Computers & Industrial Engineering 144 (2020) 106433

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Omnichannel logistics network design with integrated customer preference for deliveries and returns

Javier Guerrero-Lorente[a], Adriana F. Gabor[b,*], Eva Ponce-Cueto[c]

[a] Department of Management Engineering, Business Administration and Statistics, Technical University of Madrid, Madrid, Spain
[b] Research Center of Digital Supply Chain and Operations and Department of Applied Mathematics, Khalifa University, Abu Dhabi, United Arab Emirates
[c] MIT Center for Transportation and Logistics, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Keywords:
Omnichannel logistics
Network design
Customer preference
Returns
E-commerce

ABSTRACT

This paper proposes a mixed integer program (MIP) for the network design problem of a parcel carrier that manages online orders from omnichannel retailers. The network includes several types of facilities, such as city distribution centers, intermediary depots, parcel offices, as well as collect channel points, such as automated parcel stations (APS), stores or kiosks. The model formulation takes into account the influence of these collection points on consumer choice and the maximum distance customers are willing to walk to reach them. Realistic transportation costs are also considered, including detailed long haul costs and delivery costs in an area. The transportation costs inside an area are estimated via the Continuous Approximation proposed by Newell (1971) and Newell (1973), which results in a simpler problem than a location-routing problem, however, at the cost of losing linearity. For a special case of the problem, we propose a heuristic that, in our experimental setting, is on average 42.5 faster than the overall MIP and gives solutions within 1.02% of the optimum. Finally, we use the model to discuss the network design of a Spanish parcel carrier operating in Madrid.

1. Introduction

In current retail, when consumers are shifting from one trading channel to another, the boundaries between e-commerce and purchases in physical stores are hard to identify. According to e-marketer (Lipsman, 2019) statistics, at the end of 2019, the anual retail e-commerce sales would have grown worldwide by 20.7% compared to 2018, with a retail share of 14.7%, equivalent to 3.535 trillion dollars. It is expected that in 2023 the e-commerce share will reach 22%.

Digital catalogues of products and a wide variety of support services attached to them are increasing channel migrations. Rigby (2011) refered to omnichannel retailing as a process that enables retailers to interact with customers through countless channels. New delivery channels, such as Automated Parcel Stations (APS), provide added convenience to customers and help to increase their loyalty. At the same time, they are beneficial for retailers and parcel carriers, as they have the potential of reducing transportation costs and help to avoid repeated delivery visits (McLeod, Cherret, & Song, 2006).

Parcel carriers or Third Party Logistics Providers (3PLs) play a very important role in last mile distribution of online orders. They often provide customers several choices such as home delivery and click and collect (Lang & Bressolles, 2013). Click and collect include several options such as in store pick up, APSs, delivery at gas stations and kiosks. Several recent survey based studies indicated that price, location of click and collect services and trackability of parcels are main factors in the adoption of new delivery channels (de Oliveira, Morganti, Dablanc, & de Oliveira, 2017; Iwan, Kijewska, & Lemke, 2016). Although the number of facilities installed impacts the customers' preferences for a specific delivery channel through the distance that customers need to travel, this relationship has not yet been taken into account in the quantitative models used to design the delivery network of a 3PL.

Our paper focuses on the modeling of the last mile distribution network of a typical parcel carrier (or 3PL) that copes with the delivery of online orders and subsequent pick up of commercial returns through different existing and new channels (APSs, kiosks, stores). In particular, the paper takes into account customer preferences for the delivery mode, the maximum walking distance to a delivery option and the impact of channel availability on customers' chanel choice. We consider demand per chanel to be endogenoeus to the model, by assuming that more installed channels of a certain type generate more demand for that specific channel type, due to convenience and increased customer familiarity with the delivery mode. Furthermore, we take explicitely into account the distance customers are willing to walk to pick up

* Corresponding author.
E-mail addresses: javier.guerrero@upm.es (J. Guerrero-Lorente), adriana.gabor@ku.ac.ae (A.F. Gabor), eponce@mit.edu (E. Ponce-Cueto).

https://doi.org/10.1016/j.cie.2020.106433
Received 13 July 2019; Received in revised form 13 February 2020; Accepted 20 March 2020
Available online 09 April 2020
0360-8352/

deliveries/return products.

The paper is structured as follows: Section 2 discusses recent literature and the contribution of the paper. Section 3 describes the problem and summarizes it in a conceptual framework that is used to develop a non-linear mixed integer programming (MIP) model in Section 4. Section 5 proposes a fast MIP- based heuristic for a special case of the problem. In Section 6 the model is applied to a relevant Spanish parcel distribution network. Section 7 discusses the performance of the heuristic. Section 8 contains conclusions and discussions.

2. Literature review

Efficient back-end fulfillment, last mile distribution and returns pick up and processing are considered to be key logistics processes for omnichannel retailers. (see Hübner, Kuhn, & Wollenburg, 2016; Marchet, Melacini, Perotti, Rasini, & Tappia, 2018). We will next revise recent literature related to each of these processes.

Back-end Fulfillment. Different products such as fast-moving consumer goods, apparel and electronic appliances require different fulfillment allocation strategies. de Koster (2002) describes the pros and cons of different fulfillment strategies, such as distributing from a dedicated fulfilment center, from stores or from a hybrid structure.

