sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\delta^n_{ljt}(Y_{ljt}-Y^n_{ljt})+\sum_{k\in\mathcal{K}}\gamma^n_{jkt}(Z_{jkt}-Z^n_{jkt})\right)\leqslant\theta \quad \forall n\in N \tag{39}$$

$$Y_{lj,t-1}+\overline{R}_{ljt}=Y_{ljt}+\widehat{R}_{ljt} \quad \forall l\in\mathcal{L},\ j\in\mathcal{J},\ t\in\mathcal{T} \tag{40}$$

$$\sum_{l\in\mathcal{L}}Y_{ljt}\leqslant 1 \quad \forall j\in\mathcal{J},\ t\in\mathcal{T} \tag{41}$$

$$Z_{jkt}\leqslant\sum_{l\in\mathcal{L}}\left\lceil\frac{C_{lj}}{v^{cap}}\right\rceil Y_{ljt} \quad \forall(j,k)\in\mathcal{A}_2,\ t\in\mathcal{T} \tag{42}$$

$$Y_{ljt},\overline{R}_{ljt},\widehat{R}_{ljt}\in\{0,1\} \quad \forall l\in\mathcal{L},\ j\in\mathcal{J},\ t\in\mathcal{T} \tag{43}$$

$$Z_{jkt}\in Z^+ \quad \forall j\in\mathcal{J},\ k\in\mathcal{K},\ t\in\mathcal{T} \tag{44}$$

Constraints (39) are the *optimality cut* constraints. We build these constraints using the dual multipliers $\delta=\{\delta_{ljt}\geqslant 0|l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}\}$ and $\gamma=\{\gamma_{jkt}\geqslant 0|j\in\mathcal{J},k\in\mathcal{K},t\in\mathcal{T}\}$. The dual multipliers δ_{ljt} and γ_{jkt} are the corresponding dual variables associated with the following constraints which are added to [**LDCM**(**SUB**)]:

$$Y_{ljt}=Y^n_{ljt}:\delta_{ljt} \quad \forall l\in\mathcal{L},\ j\in\mathcal{J},\ t\in\mathcal{T} \tag{45}$$

$$Z_{jkt}=Z^n_{jkt}:\gamma_{jkt} \quad \forall j\in\mathcal{J},\ k\in\mathcal{K},\ t\in\mathcal{T} \tag{46}$$

The *master problem* is a relaxation of formulation [**LDCM**], as such, its solutions are used to generate lower bounds (as shown in Proposition 1). Eq. (39) are cuts that are added to the *master problem* and provide an outer approximation of the feasible region. Consequently, by solving the *master problem* iteratively, the quality of the lower bounds improves.

Constraints (42) are valid inequalities which, when added to the *master problem*, impact the convergence rate of the Benders algorithm. These constraints provide an upper bound on the number of containers moving along arcs $(j,k)\in\mathcal{A}_2$. This upper bound is impacted by the storage capacity of the multi-modal facility (C_{lj}) as well as the capacity of the cargo container (v^{cap}).

Proposition 1. *For any given subset of points* $\{P^h_{jt}\}_{\mathcal{H}^q\subset\mathcal{H}}$, **[LDCM(M)]** *provides a lower bound of the optimal objective function value* ($z^n_{MP}(\mathcal{H}^q)$) *of* **[CM]**.

Proof. **[LDCM(M)]**($\mathcal{H}^q$) is a relaxation of problem [**LDCM**]. Thus, the optimal objective function value $z^n_{MP}(\mathcal{H}^q)$ of **[LDCM(M)]** provides the lower bound to the optimal objective value of [**LDCM**]. Since problem [**LDCM**] is an approximation for problem **[CM]**, then $z^n_{MP}(\mathcal{H}^q)$ will also provide a valid lower bound for optimal objective function value of **[CM]**. □

The proposed Benders decomposition has two parts: an *inner* part and an *outer* part. The *inner* part of the Benders algorithm works as follows: it starts by solving the master problem [**LDCM**(**M**)]. The corresponding objective function value (z^n_{MP}) provides a valid lower bound for the overall problem, referred to as *global lower bound* LB_g (as shown in Proposition 1). Next, subproblem [**LDCM**(**SUB**)] is solved. To solve [**LDCM**(**SUB**)], the algorithm sets the values of $\{\widehat{Y}_{ljt}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ and $\{\widehat{Z}_{jkt}\}_{j\in\mathcal{J},k\in\mathcal{K},t\in\mathcal{T}}$ equal to the corresponding solutions obtained from solving [**LDCM**(**M**)]. The solutions from [**LDCM**(**M**)] and [**LDCM**(**SUB**)] are collectively used to calculate a *local upper bound* (UB_l) for [**LDCM**]. We call this upper bound as *local upper bound* since it may not necessary be an upper bound for **[CM]**. The quality of the solution obtained is evaluated by calculating the relative gap between the best local upper bound and global lower bound generated. When this gap is smaller than a threshold value ϵ_i, the inner part of the algorithm terminates.

We use the solution obtained from the inner part of the algorithm to compute an upper bound for the overall problem **[CM]**, referred to as the *global upper bound* (UB_g). Proposition 2 shows that the upper bound obtained from our algorithm is indeed an upper bound for **[CM]**. We now evaluate the quality of the solutions obtained from the global upper and lower bounds. If the gap falls below a threshold value ϵ_o, the outer part of the algorithm terminates. Otherwise, the value of P^h_{jt} is updated (using Eq. (15)) and the process continues.

A pseudo-code of the algorithm is provided in Figure Algorithm 2. In this Figure,

$$z^n_{MAS}=\sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}\overline{R}^n_{ljt}-\eta_{ljt}\widehat{R}^n_{ljt}\right)+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z^n_{jkt}\right).$$

θ^n denotes the solution of subproblem [**LDCM**(**SUB**)] in iteration n of the algorithm. A Schematic representation of the Benders decomposition algorithm is shown in Fig. 2.

Algorithm 2. Benders decomposition

$UB_g^n \leftarrow +\infty, UB_l^n \leftarrow +\infty, LB_g^n \leftarrow -\infty, n \leftarrow 1, r \leftarrow 1, \epsilon_i, \epsilon_o, terminate_i \leftarrow$ **false**, $terminate_o \leftarrow$ **false**

Choose an initial set of points: $\{P_{jt}^h\}_{j\in\mathcal{J},t\in\mathcal{T},h\in\mathcal{H}}$
while ($terminate_o =$ **false**) **do**
  $n \leftarrow 1$
  $terminate_i \leftarrow$ **false**
  while ($terminate_i =$ **false**) **do**
    Solve [**LDCM**(**M**)] to obtain $\{Y_{ljt}^n\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}, \{Z_{ijt}^n\}_{(i,j)\in\mathcal{A}^2,t\in\mathcal{T}}, (z_{MP}^n(\mathcal{H}^q)), z_{MAS}^n$
    if ($z_{MP}^n > LB_g^n$) **then**
      $LB_g^n \leftarrow z_{MP}^n$
    end if
    Solve [**LDCM**(**SUB**)] to obtain θ^n
    if ($z_{MAS}^n + \theta^n < UB^n$) **then**
      $UB_l^n \leftarrow z_{MAS}^n + \theta^n$
    end if
    if ($(UB_l^n - LB_g^n)/UB_l^n \leqslant \epsilon_i$) **then**
      $terminate_i \leftarrow$ **true**
    else
      Calculate δ_{ljt}^n and γ_{jkt}^n using Eqs. (45) and (46)
      Add (39) to [**LDCM**(**M**)]
    end if
    $n \leftarrow n + 1$
  end while
  Calculate UB_g^n using Eq. (47)
  if ($(UB_g^n - LB_g^n)/UB_g^n \leqslant \epsilon_o$) **then**
    $terminate_o \leftarrow$ **true**
  else
    Let $P_{jt}^{h_{new}} = \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijkt}^r}{\sum_{l\in\mathcal{L}} C_{lj} Y_{ljt}^r - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X_{ijkt}^r}$
    $P_{jt}^{h,r+1} = P_{jt}^{h,r} \cup \{P_{jt}^{h_{new}}\}$
  end if
  $r \leftarrow r + 1$
end while

Proposition 2. *For any given subset of points* $\{P_{jt}^h\}_{\mathcal{H}^q\subset\mathcal{H}}$, (47) *provides an upper bound for the optimal objective function value of* **[CM]**.

$$UB_g = \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\psi_{ljt}Y_{ljt} + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt} + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}c_{ikt}X_{ikt} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}c_{ijkt}X_{ijkt} + \sum_{j\in\mathcal{J}}c_0\left(\frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}{\sum_{l\in\mathcal{L}}C_{lj}Y_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}X_{ijkt}}\right) + \sum_{k\in\mathcal{K}}\pi_{kt}U_{kt}\right) \tag{47}$$

Proof. The feasible region of [**LDCM**] is a subset of the feasible region of **[CM]** since [**LDCM**] has all constraints that **[CM]** has and also has a few more (26). So a feasible solution to [**LDCM**] is feasible for **[CM]**. □

4.3. Enhancements of Benders decomposition

This sub-section presents some techniques that were used to improve the computational performance of the basic Benders decomposition algorithm when solving model [**LDCM**].