Zhang, Lee, Wu, and Choy (2016) and Yadav, Tripathi, and Singh (2017) compare a flexible network design in which customer orders can be fulfilled directly from the most appropriate echelon (manufacturer, central distribution center, regional distribution center) to a traditional design, in which customers are served only from regional DCs. Due to better facilities utilization, the flexible network design proves to be superior to the traditional design, under different objectives, such as: facilities and transportation costs, environmental cost and maximum customer coverage. A similar conclusion is reached in Yadav, Tripathi, and Singh (2019), where a flexible network design is compared to a traditional one, under uncertainty of the online demand. As omnichannel retailers usually have very large networks, finding optimal solutions to the network design problem in reasonable time can be difficult. Zhang et al. (2016) and Zhang, Zhu, Li, and Wang (2019) develop very efficient meta-heuristics for dealing with the design of a flexible distribution network, under both deterministic and random demand. The model we consider in this paper can be seen as a hybrid model. There is a flexibility for distribution to the new channels via a nearby store or via the intermediate(regional) depots, which in turn get the merchandise from the central distribution centers. However, distribution to existing channels can be done only from the CDCs, due to economy of scale in transportation, while distribution to new channels is done via intermediate depots, due to the advantage of aggregating orders to a certain district.

Another recent literature stream focusses on exploiting the advantages obtained by an omnichannel retailer from pooling online demand and deciding from which facility to fulfill it (see Acimovic & Graves, 2014; Govindarajan, Sinha, & Uichanco, 2018; Mahar & Wright, 2009). Bretthauer, Mahar, and Venakataramanan (2010) proposes a non-linear MIP model for a two echelon problem, concerned with deciding whether a location should handle both online fulfillment and traditional sales and the necessary inventory at each location. The authors illustrate, via numerical experiments, the impact of the percentage of online demand and of the different costs considered (fixed costs, holding, backorder and transportation costs) on the network design.

Returns management and network design. Network design in the presence of returns has been extensively studied in the context of reverse logistics. Fleischmann, Beullens, Bloemhof-ruwaard, and Van Wassenhove (2001) was among the first to study the impact of reverse logistics on the design of a distribution network. By modeling the design problem as a MILP, the authors show that a high percentage of returns impacts the network design considerably. Most of the papers in the reverse logistics field focus on a production environment and separate the design of the forward and reverse networks (Alumur, Nickel, Saldanha-da Gama, & Verter, 2012; Jayaraman, Patterson, & Rolland, 2003; Min, Ko, & Ko, 2006). To the best of our knowledge, there are only a few papers that discuss the design of a network that integrates both forward and reverse logistics processes. Krikke, Bloemhof-Ruwaard, and Van Wassenhove (2003) proposes a MILP model for designing the forward and reverse logistics network for a refrigerator manufacturer. A multiperiod MILP model that combines strategic decisions (forward and reverse network design) and tactical decisions (production, distribution, storage) is proposed in Salema, Póvoa, and Novais (2009). Guerrero-Lorente, Ponce-Cueto, and Blanco (2017), introduces a simple general MILP model for a distribution network of a retailer with fulfillment facilities, return facilities and product/returns exchange points to deliver online orders or pick up returns. However, this model does not include the last mile deliveries, which are usually handled by a parcel carrier. Several papers address the issue of network design when demand and returns are uncertain (see among others, Listeş & Dekker (2005), Khatami, Mahootchi, & Farahani (2015)). For a comprehensive review of models and techniques for reverse logistics network design we refer to Agatz, Fleischmann, and Van Nunen (2008) and Govindan, Soleimani, and Kannan (2015). In this paper, we study the integrated direct and reverse logistic network of a parcel carrier. In our model, demand and returns mainly interact by sharing the capacity of the same distribution channels. Our model is deterministic, as we assume that average estimates of online demand for an area can be accurately predicted, as well as the percentage of returns. The distinguishing feature of our model is that demand for new channels depends on the number of channels installed and to the distance customers are willing to walk, unlike in a production environment, where such dependency does not exist.

Last mile distribution. Once the strategy for allocation of the fulfillment and returns processing is decided by retailers, last mile distribution processes ensure that orders and commercial returns are shipped to and collected from consumers.

The integration of new channels in the distribution network has pushed couriers to redesign their network, fleet capacity and delivery modes to cope with increasing demand for e-commerce channels. Lim, Rabinovich, Rogers, and Lasester (2016) provide a taxonomy of last-mile networks for omni-channel retailers, depending on the service speed and the variety of products offered. Lim and Shiode (2011) studies the case of a 3PL with a hub and spoke network that serves an omnichannel retailer. Via a discrete event simulation, they found that the network is likely to evolve into a more centralized structure in the long run if online demand via home delivery increases consistently. Deutsch and Golany (2018) propose a MIP model for the problem of finding the parcel lockers locations and the needed capacity in order to minimize the network costs, comprised from the fixed set-up costs of the lockers and the discounts given to customers for their willingness to travel. They discuss their model for the city of Toronto.

While network design is a strategic decision, last mile delivery is an operational one. Although strategic and operational decisions are usually studied independent of each other, due to the high cost of last mile deliveries, transportation costs should be accounted for when designing the logistic network of a 3PL provider. The problem of simultaneously optimizing the costs of facilities and transportation costs to visit facilities and customers, the so called location-routing problem, has extensively been studied in the OR literature. Several MIP formulations and heuristics have been proposed for variants of the location- transportation problem with one or more echelons. For extensive literature reviews on location-routing problems, we refer the reader to Nagy and Salhi (2007), Drexl and Schneider (2015) and Prodhon and Prins (2014).

However, in many situations, a detailed planning of the routes is not necessary at the moment of network design. In such situations, often the so called Continuous Approximation is used. The approximation has been initially proposed by Newell (1971) and Newell (1973) and allows fairly accurate estimations of the cost in homogeneous areas. The approximation has been later on used and improved to tackle several problems related to location, routing and inventory. For a thorough review of the method we refer to Ansari, Başdere, Li, Ouyang, and Smilowitz (2018). In our paper, we will use the continuous approximation proposed by Winkenbach, Kleindorfer, and Spinler (2015), which builds on an approximation in Daganzo (1999). Winkenbach et al. (2015) embeds this approximation in a MIP to find the optimal network design and fleet structure for La Poste, a carrier in France.

Customer choice and facility location models From the perspective of mathematical modelling, our paper is most related to the field of multi-level (hierarchical) facility location problems. According to the classification in Farahani, Hekmatfar, Fahimnia, and Kazemzadeh (2014), our model is a variant of a fixed cost, multi-flow, capacitated, deterministic facility location problem. We refer to Farahani et al. (2014) and Şahin and Süral (2007) for a very comprehensive overview of this type of models and solution methods. The literature on incorporating customer choice in facility location and last mile delivery is quite scarce. So far, the main focus of literature has been on the choice of delivery time slots and how pricing can be used as an incentive for designing more efficient routes (Asdemir, Jacob, & Krishnan, 2009; Campbell & Savelsbergh, 2006; Yang, Strauss, Currie, & Eglese, 2016). To the best of our knowledge, the dependence between the number of channels installed and demand has not been studied before in hierarchical facility locations.

Contribution to the literature. This paper describes a MIP model for the network design problem of a parcel carrier serving omnichannel retailers, that consider both strategical decisions (opening of facilities) and operational ones (last mile transportation costs). A distinguishing feature of our model is the inclusion of the interaction between demand and the number of facilities installed; the more new delivery channels are in an area, the higher the demand for delivery through that channel. We also incorporate a detailed last mile transportation costs function, that is approximated via a piecewise linear function. For a specific case, we propose a fast MIP heuristic, that in our scenarios gives solutions within 1% of the optimum and is much faster than the MIP model. Finally, we use the model to discuss the network design problem of a parcel carrier that serves online orders and the associated returns in the city of Madrid. Extensive numerical results study the impact of the problem's parameters and the quality of the heuristic.

3. Problem description

3.1. Physical flow of products in omnichannel retail

On the websites of omnichannel retailers, several delivery options are offered for both purchase orders and commercial returns management. Consumers make a choice depending on urgency, location to pick up or drop off the goods and transportation fees.

Online orders are prepared for shipment at order fulfillment facilities, such as: *distribution centers*, that combine store replenishment tasks with online orders fulfillment; *fulfillment centers* which are dedicated warehouses for online orders fulfillment, *stores*, that allow traditional purchases and a limited capacity to prepare online orders.

Orders prepared in distribution and fulfillment centers are picked up by a carrier to be delivered to a City Distribution Center (CDC), a facility where orders are sorted and consolidated by destination. The typical delivery network of carriers consist of different channels: *existing traditional channels* such as postal or parcel offices, and *new channels*, such as automated parcel stations (APS), small convenience stores and home delivery. Traditional channels usually have large capacity to hold goods, facilitating direct delivery with full trucks from the CDC. However, more convenient locations and flexibility with respect to delivery and drop off times make the new channels increasingly popular. Distribution to new channels, which have smaller capacity, require satellite intermediate depots (ID) where goods that travel from the CDC are deconsolidated and sorted by delivery area. Most of these intermediate depots already exist to manage home delivery orders.

In order to save distribution time, brick and mortar retailers may choose to fulfill urgent online orders (with a lead time of two or three hours) from the nearest store. It is customary that urgent orders are picked up and delivered by the same vehicles that travel between ID and customers. Moreover, most customers choose the same distribution channels for both deliveries and returns. Commercial returns collected by the carrier are sent to the retailer's facility where the orders were initially prepared.

To remain competitive, carriers need to determine the right network configuration and mix of distribution channels so that demand is satisfied while minimizing the overall costs. Carriers who fail to do so might face competition from retailers who choose a competitor or set up their own delivery network when economies of scale for deliveries are reached.

3.2. Problem description

In this paper we address the problem faced by a parcel carrier that wishes to redesign its distribution network and to incorporate a diversified set of options for order delivery and pick up of returns.

The parcel carrier has a network of existing channels through which parcels are delivered to a certain region. Each existing channel covers a certain demand area, for example an area identified by a certain postal code. In the case of a parcel carrier, an existing channel is the post office in the area. Without loss of generality, the delivery region is divided in L smaller areas, each corresponding to the service area of an existing channel.

The carrier wishes to extend its services to deliver items for omnichannels retailers. As such, it is interested in opening in each area new delivery and returns channels (APSs, delivery through existing stores or kiosks, home delivery) to offer more flexibility to customers. To support these new operations, the carrier has to open new central distribution centers (CDCs) and intermediary depots (IDs).

A schematic representation of the network is given in Fig. 1.

Customers in an area may place two types of orders: *regular* and *urgent* orders.

Regular orders follow two routes: they can be delivered directly from city distribution centers (CDCs) to the existing channels, or they can be delivered to the new channels via intermediary depots, where orders are transferred to smaller vehicles for deliveries in the demand areas. Regular orders may be rejected, case in which a penalty is payed. The returns of regular orders follow the two paths described above in reverse.

Urgent orders are picked and packed at the stores and delivered to new channels by the same vehicles that transport items between the IDs and the new channels. As urgent deliveries are more expensive, we assume that they have priority above regular orders, and therefore cannot be rejected by the carrier.