4.3.1. Multi-cuts

The basic Benders decomposition algorithm adds only one optimality cut to the master problem [**LDCM**(**M**)] in each iteration of the algorithm. Recall that problem [**LDCM**(**SUB**)] can be decomposed into $|\mathcal{T}|$ independent subproblems. Using the dual variables from each subproblem, one can generate and add $|\mathcal{T}|$ optimality cuts to [**LDCM**(**M**)] in each iteration of the algorithm.

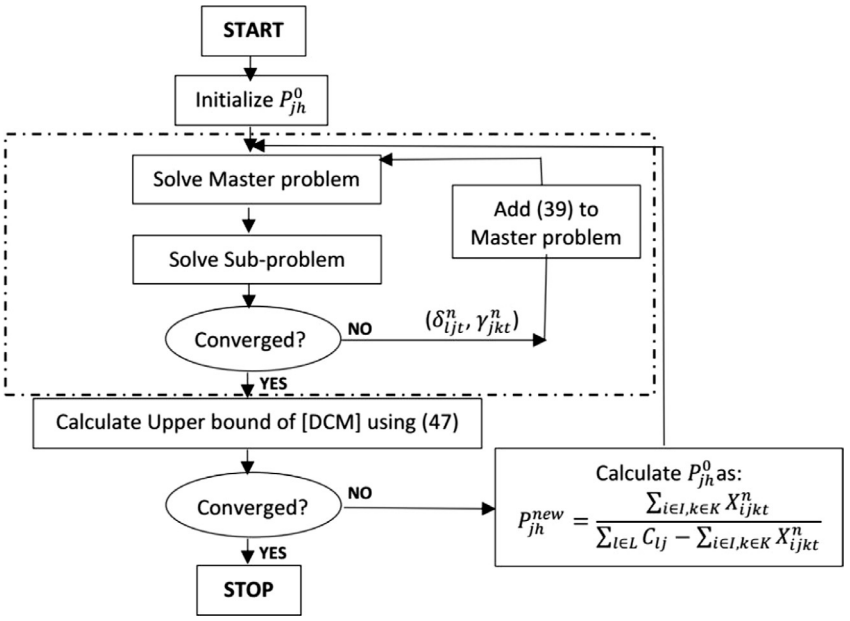


Fig. 2. Schematic representation of the Benders decomposition algorithm.

Adding multiple cuts, rather than the one aggregated cut, to the master problem impacts the quality of the lower bound generated. Consequently, this procedure can result in a faster convergence of the algorithm (Birge and Louveaux, 1997). The revised master problem formulation is shown below:

$$[\mathbf{LDCM}(\mathbf{MT})] \quad \textit{Minimize} \quad \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}\overline{R}_{ljt} - \eta_{ljt}\widehat{R}_{ljt}\right) + \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt} + \theta_t\right)$$

Subject to: (40)–(44)

$$\theta_t \geqslant \theta_t^n + \sum_{j\in\mathcal{J}}\left(\sum_{l\in\mathcal{L}}\delta_{ljt}^n(Y_{ljt} - Y_{ljt}^n) + \sum_{k\in\mathcal{K}}\gamma_{jkt}^n(Z_{jkt} - Z_{jkt}^n)\right) \quad \forall t \in \mathcal{T},\ n \in N \tag{48}$$

4.3.2. Knapsack inequality

In order to speed up the branch-and-bound procedure used by CPLEX solver, we add the following knapsack inequality to the master problem [**LDCM**(**M**)]. A number of research papers in the area (Santoso et al., 2005; Marufuzzaman et al., 2014) point out that the state-of-the-art solvers can derive a variety of valid inequalities from the knapsack inequality. This results in improvements of the convergence rate for the Benders decomposition algorithm.

Let LB^n denote the best known lower bound obtained so far. The following cut is added to the master problem [**LDCM**(**M**)] in iteration $n+1$:

$$LB^n \leqslant \sum_{t\in\mathcal{T}}\left(\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}\overline{R}_{ljt} - \eta_{ljt}\widehat{R}_{ljt}\right) + \sum_{(i,j)\in\mathcal{A}_2}\xi_{ijt}Z_{ijt}\right) + \theta. \tag{49}$$

4.3.3. Integer cuts

Solving the master problem often provides the same solution over subsequent iterations. This behavior of the algorithm impairs its ability to converge, and, its running time. In order to reduce the search space and expedite the running time of the overall algorithm, we add the following inequalities. These inequalities are generated using a local branching technique proposed by Fischetti and Lodi (2003). The inequalities force the master problem to produce a solution different from the solutions generated in the previous iterations.

Let $\mathcal{Y}_1^n = \{(l,j,t)|\widehat{Y}_{ljt}^n = 1, \forall l \in \mathcal{L}^h, j \in \mathcal{J}, t \in \mathcal{T}\}$ be the solutions obtained from solving the master problem in iteration n. Then, the following constraints are added to the master problem in iteration $n+1$:

$$\sum_{(l,j,t)\in\mathcal{Y}_1^n}(1 - Y_{ljt}) + \sum_{(l,j,t)\notin\mathcal{Y}_1^n} Y_{ljt} \geqslant 1 \tag{50}$$

4.3.4. Heuristic improvements

Dynamically Updating the Quality of Solutions Obtained from the Master Problem: In the initial stages of the Benders decomposition algorithm, the master problem typically produces low-quality solutions. This happens until sufficient information from the subproblems is passed to the master problem via constraints (39). Furthermore, the master problem is an integer programming problem, and as such, identifying the optimal solution even for a moderate size network problem, is a challenging task.

We solve the master problem using CPLEX. Experimental results indicate that CPLEX takes time to find high quality solutions. In order to reduce the running time of the algorithm, we initially stop CPLEX when a solution within 5% error gap is found. The error gap is gradually reduced as the algorithm progresses. For instance, the error gap is reduced to 1% when the gap between the upper and lower bound found by the Benders decomposition algorithm falls below 10%.

Setting Branching Priorities: In order to accelerate the running time of the integer master problem, we set proper branching priorities for decision variables Z_{jkt} and Y_{ljt}. Setting branching priorities provides CPLEX an order for branching these variables. Numerical analysis indicates that branching on Z_{jkt} first, followed by Y_{ljt}, reduces the computational time necessary to solve the master problem.

4.4. Combinatorial Benders decomposition algorithm

In order to reduce the complexity of the integer master problem [**LDCM**(**M**)] within the framework of Benders decomposition algorithm, Naoum-Sawaya and Elhedhli (2010) propose a combinatorial Benders decomposition algorithm. This algorithm reduces the complexity of the integer master problem by transferring some of the integer variables to the subproblem. Doing this results in a relatively easier master problem at the expense of solving a harder integer subproblem.

We implement the combinatorial Benders decomposition algorithm by shifting variables Z_{jkt} from the master problem to the subproblem. Therefore, for fixed values of $\{Y_{ljt}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$, the subproblem [**CB**(**SUB**)] can now be written as follows:

$$[\mathbf{CB}(\mathbf{SUB})] \quad \textit{Minimize} \quad \sum_{t\in\mathcal{T}}\left(\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\xi_{jkt}Z_{jkt}+\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}}c_{ikt}X_{ikt}+\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}c_{ijkt}X_{ijkt}+\sum_{j\in\mathcal{J}}c_0P_{jt}+\sum_{k\in\mathcal{K}}\pi_{kt}U_{kt}\right)$$

Subject to: (32) ((33) (), (35)–(37), (3))

$$\sum_{i\in\mathcal{I}}X_{ijkt}\leqslant v^{cap}Z_{jkt}\quad \forall(j,k)\in\mathcal{A}_2,\ t\in\mathcal{T} \tag{51}$$

$$Z_{jkt}\leqslant\sum_{l\in\mathcal{L}}\left\lceil\frac{C_{lj}}{v^{cap}}\right\rceil\widehat{Y}_{ljt}\quad \forall(j,k)\in\mathcal{A}_2,\ t\in\mathcal{T} \tag{52}$$

$$Z_{jkt}\in Z^{+}\quad \forall j\in\mathcal{J},\ k\in\mathcal{K},\ t\in\mathcal{T} \tag{53}$$

Subproblem [**CB**(**SUB**)] is now an integer program. Similar to [**LDCM**(**SUB**)], subproblem [**CB**(**SUB**)] remains feasible (due to constraints (33)) despite of the values of $\widehat{Y}_{ljt}$. The master problem [**CB**(**M**)] can now be written as follows:

$$[\mathbf{CB}(\mathbf{M})] \quad \textit{Minimize} \quad \sum_{t\in\mathcal{T}}\sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{J}}\left(\psi_{ljt}\overline{R}_{ljt}-\eta_{ljt}\widehat{R}_{ljt}\right)+\theta$$

Subject to: (40), (41), (43)

$$M^n\sum_{(l,j,t)\notin\mathcal{Y}_1^n}Y_{ljt}+M^n\sum_{(l,j,t)\in\mathcal{Y}_1^n}(1-Y_{ljt})+\theta\geqslant\theta^n\quad \forall n\in N \tag{54}$$

Constraints (54) are the optimality cut constraints. Here M^n is a large number which turns constraints (54) active only when $Y_{ljt}=Y_{ljt}^n,\forall l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}$.