When (re)designing its logistics network, a carrier needs to decide the location of IDs and CDCs, as well as the capacity that needs to be installed in new channels, such that total costs are minimized. The costs that need to be accounted for are: the fixed costs for facilities, the

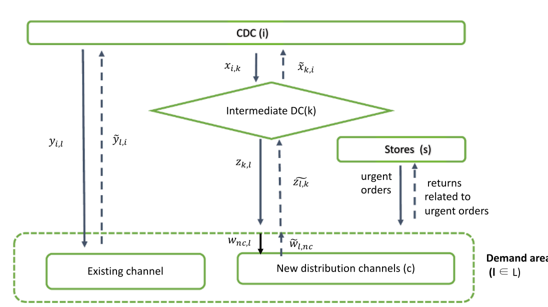


Fig. 1. Conceptual framework for an omnichannel distribution network.

transportation costs, the processing costs of the orders and returns and the penalty costs for undelivered items or unprocessed returns. When planning an adequate network, the carrier needs to consider customer preferences for a certain channel, as well as the maximum distance a customer is willing to walk to pick/up or return an item. At the same time, the carrier has to take into account the physical restrictions imposed by the present facilities, such as processing capacity.

In our modelling approach, we made a few assumptions that are listed below.

Assumption regarding demand modelling

As noted in empirical research, location and convenience of use play an important role in the adoption of new channels (see de Oliveira et al., 2017; Iwan et al., 2016). Hence, the more new channels customers see or are exposed to in an area, the more they will be tempted to make use of them. As the number of channels installed is a variable in our model, one cannot estimate in advance the demand for a certain channel. In this paper, we propose to model this relationship via a linear regression. Below we give the exact assumptions we make regarding demand.

A1. In each demand area l, demand follows a homogeneous spatial Poisson distribution, with demand density λ_l per unit area

A2. For each new channel, the carrier can estimate the maximum distance $WD_{nc,l}$ customers in demand region l are willing to walk to reach the channel.

A3. A location where new channels are installed, can serve an area of $Area_{nc,l} = \pi WD_{nc,l}^2$, hence a maximum demand of $\lambda_l Area_{nc,l}$.

A4. Demand for a new channel depends linearly on the number of channels installed in an area and the carrier is capable of estimating the coefficients of this linear relationship, for example, by using a linear regression.

A5. The returns via existing or new channels are a known percentage of the demand.

A6. A vehicle serves one region only; there is no routing problem between demand regions, only inside a region.

Assumptions regarding new channels

Without loss of generality, we divide the new channels in 3 groups: the first group, called *PH*, such as stores and kiosks, for which one facility can be installed at each location; second, home deliveries, that do not need any facilities; finally, the third group, such as APSs, for which several facilities can be installed at one location. For the easiness of the presentation, we will assume that the third channel type is formed by one class, APS. The exact assumptions regarding the new channels are enumerated below.

A7. More APSs can be installed at one location. For all the other new channels, one facility can be installed at one location

A8. The number of locations where APSs are installed should not exceed the maximal number imposed by the distance customers are willing to walk. To capture this aspect, we assume that each region l is covered by $\frac{Area_l}{Area_{APS,l}}$ circles, where $Area_l$ is the area of region $l \in L$. This gives a maximum of $\frac{area_l}{Area_{APS,l}}$ APS locations.

A9. The carrier is more interested in the capacity in new channels needed to be installed in an area, then the exact location of the channels. For calculating the transportation costs, we assume that the locations of the new channels are uniformly distributed in the area.

A10. The carrier wants to encourage the use of new channels, hence she offers a discount for each item delivered through a new channel. Clearly, if the carrier is insensitive to the channel choice, the discount will be zero.

Assumptions regarding CDCs, IDs and existing channels

A11. As the carrier already posses the existing channels, she imposes a minimum volume that has to be direct through it.

A12. Possible locations for CDCs and IDs have been previously identified.

Assumptions regarding the network

A13. Each demand area is assigned to at most one ID and at most one CDC. Each of these facilities process the deliveries as well as the returns from an area. We assume that for each demand area, there exists at least one ID or CDC capable to handle the demand of that area.

A14. Each open ID is assigned to at most one CDC. All the deliveries and returns to/from the ID are processed by the same CDC.

4. Mathematical model

In this section we propose a nonlinear mixed integer program (MIP) to address the problem described in Section 3. We will refer to this MIP as MIP-OLN. We first introduce the notation and variables used, give the MIP and then discuss several modelling aspects.

Index sets

I: potential locations for CDCs (indexed by i)
K: potential locations for intermediary depots (indexed by k)
L: demand regions (indexed by l)
EC: set of existing channels (parcel offices)
PH: set of new channels for which one facility can be opened at each location
APS: set of new channels, for which more facilities can be opened at one location
NC: set of new channel types $PH \cup APS \cup \{home\}$ indexed by nc
C: set of distribution channels, $C = NC \cup EC$

Parameters *Transportation costs*

$tc_{a,b}$ - transportation costs for deliveries between a and b
$\widetilde{tc}_{b,a}$ - transportation costs for returns between b and a

Asset costs

f_u - fixed cost for opening facility u, where $u \in I \cup K$
$f_{c,l}$ - fixed costs for opening distribution channel c in area l, $c \in C$, $l \in L$

Capacities and facility bounds

cap_u - capacity of facility u, $u \in I \cup K$
$cap_{l,c}$ - capacity of channel c in area l
$e_{nc,l}$ - upperbound on the number of new channels that can be built in area l

Parameters related to demand and returns

d_l: demand of regular orders in area l
d_l^{urg} ($\widetilde{d}_l^{urg}$): demand (returns) of urgent orders in area l (orders delivered from store)
η_l: percentage of demand in area l that is returned
LB_l: minimum volume of demand and returns that has to be directed through existing channels
$Area_{nc}$ - area served by a new channel, calculated based on the distance a customer is willing to walk, $nc \in NC$
λ_l - demand density per area unit in region l

Parameters related to processing orders

p_a - cost for processing an order in location a, where $a \in C \cup K \cup I$

Penalty costs and discounts

pen - penalty cost for an unsatisfied order
v_{nc} - discount per order/return for using channel $nc \in NC$

Decision variables
Variables related to opening facilities/capacity offered in new channels

α_i: indicator variable taking value 1 if CDC i is open
β_k: indicator variable taking value 1 if ID k is open
$\gamma^{nr}(c, l)$: number of locations where the new channel c is installed in area l, $c \in NC\setminus\{home\}$
γ_l^{vol}: number of APS installed in demand area l

Assignment variables

$\tau_{l,k}$: indicator variable taking value 1 if demand area l is assigned to the intermediate depot k for delivery
$\rho_{k,i}$: indicator variable taking value 1 if intermediate depot k is assigned to CDC i
$\sigma_{l,i}$: indicator variable taking value 1 if area l is assigned to CDC i

Variables related to deliveries

$x_{i,k}$: units sent from the CDC i to intermediate depot $k \in K$
$y_{i,l}$: units sent from CDC i to area l
$z_{k,l}$: units delivered from processing facility k to demand area $l \in L$
$w_{nc,l}$: units delivered in area l via distribution channel $nc \in NC$

Variables related to returns

$\tilde{x}_{k,i}$: units sent from the intermediate depot $k \in K$ to CDC i
$\tilde{y}_{l,i}$: units returned from area l to CDC i
$\tilde{w}_{l,nc}$: flow of returned units picked up from customers in group $l \in L$ via channel $nc \in NC$
$\tilde{z}_{l,k}$: flow of items returned from demand area $l \in L$ to intermediate depot $k \in K$

Variables related to unsatisfied demand and returns

$u_{c,l}$: unsatisfied regular demand in area l that had to be delivered via channel $c \in C$
$\tilde{u}_{l,c}$: unsatisfied returns via new channel $c \in C$ from demand area l

We will next discuss the objective function of the model.

4.1. Objective function

Fixed cost of facility operations

$$\sum_{i\in I} f_i\alpha_i + \sum_{k\in K} f_k\beta_k + \sum_{c\in PH, l\in L} f_{c,l}\gamma_{c,l}^{nr} + \sum_{l\in L} f_{APS,l}\gamma_l^{vol}$$

Processing costs at CDC and ID

$$\sum_{i\in I} \left[\sum_{l\in L} p_i(y_{i,l} + \tilde{y}_{l,i}) + \sum_{k\in K} (x_{i,k} + \tilde{x}_{k,i})\right] +$$

Processing costs at existing and new channels

$$\sum_{l\in L, nc\in NC} p_{nc}(w_{nc,l} + \tilde{w}_{l,nc}) + \sum_{l\in L, i\in I, ec\in EC} p_{ec}(y_{i,l} + \tilde{y}_{l,i}) +$$

Penalty costs and channel discounts

$$pen \sum_{c\in C, l\in L} (u_{c,l} + \tilde{u}_{l,c}) + \sum_{nc\in NC, l\in L} v_{nc}(w_{nc,l} + \tilde{w}_{l,nc})$$

Transportation costs between CDCs and IDs

$$\sum_{i\in I, k\in K} [tc_{i,k}x_{i,k} + \tilde{tc}_{k,i}\tilde{x}_{k,i}] +$$

Transportation costs from IDs/CDC to demand areas

$$\sum_{k\in K, l\in L} (tc_{k,l}z_{k,l} + \tilde{tc}_{l,k}\tilde{z}_{l,k}) + \sum_{i\in I, l\in L} (tc_{i,l}y_{i,l} + \tilde{tc}_{l,i}\tilde{y}_{l,i}) +$$

Transportation costs inside area l, independent on the nr. of stops

$$\sum_{l\in L} tc_l^{in}\left(\sum_{nc\in NC} w_{nc,l} + \sum_{nc\in NC} \tilde{w}_{l,nc}\right) +$$

Transportation costs inside area l, related to the number of stops

$$tc_l^{st}[n_{l,stops}].$$

The objective of the carrier is to minimize the total costs, that include the fixed costs for CDCs, intermediate depots and the installed channels, proccessing costs of direct orders and returns, penalty costs associated to rejected deliveries or returns and the discount costs offered for using certain channels and transportation costs related to the direct flows of products and to the flows of returns.

Note that for the fixed costs of new channels, we distinguish between the costs of channels in *PH* (stores and kiosks) and costs for APSs, as more APSs can be installed at one location. At the CDCs, processing costs are incurred for all items delivered via existing channels, without transshipment and for the items transshipped via intermediary depots. At IDs and new channels, processing costs are incurred only for the direct and reverse flows of items delivered through the respective channels. Penalty costs are incurred for all the unfulfilled direct orders and returns, while discounts are given only for the use of new channels (see assumption A10).

The transportation costs are calculated following Winkenbach et al. (2015). We mention here only the main assumptions, an in depth discussion of the transportation costs can be found in Section 6. The costs incurred for transporting items from an ID to the delivery point are divided in two parts: one part accounting for transportation from ID to region l, and the transportation costs inside region l, where the vehicle makes a tour visiting the new channels. For all the deliveries with a line haul trip, where the vehicle does not stop between origin and destination, the transportation costs are linear in the number of orders delivered/returned. This is the case for the transportation costs between CDCs and IDs, CDCs and demand regions and IDs and demand regions.