A good value for M^n can be calculated by setting $M^n=\theta^n-c^T(LB)_{Z,X,P,U}$. Here, θ^n denotes the objective function value of subproblem [**CB**(**SUB**)] in iteration n; $(LB)_{Z,X,P,U}$ represents the lower bounds of the decision variables $Z_{jkt},X_{ikt},P_{jt},U_{kt}$; and c^T represents the cost coefficients associated with these variables (Naoum-Sawaya and Elhedhli, 2010).

Like Benders decomposition algorithm, combinatorial Benders decomposition algorithm also solves problem **[CM]** into two phases: an *inner* part and an *outer* part. In each iteration of the inner part of the algorithm the master problem [**CB**(**M**)] is solved. The corresponding objective function value provides a lower bound (LB_g) for problem **[CM]**. Next, the values of $\{\widehat{Y}_{ljt}\}_{l\in\mathcal{L},j\in\mathcal{J},t\in\mathcal{T}}$ are fixed based on the solutions obtained from the master problem, and passed to the subproblem. The subproblem [**CB**(**SUB**)] is then solved. The solutions from the master problem and subproblem are used to calculate an upper bound for problem [**LDCM**]. We call this upper bound a *local upper bound* UB_l since it is not guaranteed to be an upper bound for **[CM]**. The relative difference between the local upper and global lower bound is calculated. When this gap is smaller than a threshold value ϵ_i, the inner part of the algorithm terminates. We use the solution obtained from the inner part of the algorithm to calculate an upper bound for **[CM]** which we call *global upper bound* (UB_g). Proposition 2 shows that the upper bound obtained from this process guarantees to be an upper bound for the original problem **[CM]**. We now evaluate the qual-

ity of the solutions obtained from the global upper and lower bound. If the gap falls below a threshold value ϵ_o, the outer phase of the algorithm terminates. Otherwise, the value of P^h_{jt} is updated (using Eq. (15)) and the process continues. A pseudo-code of the basic Benders decomposition algorithm is provided in Figure Algorithm 3. In this Figure,

$$z^n_{MAS} = \left(\sum_{t\in\mathcal{T}} \sum_{l\in\mathcal{L}} \sum_{j\in\mathcal{J}} \left(\psi_{ljt} \overline{R}_{ljt} - \eta_{ljt} \widehat{R}_{ljt} \right) \right).$$

Algorithm 3. Combinatorial Benders decomposition

$UB^n_g \leftarrow +\infty, UB^n_l \leftarrow +\infty, LB^n_g \leftarrow -\infty, n \leftarrow 1, r \leftarrow 1, \epsilon_i, \epsilon_o, terminate_i \leftarrow$ **false**, $terminate_o \leftarrow$ **false**

Choose an initial set of points: $\{P^h_{jt}\}_{j\in\mathcal{J},t\in\mathcal{T},h\in\mathcal{H}}$
while ($terminate_o =$ **false**) **do**
  $n \leftarrow 1$
  $terminate_i \leftarrow$ **false**
  while ($terminate_i =$ **false**) **do**
    Solve [**CB**(**M**)] to obtain $\{Y^n_{ljt}\}_{l\in\mathcal{L}j\in\mathcal{J},t\in\mathcal{T}}, z^n_{MP}(\mathcal{H}^q))$, and z^n_{MAS}
    if ($z^n_{MP}(\mathcal{H}^q)) > LB^n_g$) **then**
      $LB^n_g \leftarrow z^n_{MP}(\mathcal{H}^q))$
    end if
    Solve [**CB**(**SUB**)] to obtain $\{Z^n_{ijt}\}_{(i,j)\in\mathcal{A}^2,t\in\mathcal{T}}$ and θ^n
    if ($z^n_{MAS} + \theta^n < UB^n$) **then**
      $UB^n_l \leftarrow z^n_{MAS} + \theta^n$
    end if
    if ($(UB^n_l - LB^n_g)/UB^n_l \leqslant \epsilon_i$) **then**
      $terminate_i \leftarrow$ **true**
    else
      Add (54) to [**CB**(**M**)]
    end if
    $n \leftarrow n + 1$
  end while
  Calculate UB^n_g using Eq. (47)
  if ($(UB^n_g - LB^n_g)/UB^n_g \leqslant \epsilon_o$) **then**
    $terminate_o \leftarrow$ **true**
  else
    Let $P^{h_{new}}_{jt} = \frac{\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X^r_{ijkt}}{\sum_{l\in\mathcal{L}} c_{lj} Y^r_{ljt} - \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} X^r_{ijkt}}$
    $P^{h,r+1}_{jt} = P^{h,r}_{jt} \cup \{P^{h_{new}}_{jt}\}$
  end if
  $r \leftarrow r + 1$
end while

Let $\mathbb{Y} = \{Y_{ljt} | l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}\}$ denote the set of the decision variables in the master problem. This set is bounded since these variables are binary. Furthermore, an optimal solution satisfies the following $\sum_{l\in\mathcal{L}} Y_{ljt} \leqslant 1; \forall j \in \mathcal{J}, t \in \mathcal{T}$. This implies that, the combinatorial Benders decomposition algorithm will converge in a finite number of iterations. However, the convergence of this algorithm is slow mainly due to constraints (54). These constraints are redundant if two consecutive solutions of the master problem are not equal (that is, $Y_{ljt} \neq Y^n_{ljt}, \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$). These constraints are active otherwise. Consequently, constraints (54) do not provide information as to how to improve the Y^n_{ljt} values in most of the iterations of this algorithm. For this reason, we add an integer cut (discussed in Section 4.3.3) and the following knapsack inequalities in each iteration of the master problem [**CB**(**M**)] to accelerate its convergence.

$$LB^n \leqslant \sum_{t\in\mathcal{T}} \sum_{l\in\mathcal{L}} \sum_{j\in\mathcal{J}} \left(\Psi_{ljt} \overline{R}_{ljt} - \eta_{ljt} \widehat{R}_{ljt} \right) + \theta \tag{55}$$

In this equation, LB^n denotes the best known lower bound obtained so far.

4.5. Benders decomposition with rolling horizon heuristics

We propose two algorithms which integrate the rolling horizon approach within the Benders decomposition framework. The first algorithm integrates the rolling horizon algorithm presented in Section 4.1 with the Benders decomposition proposed in Section 4.2. The rolling horizon algorithm here is used to solve the master problem. The second algorithm integrates the rolling horizon with the combinatorial Benders decomposition algorithm proposed in Section 4.4. The rolling horizon algorithm here is used to solve the subproblem.

The motivation behind using a rolling horizon algorithm to solve the master problem within the Benders decomposition framework is the fact that the master problem is an integer program, and as such, it takes much longer to solve as compared to solving $|\mathcal{T}|$ independent subproblems. Therefore, in the first algorithm proposed, the rolling horizon approach is adopted to solve the master problem. This algorithm decomposes the master problem into a series of smaller subproblems which are solved sequentially using a rolling horizon approach.

The motivation behind using a rolling horizon algorithm to solve the subproblems obtained from the combinatorial Benders decomposition algorithm is the fact that these subproblem are indeed dynamic network design problems of large size. For this reason, solving these subproblems takes time, and often, it takes longer as compared to solving the integer master problem. The algorithm solves these dynamic network design problems repeatedly over time as new information becomes available. Here are the details on how this algorithm works. Solving the master problem provides investment decisions, in other words, values for the decision variables Y_{ljt}. These decisions are then fixed, and the algorithm continues by solving the corresponding subproblem. The subproblem is decomposed into a series of smaller subproblems which are solved sequentially using the rolling horizon approach. Each smaller subproblem comprises a few consecutive time periods of the overall planning horizon. When all the subproblems are solved, the rolling horizon approach provides an upper bound of the cost of the overall subproblem. The feasible solution obtained from solving the subproblem can be used to calculate an upper bound for the overall problem. Clearly, this upper bound will be higher as compared to using the optimal solution of the subproblems. However, the procedure results in computational savings.

5. Case study: biomass supply chain

In order to test the performance of the algorithms proposed in this paper we develop a case study. The case study is focused on the biomass supply chain. We consider the scenario when a number of power plants co-fire coal and biomass to produce renewable electricity. Biomass co-firing is a strategy that leads to reduced greenhouse gas (GHG) emissions in coal-fired power plants. Currently, 40 of the 560 coal-fired power plants in the USA are co-firing biomass. To learn more about the impacts of co-firing to GHG emission reductions, and the associated costs, see the recent report by the Idaho and Pacific Northwest National Laboratories (Boardman et al., 2013).