To estimate the transportation costs inside area l exactly, one would need to solve a vehicle routing problem between the locations where new channels are installed and between the locations where home deliveries are required. Even if the location of the home deliveries were known, the resulting problem is a large instance of a location-routing problem, which is hard to solve. An alternative for using heuristics to solve large location-routing problems, is to use the Continuous Approximation (CA) proposed in Newell (1971) and Newell (1973). The CA has been used to approximate several variants of vehicle routing problems, such as traveling salesman problem (TSP) Daganzo (1984), multi-echelon distribution with intermediate consolidation and transshipment facilities (Daganzo, 1988) or has been incorporated in facility location problems, such as in (Cachon, 2014). The approximation is known to work well for routing problems with a large density of locations that need to be visited (Ansari et al., 2018). A review of the development and applications of CA can be found in Ansari et al. (2018). In this paper we adopt the continuous approximation of transportation costs in a demand area proposed in Winkenbach et al. (2015). The approximation is based on the method proposed by Daganzo in Daganzo (1999) (see Chapter 4).

The main idea is to estimate a demand area $Area_l$ by $n_{l,stops}d^2$, where $n_{l,stops}$ represents the number of stops in area l and d is the expected distance between 2 stops, assuming locations are uniformly distributed in the area. As such, the distance between 2 stops is approximated by $\sqrt{\frac{Area_l}{n_{l,stops}}}$. This approximation works well when the number of stops is large, which is the case for omni-channel retailers. The total cost related to transportation *between stops* is thus equal to

$$t_l tc_v \sqrt{Area_l n_{l,stops}} \tag{1}$$

where tc_v represent the costs per time unit for operating a vehicle and t_l represents the estimated time needed to travel between two consecutive stops. For new channels $nc \in NC\setminus\{home\}$, there is one stop at each facility installed, hence there are $\sum_{nc\in NC\setminus\{home\}} \gamma_{nc,l}^{nr}$ stops. For home deliveries and returns, there is a stop for each corresponding order/return, hence there will be $w_{home,l} + \tilde{w}_{l,home}$ stops. This gives

$$n_{l,stops} = \sum_{nc \in NC} \gamma_{nc,l}^{nr} + w_{home,l} + \tilde{w}_{l,home}, \quad (2)$$

The quality of this approximation is assessed in Winkenbach et al. (2015), where the results of the approximation are compared to the results of the Tabu Search procedure proposed in Côté and Potvin (2009). The authors show that for large number of deliveries (above 40 per 1 km^2), the continuous approximation deviates from the Tabu Search by less than 5%. Note that in case of an omnichannel retailer, this amount of deliveries is easily achieved. A disadvantage of the approximation is that the function obtained contains the square root of the number of stops, which is a decision variable in our model. However, the square root function is a concave function, and can be well approximated via a piecewise linear function, a standard procedure in convex/concave optimization. For a detailed description on how one can model piecewise linear objectives via liniar programming we refer to Hillier and Lieberman (1995) and Taha (2007). Details on the implementation of the approximation can be found in Section 6.

Additional to the costs to travel between stops, there are operational costs related to a stop, t_{opstop}.

Thus, the total costs related to the nr. of stops in a demand area l are estimated by

$$tc_l^{st}(n_{l,stops}) = t_{op\ stop} n_{l,stops} + t_l tc_v \sqrt{Area_l n_{l,stops}}, \quad (3)$$

Finally, we omit the costs of transporting the urgent items between stores and demand areas from the objective function, as we assume that all the urgent orders need to be satisfied and are delivered by the same vehicles that transport items from IDs to the demand area.

4.2. Constraints

Constraints related to demand/returns satisfaction and channel choice

$$\sum_{i \in I} y_{i,l} + \sum_{nc \in NC} w_{nc,l} + \sum_{nc \in NC} u_{nc,l} + u_{ec,l} = d_l + d_l^{urg}, \qquad l \in L \quad (4)$$

$$w_{nc,l} + u_{nc,l} \geqslant m_{nc} \gamma_{nc,l}^{nr}, \qquad nc \in PH,\ l \in L \quad (5)$$

$$w_{nc,l} + u_{nc,l} \geqslant m_{nc} \gamma_l^{vol}, \qquad nc \in APS,\ l \in L \quad (6)$$

$$w_{nc,l} + u_{nc,l} \leqslant \lambda_l Area_{nc} \gamma_{nc,l}^{nr}, \qquad nc \in NC \setminus \{home\},\ l \in L \quad (7)$$

$$\tilde{w}_{l,nc} + \tilde{u}_{l,nc} = \eta_l w_{nc,l}, \qquad l \in L \quad (8)$$

$$\tilde{y}_{l,i} + \tilde{u}_{l,ec} = \eta_l y_{i,l}, \qquad l \in L,\ i \in I \quad (9)$$

Constraints (4) ensure that the flow of satisfied and unsatisfied demand (from CDC or via new channels) equals the total demand in area l. Constraints (5) and (6) model the assumption that the more new channels people see, the more they will be tempted to use them (see assumption A4).

Constraints (7) imposes that the maximum demand served by one new channel (except the home delivery) is equal to $\lambda_l Area_{nc,l}$, where λ_l is the demand density in region l and $Area_{nc,l} = \pi WD_{nc,l}^2$ is the area served by the new channel (see assumptions A1-A3). Note that this demand model is obtained when the demand in region l is assumed to follow a spatial Poisson process with density λ_l. In this case, the right hand side represents the total demand in the service areas around the new installed channels.

Constraints (8) and (9) and impose that returns to CDCs and to new channels are a percentage of the direct demand delivered through the respective channel (see assumption A5).

Assignment constraints

$$\sum_{k \in K} \tau_{l,k} \leqslant 1 \qquad l \in L \quad (10)$$

$$\sum_{i \in I} \sigma_{l,i} \leqslant 1 \qquad l \in L \quad (11)$$

$$\tau_{l,k} \leqslant \beta_k \qquad l \in L,\ k \in K \quad (12)$$

$$\sigma_{l,i} \leqslant \alpha_i \qquad l \in L,\ i \in I \quad (13)$$

$$\rho_{k,i} \leqslant \alpha_i \qquad i \in I,\ k \in K \quad (14)$$

$$\sum_{i \in I} \rho_{k,i} = \beta_k \quad k \in K \quad (15)$$

$$z_{k,l} + \tilde{z}_{l,k} \leqslant d_l (1 + \eta_l) \tau_{l,k}, \qquad l \in L,\ k \in K \quad (16)$$

Constraints (10) and (11) ensure that a demand area $l \in L$ can only be served from at most one intermediate depot $k \in K$ (see assumption A13) and at most one CDC. Constraints (12) and 13 ensure that a demand region can be assigned only to an open ID and CDC respectively.

Further, constraints (14) impose that if ID k is assigned to CDC i, CDC i has to be opened. Constraints (15) ensure that each open ID has to be assigned to one CDC (see assumption A14).

Constraints (16) ensure that orders can be transported between area l and intermediary depot k only if l was assigned to k. In the same time, as each area is assigned to only one depot, it imposes that both deliveries and returns in an area should be done through the same ID (assumption A13).

Capacity constraints

$$\sum_{k \in K} (x_{i,k} + \tilde{x}_{k,i}) + \sum_{l \in L} (y_{i,l} + \tilde{y}_{l,i}) \leqslant cap_i \alpha_i, \qquad i \in I \quad (17)$$

$$x_{i,k} + \tilde{x}_{k,i} \leqslant cap_k \rho_{k,i}, \qquad k \in K,\ i \in I \quad (18)$$

$$y_{i,l} + \tilde{y}_{l,i} \leqslant cap_{l,ec} \sigma_{l,i}, \qquad i \in I,\ l \in L,\ ec \in EC \quad (19)$$

$$w_{nc,l} + \tilde{w}_{l,nc} \leqslant cap_{nc,l} \gamma_{nc,l}^{nr}, \qquad nc \in NC \setminus \{APS\},\ l \in L \quad (20)$$

$$w_{APS,l} + \tilde{w}_{l,APS} \leqslant cap_{APS,l} \gamma_l^{vol}, \qquad l \in L \quad (21)$$

Constraints (17)–(21) enforce that the total flow through each *open* facility or new channel does not exceed its processing capacity. Observe that for APSs the volume installed at each location is considered. Finally, constraints (18) ensure that orders can be transported from ID k to CDC i only if k is assigned to i. This constraint also imposes that deliveries and returns to/from an ID are processed at the same CDC (assumption A14).

Flow Constraints

$$\sum_{i \in I} x_{i,k} = \sum_{l \in L} z_{k,l}, \qquad k \in K \quad (22)$$

$$\sum_{i \in I} \tilde{x}_{k,i} = \sum_{l \in L} \tilde{z}_{l,k}, \qquad k \in K \quad (23)$$

$$\sum_{nc \in NC} w_{nc,l} = \sum_{k \in K} z_{k,l} + d_l^{urg}, \qquad l \in L \quad (24)$$

$$\sum_{nc \in NC} \tilde{w}_{l,nc} = \sum_{k \in K} \tilde{z}_{l,k} + \tilde{d}_l^{urg}, \qquad l \in L \quad (25)$$

Constraints (22) and (23) are the flow conservation constraint for intermediate depots, while constraints (24) are flow conservation constraints for deliveries through existing channels in a demand area l (the flow of direct deliveries from CDC to area l is equal to the flow of items delivered through existing channels in area l). Constraints (25) have a similar interpretation for returns. Constraints (24) and (25) take care that the flow of orders/returns delivered (picked up) through *new channels* is equal to the flow of orders/returns transported between an ID and a demand area plus the urgent orders (returns). Note that the urgent deliveries are included in the variables $w_{c,l}$ and the returns related to urgent deliveries are included in $\tilde{w}_{l,c}$. Thus, they must be satisfied and are considered as occupying capacity in the new and existing channels through (20) and (21).

Bounds on variables

$$\gamma_{nc,l}^{nr} \leqslant e_{nc,l}, \qquad nc \in NC,\ l \in L \quad (26)$$

$$\gamma_{APS,l}^{nr} \leqslant \gamma_l^{vol}, \qquad l \in L \quad (27)$$

$$\gamma^{nr}_{APS,l} \leqslant \frac{d_l}{Area_{APS,l}}, \qquad l \in L \tag{28}$$

$$\sum\nolimits_{i \in CDC} (y_{i,l} + \tilde{y}_{l,i}) \geqslant LB_l, \qquad l \in L \tag{29}$$

$$x_{i,k}, y_{i,l}, z_{k,l}, w_{nc,l}, u_{c,l}, \gamma_{nc,l} \geqslant 0, i \in I, k \in K, nc \in \text{NC}, c \in C, l \in L \tag{30}$$

$$\tilde{x}_{k,i}, \tilde{y}_{l,i}, \tilde{z}_{l,k}, \tilde{w}_{l,nc}, \tilde{u}_{l,c} \geqslant 0, i \in I, k \in K, nc \in NC, c \in C, l \in L \tag{31}$$

$$\alpha_i, \beta_k, \rho_{k,i}, \sigma_{i,l}, \tau_{l,k} \in \{0, 1\}, \qquad i \in I, k \in K, l \in L \tag{32}$$

$$\gamma^{nr}_{c,l}, \gamma^{vol}_l \in \mathbf{Z}_+, \qquad c \in NC, l \in L \tag{33}$$