The power plants receive shipments from a number of suppliers. We assume that biomass is pre-processed before shipping. Depending on the distance traveled, some of the suppliers could use truck and rail to deliver the product. Since coal (similar to biomass) is a bulk product, existing coal-fired power plants do typically have access to rail transportation. Therefore, we assume power plants are not investing in building the necessary infrastructure. The main goal is to identify which suppliers to use, what mode of transportation, and which multi-modal hubs so that the system-wide costs are minimized.

This section provides details about the data used to develop the case study. Next, it summarizes the results from the numerical analysis. Finally, an interpretation of the results and a few managerial insights are presented.

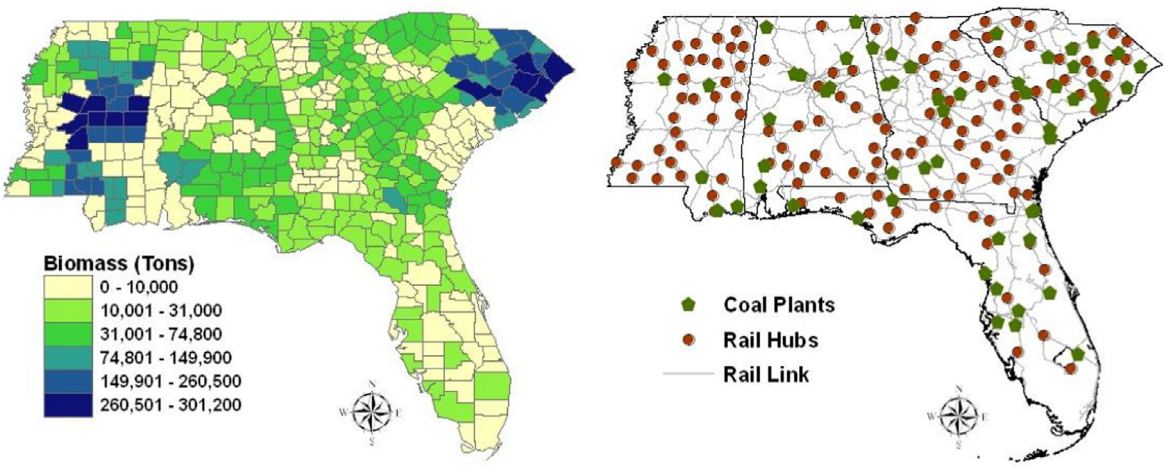
5.1. Data description

5.1.1. Biomass supply

The region under study consists of five states located in the Southeast USA: Mississippi, Alabama, Georgia, Florida, and South Carolina. The two main biomass feedstocks in this region are corn stover and forest residues. The biomass availability data, at the county level, is provided by the Knowledge Discovery Framework (KDF) database of United States Department of Energy (Bioenergy Knowledge Discovery Framework, 2013). The database includes 274 counties within this region. The data was further processed at Idaho National Lab (INL) to calculate the amount of densified biomass available in this region. We focus on densified biomass since it can be delivered in high volume and long distance via modes of transportation such as rail. Fig. 3(a) shows the distribution of densified biomass in the Southeast. Based on the availability of biomass in the region, it is expected that about 14.87 million tons (MT) of densified biomass can be produced annually.

5.1.2. Biomass demand

The data about coal-fired power plant locations and capacities is obtained from the National Energy Technology Laboratory (The National Energy Technology Laboratory, 2005). Within the study region, there are 59 coal-fired power plants with an overall production capacity of about 58,503 MW. Fig. 3(b) presents the distribution of these plants. We assume that power plants will displace on the average 6% of coal with densified biomass to produce renewable electricity. This results in a demand equal to 17.19 million tons per year (MTY) of densified biomass. This amount is estimated based on the name-



(a) Biomass distribution (b) Network representation

Fig. 3. Biomass distribution and facility locations.

plate capacity, efficiency and operating hours of the plant, and the lower heating values of coal and biomass (Karimi et al., 2014).

5.1.3. Investment costs

There is a total of 119 rail ramps in the Southeast USA. Fig. 3(b) presents the locations of these ramps. Each rail ramp could be a potential multi-modal facility where truck shipments are consolidated. The annualized fixed cost per rail car at a rail ramp of capacity of 1.05 MTY is estimated to be $54,949 per year (Mahmudi and Flynn, 2006). This cost represents the fixed cost of using a multi-modal facility (ψ_{ljt}). Using this information, we estimate the fixed costs per MTY, and use this estimate to calculate investment costs for other rail ramp capacities. We consider five different capacities, which are: l = 0.6 MTY, 0.8 MTY, 0.9 MTY, 1.05 MTY, and 1.20 MTY. We assume a lifetime of 30 years, and a discount factor of 10%. We acknowledge that, the actual fixed cost of using a multi-modal facility may vary by location. However, since this analysis is focused in Southeast USA, these differences are small. Thus, the same fixed costs are used independent of the location of the plant within the region.

5.1.4. Transportation costs

This study assumes that trucks are used to transport biomass from feedstock suppliers to multi-modal facilities. Trucks are also an option to transport biomass directly from feedstock suppliers to power plants. The input data necessary to calculate the unit truck transportation cost is obtained from a study by Parker et al. (2008). This data is summarized in Table 2.

The data necessary to calculate the cost of rail transportation is obtained using the regression models developed by Gonzales et al. (2013). These regression equations represent the relationship between the total transportation cost and transportation distance for a single rail car of capacity 100 tons. The fixed shipment cost per rail car equals $2248 ($\xi_{jkt}$). The unit transportation cost per mile traveled by the rail car is estimated $1.12. We assume that a shipment is delivered from its source to its destination using the shortest path. Arc GIS Desktop 10 is used to present the existing railway transportation network in Southeast USA. We use this network to identify the shortest paths between rail ramps. In addition to existing railway lines, this network also includes local, rural, urban roads, and major highways in the Southeast USA.

5.2. Experimental results

The numerical results presented next are obtained by solving the algorithms proposed above. These algorithms are coded in GAMS 24.2.1 (General Algebraic Modeling System, 2013) and executed on a desktop computer with Intel Core i7 3.50 GHz processor and 32.0 GB RAM. The optimization solver used is ILOG CPLEX 12.6.

Table 2
Truck transportation cost components.

Item	Value	Unit
Loading/unloading	5.0	$/wet ton
Time dependent	29.0	$/hr/truckload
Distance dependent	1.20	$/mile/truckload
Truck capacity	25	wet tons/truckload
Average travel speed	40	miles/hour

5.2.1. Analyzing the impacts of dynamic facility location on congestion

To evaluate the impacts of dynamic multi-modal facility locations on mitigating congestion, we conducted two different experiments. First, we solved model [**LDCM**] to determine the impacts of the proposed dynamic facility location model on system's performance. We used the following system performance measures: number of containers transported, amount of biomass shipped, total costs, etc. Next, we resolved model [**LDCM**] adding one more constraint. This constraint forces a multi-modal facility in operation to remain active until the end of the planning horizon. This constraint renders our model static. The constraint we add to model [**LDCM**] is: $Y_{lj,t-1} \leqslant Y_{lj,t}; \forall l \in \mathcal{L}, j \in \mathcal{J}, t \in \mathcal{T}$.

Fig. 4 presents the number of hubs operating each month, the number of containers transported between arcs $(j,k) \in \mathcal{A}_2$, and the amount of biomass shipped through highways. This information is essential to support strategic decisions and to guide the planning of manpower and equipment during the year. In the experiments conducted, we set $t = 1$ to represent the month of July. Corn stover is typically harvested from September ($t = 3$) until November ($t = 5$) in a given calender year. Forest residues are harvested all year around, except during the three months of winter, December to February ($t = 6$–8), due to humid weather. The data suggests that September ($t = 3$) to November ($t = 5$) is the peak biomass production season in a given year. Neither corn stover, nor woody biomass are available during the three months of winter ($t = 6$–8), thus, this is the off-peak production season for biomass. The multi-modal facilities are likely to become congested during the peak biomass production seasons. To support the delivery of biomass to power plants from September to November, additional facilities and containers are used. During this period, the amount of biomass shipped by trucks via highways also increases. These results reinforce the need for adjusting dynamically the short and mid-term supply chain decisions.

Figs. 5 and 6, respectively, are examples of the network design during the peak and off-peak biomass production seasons. The results indicate that, supply chain decisions in the static model are insensitive to changes in biomass production. The

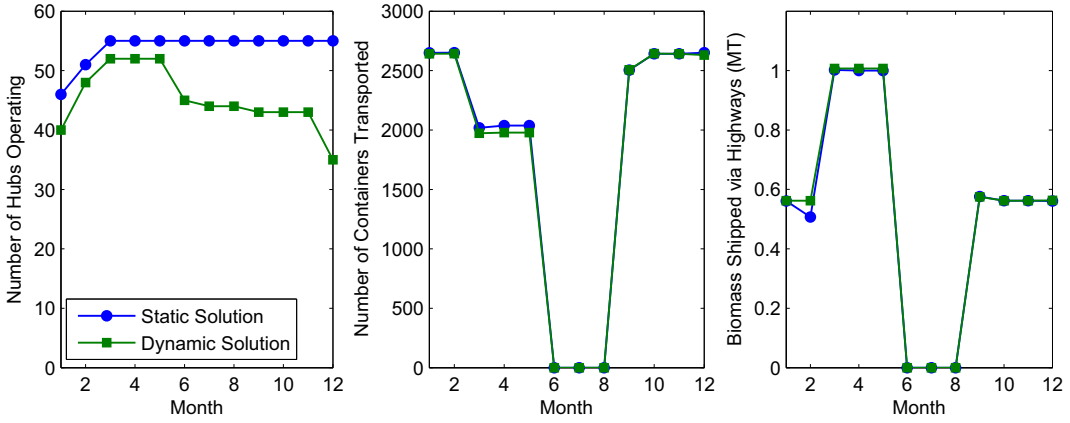
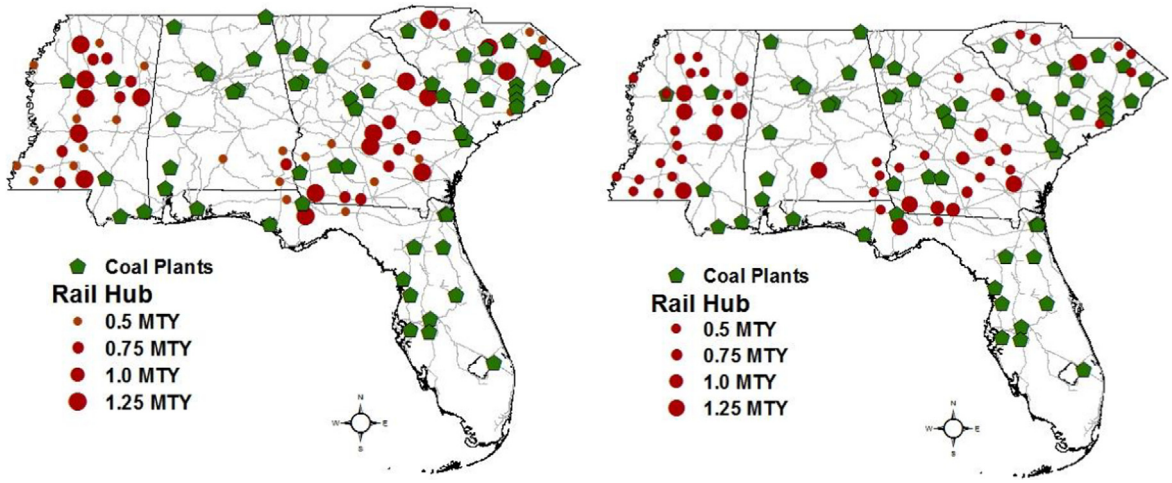
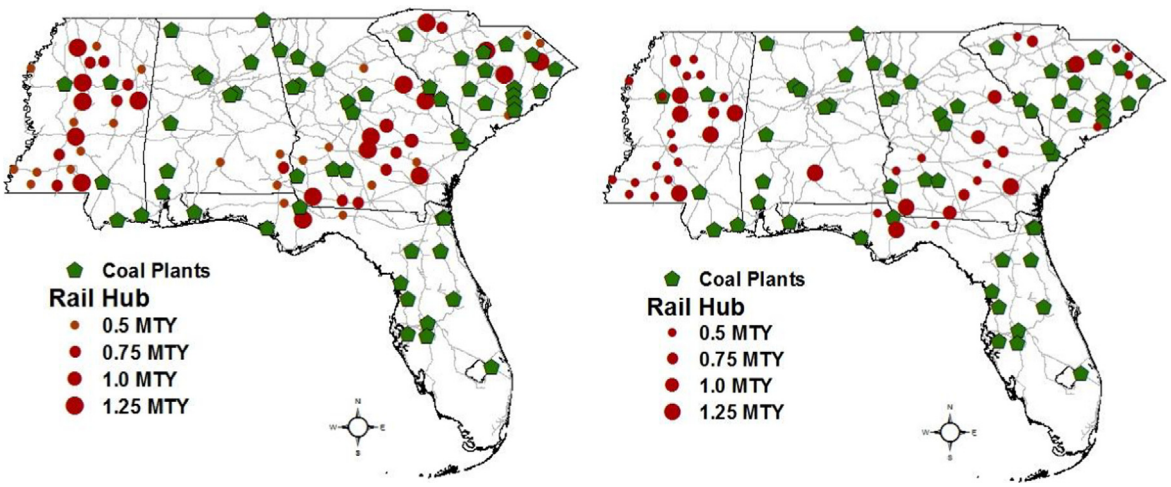


Fig. 4. Impact of dynamicity on system performance.



(a) Static solution (b) Dynamic solution

Fig. 5. Network representation under peak biomass production season ($t = 3$).



(a) Static solution (b) Dynamic solution

Fig. 6. Network representation under low biomass production season ($t = 7$).

dynamic model does reconfigure the supply chain network in order to make the best use of the operating facilities in a given time period. The results from Fig. 4 indicate that both, the static and dynamic models, use about the same number of facilities until period 6. However, the static model does not discontinue use of these facilities during the months of the low supply season. Therefore, the annual facility operating costs are much higher for the static as compared to the dynamic model. The dynamic model does a better job managing congestion at multi-modal facilities by using alternative transportation options. For this reason, the total unit cost of the dynamic model is \$22.65/ton of biomass, and the corresponding total unit cost provided by the static model is \$23.35/ton. The unit total cost from the dynamic model, assuming that facility congestion is not present (i.e., $c_0 = 0$), is \$22.23/ton.

5.2.2. Analyzing the impacts of congestion on system's performance

Fig. 7 shows the impact of facility congestion cost on biomass supply network. The results from the first sub-figure of Fig. 7 indicate that increasing congestion cost negatively impacts the number of multi-modal facilities being used in the supply chain. As the congestion cost decreases, more multi-modal facilities are used. When congestion cost is decreased below a certain threshold level, it does not impact the number of operating facilities. This implies that, when facility congestion costs are negligible, then dynamically allocating facilities has little impact on costs (see Table 3). The dynamic model we propose greatly impacts the supply chain performance when congestion costs are high. In this case, routing biomass via highways becomes a preferred mode of transportation (see the second and third sub-figures in Fig. 7). In summary, the results indicate that accounting for the impacts of multi-modal facility congestion in the supply chain performance results in costs savings.

5.2.3. Analyzing the impacts of biomass supply on system's performance

Fig. 8 summarizes the impact of changes in biomass supply on the supply chain performance. The results indicate that the number of multi-modal facilities used, and the number of containers transported between facilities is affected by fluctua-

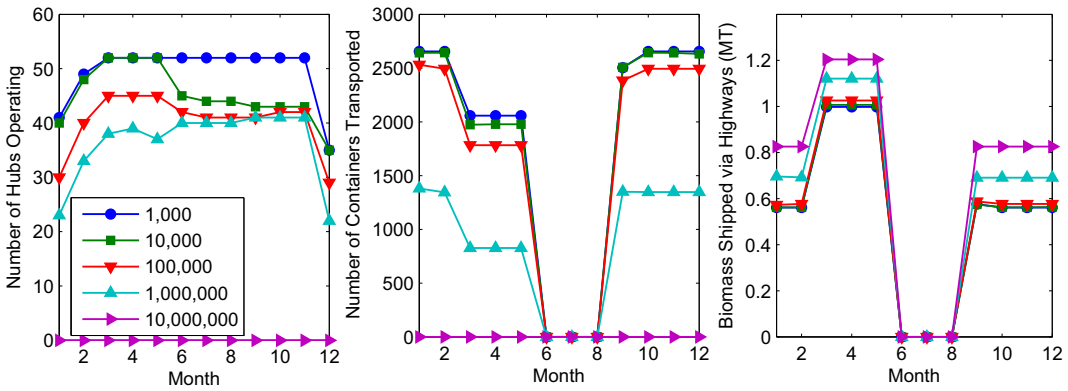


Fig. 7. Impact of congestion cost on system performance.

Table 3
Unit cost under different hub congestion cost.

Hub congestion cost ($/year)	Annual system cost (Million $)	Unit cost ($/ton)
0	1053.71	22.23
1000	1053.74	22.25
10,000	1054.02	22.28
100,000	1057.27	22.65
1,000,000	1071.08	24.26
10,000,000	1079.71	25.27

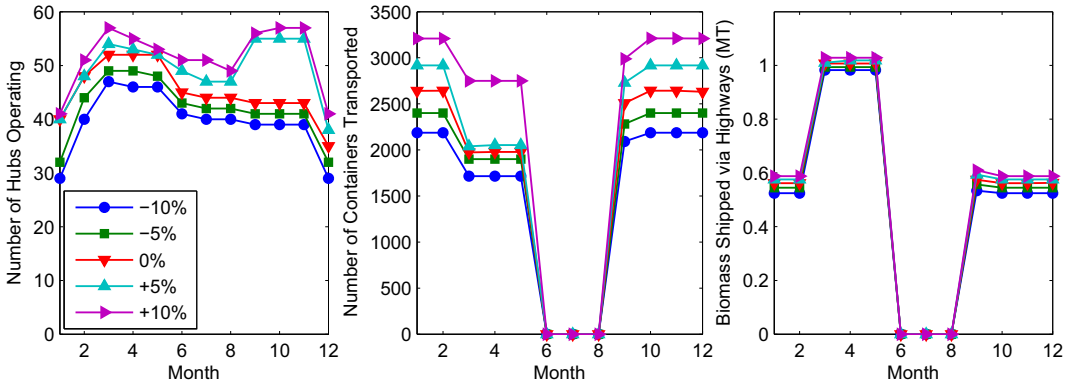


Fig. 8. Impact of supply changes on system performance.

tions in biomass supply. The impact of biomass supply fluctuations on system's performance is more evident when facilities in the supply chain are congested. For instance, when facility congestion is negligible, a 10% increase in biomass supply results in a 34.28% increase of containers used. When facilities in the network are congested, a 10% increase in biomass supply results in a 26.14% increase in the number of containers used. In this case, some amount of biomass is delivered to power plants via highways in order to alleviate congestion at multi-modal facilities.

5.2.4. Analyzing the performance of the algorithms proposed

This section summarizes our computational experience in solving model [**LDCM**] using the algorithms proposed in Section 4. The following criteria are used to stop the algorithms: (a) the gap between the upper and lower bound falls below a threshold limit ϵ, i.e., $\epsilon = |UB - LB|/UB = 0.01$ (b) the maximum time limit is reached i.e., $t_{max} = 36,000$ s and, (c) the maximum iteration limit is reached i.e., $n = 500$.

The first set of experiments (reported in Table 4) conduct an optimality test for the algorithms proposed in this study. To perform these experiments, we vary $|\mathcal{J}| = \{10, 15, 20, 25, 30\}$ while fixing $|\mathcal{I}| = |\mathcal{K}| = 10, |\mathcal{L}| = 5$, and $|\mathcal{T}| = 4$ to obtain five different problem instances. We set $\epsilon = 0.0, t_{max} = 3600$, and $n = 100$ to run these experiments. Results indicate that CPLEX is capable of solving all the problem instances (5 out of 5) in 0.0% optimality gap within the specified time limit. On the other hand, the algorithms proposed in this study (e.g., RH, RH-Benders, RH-CB) is capable of providing high quality solutions in solving model [**LDCM**] but fails to provide an optimal solution within our experimental range. Benders and combinatorial Benders decomposition algorithm provides optimal solution to 4 and 3 out of 5 problem instances, respectively; however, these solutions are achieved by sacrificing some running time of the algorithms. The last column of Table 4 shows the num-

Table 4
Optimality test of different solution approaches.

Method	#opt	Avg. Gap (%)	Avg. Time (s)	$\lvert\mathcal{H}\rvert$
CPLEX	5/5	0.00	1621.2	3
RH	0/5	1.93	89.6	2.5
Benders	4/5	0.04	421.8	3
CB	3/5	0.16	818.9	2.5
RH-Benders	0/5	0.21	201.4	2.5
RH-CB	0/5	0.27	326.2	3

ber of points in set $|\mathcal{H}|$ are needed to obtain an average optimality gap reported in the third column of the same table. Note that all the algorithms require few points ($|\mathcal{H}| \leqslant 3$) to produce the reported optimality gap. In overall, CPLEX is capable of providing optimal solution to all the small-sized problem instances while the other algorithms hold the promise to solve large-scale problem instances for which CPLEX fails to provide a feasible solution in a reasonable amount of time.

Table 5 summarizes the characteristics and size of the real world problems solved. We generated these problems by changing the number of potential multi-modal facilities to use, and the length of the planning horizon. Tables 6 and 7 present the error gap (ϵ), the running time (t_{max}) and the total number (n) and the average number of iterations ($\overline{n}$) of the algorithms proposed. Note that $\overline{n}$ are computed by taking the average of the number of iterations needed to solve the inner part of the Benders and combinatorial Benders decomposition algorithm. Since the outer part of the Benders and combinatorial Benders decomposition algorithm were solved in less than 3 iterations (r) in all test instance, we exclude this column from Tables 6 and 7. We use the value of the maximum lower bound generated using the Benders decomposition and combinatorial Benders decomposition algorithm to calculate the error gap of the rolling horizon algorithm and Benders-based rolling horizon algorithm. We are doing this because these algorithms generate upper bounds only. For each problem we highlight the results from the best algorithm. Since the stopping criteria for all problems is an error gap less than 1%, then, the best running time per problem is identified using boldface numbers. In the case when the algorithms provides solutions with an error gap greater than 1%, then, the running time of the problem with the smallest optimality gap is highlighted.

Results in Tables 6 and 7 indicate that the combinatorial Benders decomposition algorithm (CB) provides solutions with less than 1% error gap for 18 out of 30 problem instances. The Benders decomposition algorithm provides solutions with less

Table 5
Experimental problem sizes.[a]

Case	$\lvert\mathcal{I}\rvert$	$\lvert\mathcal{J}\rvert$	$\lvert\mathcal{K}\rvert$	$\lvert\mathcal{L}\rvert$	$\lvert\mathcal{T}\rvert$	Binary variables	Integer variables	Continuous variables	Total variables	Total constraints
1	274	25	59	5	4	1500	5900	1,682,100	1,689,500	8532
2	274	25	59	5	8	3000	11,800	3,364,200	3,379,000	17,064
3	274	25	59	5	12	4500	17,700	5,046,300	5,068,500	25,596
4	274	50	59	5	4	3000	11,800	3,299,300	3,314,100	15,732
5	274	50	59	5	8	6000	23,600	6,598,600	6,628,200	31,464
6	274	50	59	5	12	9000	35,400	9,897,900	9,942,300	47,196
7	274	75	59	5	4	4500	17,700	4,916,500	4,938,700	22,932
8	274	75	59	5	8	9000	35,400	9,833,000	9,877,400	45,864
9	274	75	59	5	12	13,500	53,100	14,749,500	14,816,100	68,796
10	274	100	59	5	4	6000	23,600	6,533,700	6,563,300	30,132
11	274	100	59	5	8	12,000	47,200	13,067,400	13,126,600	60,264
12	274	100	59	5	12	18,000	70,800	19,601,100	19,689,900	90,396
13	274	119	59	5	4	7140	28,084	7,762,772	7,797,996	35,604
14	274	119	59	5	8	14,280	56,168	15,525,544	15,595,992	71,208
15	274	119	59	5	12	21,420	84,252	23,288,316	23,393,988	106,812

[a] Problem size reported by setting $|\mathcal{H}| = 1$

Table 6
Comparison of different solution approaches ($c_0 = 10,000$).

	RH			Benders			CB			RH-Benders			RH-CB		
Case	ϵ	t_{max}	n	ϵ	t_{max}	$\overline{n}$	ϵ	t_{max}	$\overline{n}$	ϵ	t_{max}	$\overline{n}$	ϵ	t_{max}	$\overline{n}$
1	2.02	153.6	2	0.94	3200.6	99	0.49	2488.6	143	1.09	2400.8	88	0.52	**2133.7**	122
2	2.29	617.2	2	0.77	5124.7	138	0.62	2744.9	174	1.12	3924.7	118	0.77	**2412.4**	162
3	2.24	944.4	2	0.84	6788.2	192	0.87	4529.8	247	1.03	5897.6	172	0.97	**3989.7**	242
4	1.89	438.6	2	0.71	7437.3	84	0.82	8009.2	241	0.94	6704.8	78	0.93	**6222.3**	209
5	1.47	1284.8	2	0.88	**10021.9**	118	0.89	17425.5	217	1.07	9211.7	102	0.89	13247.4	211
6	1.52	2248.4	2	0.62	13221.8	132	0.97	24589.4	196	0.91	**12051.7**	99	0.84	19557.6	187
7	1.44	610.9	2	0.59	9542.7	79	0.86	13542.7	209	0.87	**8742.8**	72	0.97	10859.7	189
8	1.24	1742.4	2	0.94	14598.1	72	0.91	23011.4	197	0.93	**12987.1**	64	0.94	15147.8	172
9	1.29	3424.1	2	0.86	18245.8	67	1.12	36000.0	214	0.96	**16118.4**	58	0.87	22487.9	188
10	0.98	**987.3**	2	0.76	14227.9	61	0.83	22024.7	211	1.07	12589.2	52	0.97	15475.8	127
11	1.21	2948.7	2	0.91	20411.7	56	1.09	36000.0	182	0.84	17248.7	48	0.87	**16478.5**	108
12	1.08	7201.8	2	0.89	26438.4	51	2.22	36000.0	148	0.99	**23895.7**	42	0.96	26887.4	111
13	1.32	1412.8	2	0.84	20712.5	37	6.54	36000.0	199	0.98	**14596.5**	34	1.52	36000.0	221
14	1.29	4321.2	2	0.94	**28096.9**	36	8.01	36000.0	164	1.08	21888.6	33	2.55	36000.0	186
15	1.18	9606.4	2	0.98	**33230.1**	32	14.22	36000.0	114	1.04	26667.8	27	2.78	36000.0	137
Average	1.50	2529.5	2	0.83	15419.9	82	2.70	22291.1	190	0.99	**12992.7**	72	1.16	17526.7	171

Table 7
Comparison of different solution approaches ($c_0 = 100,000$).

	RH			Benders			CB			RH-Benders			RH-CB		
Case	ϵ	t_{max}	n	ϵ	t_{max}	$\bar{n}$	ϵ	t_{max}	$\bar{n}$	ϵ	t_{max}	$\bar{n}$	ϵ	t_{max}	$\bar{n}$
1	2.51	225.9	3	0.92	3267.2	98	0.67	2516.8	144	1.22	2240.4	84	0.63	**2210.2**	129
2	2.08	706.14	3	0.85	5626.7	147	0.88	2712.4	172	1.09	3994.5	118	0.84	**2487.4**	166
3	2.44	949.14	2	0.87	6922.9	194	0.93	4219.3	233	0.99	6022.1	177	0.92	**3762.4**	238
4	1.66	586.4	3	0.78	7627.9	85	0.92	8111.2	243	1.14	6922.9	83	0.96	**6427.7**	211
5	1.47	1755.2	3	0.84	**10866.3**	122	0.96	17465.7	217	1.06	9024.5	99	0.88	13017.1	209
6	1.44	4213.6	4	0.93	13198.4	110	0.86	24377.1	195	0.97	**12346.6**	101	0.93	20004.6	201
7	1.84	953.0	3	0.95	10052.7	81	0.88	13701.8	208	1.04	**8931.7**	78	0.76	10628.4	185
8	1.15	2412.7	3	0.91	14756.8	73	0.94	23178.7	198	0.97	**12855.7**	62	0.59	15772.1	179
9	1.28	8063.3	5	0.99	18474.7	69	1.58	36000.0	201	1.03	**16427.4**	57	0.94	22948.5	195
10	1.27	1459.7	3	0.87	14124.8	60	0.95	24057.8	214	1.11	**12175.8**	50	0.87	15897.3	131
11	1.04	3674.3	3	0.69	21478.4	55	1.17	36000.0	178	0.93	17514.4	47	0.95	**16899.4**	114
12	1.11	9245.8	4	0.94	26134.7	50	3.59	36000.0	252	0.88	**22957.3**	39	0.82	28222.2	113
13	1.33	1725.6	3	0.98	22814.7	39	3.62	36000.0	201	0.85	**15121.7**	35	1.47	36000.0	224
14	1.27	4917.4	3	2.54	36000.0	44	7.44	36000.0	161	**1.11**	22210.1	31	2.12	36000.0	172
15	**1.16**	10704.1	3	6.57	36000.0	36	17.54	36000.0	112	1.36	28011.5	27	3.14	36000.0	126
Average	1.54	3439.5	3	1.38	16489.7	84	2.86	22422.7	189	**1.05**	13117.1	73	1.12	17751.8	173

than 1% error gap for 28 out of 30 problem instances. On the average, the running time of Benders decomposition algorithm is 28.64% shorter as compared to the combinatorial Benders decomposition algorithm. Computational results indicate that the combinatorial Benders decomposition outperforms other algorithms when solving problems of small size. However, as the number of facilities increases, the Benders decomposition algorithm outperforms the combinatorial Benders decomposition algorithm.

Results can be further improved by incorporating rolling horizon heuristic (RH) in the Benders decomposition and combinatorial Benders decomposition frameworks. The combinatorial Benders decomposition algorithm nested with rolling horizon heuristics provides solutions with less than 1% error gap for 24 out of 30 problems. This hybrid algorithm takes on the average 21.10% less time as compared to the combinatorial Benders decomposition algorithm. The Benders-based rolling horizon algorithm provides solutions with less than 1% error gap for all the problems solved. The error gap presented in Tables 6 and 7 is calculated using the upper bound from the hybrid Benders-based rolling horizon algorithm and the lower bound from the Benders decomposition algorithm. On the average, the running time of the Benders-based rolling horizon algorithm is 25.99% shorter compared to the combinatorial Benders-based rolling horizon algorithm.

Experimental results indicate that the running time of the stand alone rolling horizon heuristic is 77.14% shorter compared to the Benders-based rolling horizon algorithm. The average error gap reported by the rolling horizon heuristic is 1.52%, as compared to 1.02% provided by the Benders-based rolling horizon algorithm. The running time of the rolling horizon heuristic is 83.08% shorter compared to the combinatorial Benders-based rolling horizon algorithm. This decrease in running time is achieved without much sacrifice of solution quality. Overall, the rolling horizon heuristic offers high quality solutions.

6. Conclusion

This paper analyzes the impacts of congestion on supply chain design and management decisions. We propose a dynamic transportation mode selection and multi-modal facility location model in order to alleviate the impacts of congestion on supply chain performance. That means, different modes of transportation, and therefore, a different set of multi-modal facilities are used in different seasons in order to better manage the flow of products in the network and reduce the impacts of congestion. This model is a mixed integer linear program called [**LDCM**]. The model captures the trade-offs that exist between investment, transportation and congestion costs.

This study proposes a number of solution algorithms to solve the problem. These algorithms are: an accelerated Benders decomposition, a combinatorial Benders decomposition, and a heuristic method commonly known as rolling horizon algorithm. The algorithms were further enhanced by incorporating rolling horizon heuristics in the Benders and combinatorial Benders decomposition framework. We test the performance of these algorithms in a case study we developed about the biomass supply chain. To develop the case study we used data from the Southeast USA. Numerical results indicate that both the Benders-based rolling horizon algorithm and the stand alone rolling horizon algorithm offer high quality solutions in a reasonable amount of time. Numerical analysis indicates that, the impacts of congestions in the supply chain can be reduced via dynamically changing the transportation modes and multi-modal facilities to use.

In summary, the contributions of the paper to the literature are manifold. First, we present a dynamic multi-modal facility location model to alleviate the impacts of congestion on biomass supply chain performance. To the best of the author's knowledge, no prior research proposes any mathematical model that offers dynamic allocation of multi-modal facilities

to alleviate congestion caused by seasonality of biomass. Second, we propose multiple customized solution approaches to find high quality solutions to large instances of our problem in a time efficient manner. Finally, a real-world case study of the model is presented that reveals the impact of multi-modal facility congestion on the biomass supply chain network. The findings can be used by decision makers to design and manage biomass supply chain network economically and efficiently.

This work can be extended to consider congestion caused by multi-modal facility disruptions. Further, it will be interesting to see how the model behaves under system uncertainty (e.g., uncertainty in supply, demand, fuel cost). These issues will be addressed in future studies.

Acknowledgments

The authors are grateful to two anonymous referees for their constructive comments that have contributed to improve the quality of this paper. This work was supported in part by the National Center for Intermodal Transportation and Economic Competitiveness (NCITEC) through project USDOT 009067-019 and National Science Foundation through grant NSF 1462420. This support is gratefully acknowledged.

References

Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1993. Network Flows: Theory, Algorithms, and Applications. Prentice-Hall, New Jersey.
An, H., Wilhelm, W.E., Searcy, S.W., 2011. A mathematical model to design a lignocellulosic biofuel supply chain system with a case study based on a region in central Texas. Bioresour. Technol. 102, 7860–7870.
Bai, Y., Hwang, T., Kang, S., Ouyang, Y., 2011. Biofuel refinery location and supply chain planning under traffic congestion. Transp. Res. Part B 45, 162–175.
Balasubramanian, J., Grossmann, I., 2004. Approximation to multistage stochastic optimization in multiperiod batch plant scheduling under demand uncertainty. Ind. Eng. Chem. Res. 43 (14), 3695–3713.
Benders, J.F., 1962. Partitioning procedures for solving mixed-variables programming problems. Numer. Math. 4, 238–252.
Bioenergy Knowledge Discovery Framework (KDF), 2013. Available from: <https://bioenergykdf.net/taxonomy/term/1036>.
Birge, J.R., Louveaux, F.V., 1997. Introduction to Stochastic Programming. Springer, New York.
Boardman, R., et al., 2013. Logistics, costs, and ghg impacts of utility-scale cofiring with 20percent biomass. Technical Report INL/EXT-12-25252 and PNNL-23492, Idaho National Laboratory and Pacific Northwest National Laboratory, <http://www.pnnl.gov/main/publications/external/technicalreports/PNNL-23492.pdf>.
Brower, M., 2010. Woody biomass Economics. Available from: <http://www.mosaicllc.com/documents/34.pdf>.
Camargo, R.S., Miranda Jr., G., Ferreira, R., Luna, H.P., 2009. Multiple allocation hub and spoke network design under hub congestion. Comput. Oper. Res. 36 (12), 3097–3106.
Chen, C.-W., Fan, Y., 2012. Bioethanol supply chain system planning under supply and demand uncertainties. Transp. Res. Part E 48, 150–164.
Ebery, J., Krishnamoorthy, M., Ernst, A., Bolanf, N., 2000. The capacitated multiple allocation hub location problem: formulations and algorithms. Eur. J. Oper. Res. 120 (3), 614–631.
Eksioglu, S.D., Acharya, A., Leightley, L.E., Arora, S., 2009. Analyzing the design and management of biomass-to-biorefinery supply chain. Comput. Ind. Eng. 57, 1342–1352.
Eksioglu, S.D., Li, S., Zhang, S., Sokhansanj, S., Petrolia, D., 2010. Analyzing impact of intermodal facilities on design and management of biofuel supply chain. Transp. Res. Rec. 2191, 144–151.
Elhedhli, S., Hu, F.X., 2005. Hub-and-spoke network design with congestion. Comput. Oper. Res. 32, 1615–1632.
Elhedhli, S., Wu, H., 2010. A Lagrangean heuristic for hub-and-spoke system design with capacity selection and congestion. INFORMS J. Comput. 22 (2), 282–296.
Energy Information Administration, 2014. U.S. Grain Transportation: Rail Shipments Make Big Jump. Available from: <http://agfax.com/2014/12/18/u-s-grain-transportation-rail-shipments-make-big-jump/>.
Fischetti, M., Lodi, A., 2003. Local branching. Math. Program. 98, 23–47.
Gebreslassie, B.H., Yao, Y., You, F., 2012. Design under uncertainty of hydrocarbon biorefinery supply chains: multiobjective stochastic programming models, decomposition algorithm, and a comparison between CVaR and downside risk. AIChE J. 58 (7), 2155–2179.
General Algebraic Modeling System (GAMS), 2013. Available from: <http://www.gams.com/>.
Ghaderi, A., Boland, N., JabalAmeli, M.S., 2012. Exact and heuristic approaches to the budget-constrained dynamic uncapacitated facility location-network design problem. Available from: <http://www.optimization-online.org/DB-FILE/2012/02/3336.pdf>.
Gonzales, D., Searcy, E.M., Eksioglu, S.D., 2013. Cost analysis for high-volume and long-haul transportation of densified biomass feedstock. Transp. Res. Part A 49, 48–61.
Grove, G.P., OKelly, M.E., 1986. Hub networks and simulated schedule delay. Pap. Regional Sci. Assoc. 59, 103–119.
Hajibabai, L., Ouyang, Y., 2013. Integrated planning of supply chain networks and multimodal transportation infrastructure expansion: Model development and application to the biofuel industry. Comput.-Aided Civ. Infrastruct. Eng. 28, 247–259.
Hess, J.R., Wright, T.C., Kenney, L.K., Searcy, M.E., 2009. Uniform-format solid feedstock supply system: a commodity-scale design to produce an infrastructure-compatible bulk solid from lignocellulosic biomass. INL/EXT-09-15423. Available from: <http://www.inl.gov/technicalpublications/Documents/4408280.pdf>.
Hinojosa, Y., Kalcsics, J., Nickel, S., Puerto, J., Velten, S., 2008. Dynamic supply chain design with inventory. Comput. Oper. Res. 35, 373–391.
Huang, Y., Chen, C.W., Fan, Y., 2010. Multistage optimization of the supply chains of biofuels. Transp. Res. Part E 46 (6), 820–830.
Karimi, H., Eksioglu, S.D., Hu, M., 2014. A model for analyzing the impact of production tax credit on renewable electricity production. In: Proceedings of the Industrial & Systems Engineering Research Conference (ISERC), May 31–June 3, Montreal, Canada.
Kim, J., Realff, M.J., Lee, J.H., 2011. Optimal design and global sensitivity analysis of biomass supply chain networks for biofuels under uncertainty. Comput. Chem. Eng. 35, 1738–1751.
Kostina, A.M., Guillen-Gosalbeza, G., Meleb, F.D., Bagajewiczc, M.J., Jimeneza, L., 2011. A novel rolling horizon strategy for the strategic planning of supply chains. Application to the sugar cane industry of Argentina. Comput. Chem. Eng. 35, 2540–2563.
Kumar, A., Cameron, J.B., Flynn, P.C., 2005. Pipeline transport and simultaneously saccharification of corn stover. Bioresour. Technol. 96 (7), 819–829.
Magnanti, T.L., Wong, R.T., 1981. Accelerating Benders decomposition: algorithmic enhancement and model selection criteria. Oper. Res. 29, 464–484.
Mahmudi, H., Flynn, P., 2006. Rail vs. truck transport of biomass. Appl. Biochem. Biotechnol. 129 (1), 88–103.
Marianov, V., Serra, D., 2003. Location models for airline hubs behaving as M/D/c queues. Comput. Oper. Res. 30, 983–1003.
Marufuzzaman, M., Eksioglu, S.D., Hernandez, R., 2014. Environmentally friendly supply chain planning and design for biodiesel production via wastewater sludge. Transp. Sci. 48 (4), 555–574.

Marufuzzaman, M., Eksioglu, S.D., Huang, Y., 2014. Two-stage stochastic programming supply chain model for biodiesel production via wastewater treatment. Comput. Oper. Res. 49, 1–17.
Marufuzzaman, M., Eksioglu, S.D., Li, X., Wang, J., 2014. Analyzing the impact of intermodal-related risk to the design and management of biofuel supply chain. Transp. Res. Part E 69, 122–145.
Melo, M.T., Nickel, S., da Gama, F.S., 2005. Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. Comput. Oper. Res. 33, 181–208.
Miranda Jr., G., de Camargo, R.S., Pinto, L.R., Conceicao, S.V., Ferreira, R.P.M., 2011. Hub location under hub congestion and demand uncertainty: the Brazilian case study. Pesqui. Oper. 31 (2), 319–349.
Naoum-Sawaya, J., Elhedhli, S., 2010. A nested Benders decomposition approach for telecommunication network planning. Naval Res. Logist. 57 (6), 519–539.
Oosterhuis, M., Molleman, E., Vaart, T.V.D., 2005. Research methodologies in supply chain management chapter. In: Multilevel Issues in Supply Chain Management. Physica-Verlag HD, pp. 283–297.
Parker, N., Tittmann, P., Hart, Q., Lay, M., Cunningham, J., Jenkins, B., 2008. Strategic Assessment of Bioenergy Development in the West: Spatial Analysis and Supply Curve Development. Western Governors' Association.
Pels, E., Verhoef, E.T., 2004. The economics of airport congestion pricing. J. Urban Econ. 55 (2), 257–277.
Raffarin, M., 2004. Congestion in European airspace. A pricing solution? J. Transp. Econ. Policy 38 (1), 109–125.
Roni, M.S., Eksioglu, S.D., Searcy, E., Jha, K., 2014. A supply chain network design model for biomass co-firing in coal-fired power plants. Transp. Res. Part E 61, 115–134.
Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A., 2005. A stochastic programming approach for supply chain network design under uncertainty. Eur. J. Oper. Res. 167, 96–115.
Sasaki, M., Fukushima, M., 2003. On the hub-and-spoke model with arc capacity constraints. J. Oper. Res. Soc. Jpn. 46 (4), 409–428.
Searcy, E., Flynn, P., Ghafoori, E., Kumar, A., 2007. The relative cost of biomass energy transport. Appl. Biochem. Biotechnol., 639–652
SteadieSeifi, M., Dellaert, N.P., Nuijten, W., Woensel, T.V., Raoufi, R., 2014. Multimodal freight transportation planning: a literature review. Eur. J. Oper. Res. 233, 1–15.
The National Energy Technology Laboratory, 2005. Coal-Fired Power Plants in the United States. Available from: <http://www.netl.doe.gov/energyanalyses/hold/technology.html>.
Vidyarthi, N., Jayaswal, S., 2014. Efficient solution of a class of location–allocation problems with stochastic demand and congestion. Comput. Oper. Res. 48, 20–30.
Xie, W., Ouyang, Y., 2013. Dynamic planning of facility locations with benefits from multitype facility colocation. Comput.-Aided Civ. Infrastruct. Eng. 28 (9), 666–678.
Xie, F., Huang, Y., Eksioglu, S.D., 2014. Integrating multimodal transport into cellulosic biofuel supply chain design under feedstock seasonality with a case study based on California. Bioresour. Technol. 152, 15–23.