Constraints (26) give an upperbound on the number of new channels that can be built in area l, while constraint (27) ensures that the number of locations where APSs are opened cannot exceed the number of APSs installed. Recall that by assumption (A7), more APSs can be installed at one location. Constraints (28) limit the number of locations where APSs can be installed, based on the maximum area an APS can serve (see assumption (A8)). Constraints (29) ensure that a minimum flow of orders is sent through existing channels (see also assumption (A11)). This constraint is necessary as the carriers are usually interested in maintaining the existing channels operational. Constraints (30)–(33) ensure that variables are positive or integer.

The final MIP-OLN model minimizes the objective function described in (4.1) subject to constraints (4)–(33).

Remarks on the model

1. Note that if $\sum_{nc,l} w_{nc,l} > d^{urg}_l$, or, equivalently, if there are regular orders to be delivered via new channels, $\sum_k z_{k,l} > 0$ by constraint (24). Constraint (12) then ensures that there should be an open ID.
2. Constraints (16) and (10) ensure that both direct deliveries and returns from an area are processed by the same ID. Note that a similar condition is imposed for the deliveries/returns through a CDC, by constraints (19) and (11).
3. Assuming one predicts the demand based on a regression model, one would expect an equality sign in constraints (5) and (6). However, if an equality sign was used when demand exceeds capacity, a too large number of facilities will be opened.
4. The model assumes the same percentage of returns in all the area. Clearly, different percentages can be accommodated by making the parameter η dependent on l. Moreover, the model assumes that demand and returns are expected values, hence can be fractional. If this is not desired, the percentage of returns should be area dependent and defined as r^*_l/d^*_l, where d^*_l and r^*_l are the rounded values of demand and returns.

5. MIP-based heuristic

In many cases, the processing costs at different CDCs or different IDs is the same. In these situations, the processing costs play an important role in the choice of whether a product is delivered through an existing or through an intermediary channel, but not in the choice of the specific ID or CDC. In the same way, transportation between echelons is usually done with the same type of vehicle. Hence, transportation costs that are related to the vehicle usage and not to distance, are independent on the exact location of facilities. Also, observe that, outside a demand region l, the information on which new channel will be used to deliver an order is irrelevant, as all the new channels in a region share the same delivery path (CDC and ID). This justifies analyzing the problem at hand under the following assumptions:

A15. The processing costs at all *CDCs* are equal to pr_{CDC}

A16. The processing costs at all *IDs* are equal to pr_{ID}

A17. Same type of vehicle is used for transportation between facilities of same type (the types are *CDCs*, *IDs*, existing channel and new channel).

Before describing the heuristic in detail, we will discuss the consequence of the assumptions (A15)–(A17) on the optimization. Based on (22) and (24) we conclude that

$$\sum\nolimits_{k \in K, i \in I} x_{i,k} = \sum\nolimits_{k \in K, l \in L} z_{k,l} = \sum\nolimits_{l \in L, nc \in NC} w_{nc,l} - \sum\nolimits_{l \in L} d^{urg}_l \tag{34}$$

A similar relationship can be written for the returns. Hence, under assumptions (A15) and (A16), the processing costs can be rewritten as

$$PR = (pr_{CDC} + pr_{ID})(\sum\nolimits_{l \in L, nc \in NC} (w_{nc,l} + \tilde{w}_{l,nc}) - d^{urg}_l - r^{urg}_l) + \sum\nolimits_{l \in L, nc \in NC} p_{nc}(w_{nc,l} + \tilde{w}_{l,nc}) + p_{ec} \sum\nolimits_{i \in I, l \in L} (y_{i,l} + \tilde{y}_{l,i}).$$

Notice further that for each $l \in L$, the whole flow $y_{i,l}$ will be delivered from a single CDC, and as processing costs are equal at all CDCs, it is not essential which CDC is chosen in the calculation of the processing costs. Furthermore, notice that the processing costs can be written as $PR = \sum_{l \in L} PR_l$, where

$$PR_l = (pr_{CDC} + pr_{ID})(\sum\nolimits_{nc \in NC} (w_{nc,l} + \tilde{w}_{l,nc}) - d^{urg}_l - r^{urg}_l) + \sum\nolimits_{nc \in NC} p_{nc}(w_{nc,l} + \tilde{w}_{l,nc}) + p_{ec}(y_l + \tilde{y}_l).$$

where y_l and $\tilde{y}_l$ represent the orders and returns through the existing channel.

For the heuristic, it is also convenient to split the transportation costs between any origin and destination into costs independent on the travel time and dependent only on the vehicle used, and costs that depend on the travel time (see Section 6 for more details). More precisely, assumption (A17) translates into following costs structure:

$$\begin{aligned} tc_{i,k} &= tc_{CDC,ID} + tc(t_{i,k}), \text{ for } i \in CDC, k \in K \\ tc_{i,l} &= tc_{CDC,L} + tc(t_{i,l}), \text{ for } i \in CDC, l \in L \\ tc_{k,l} &= tc_{ID,L} + tc(t_{k,l}), \text{ for } k \in K, l \in L \\ tc^{in}_l &= tc(veh_{in}), l \in L \end{aligned}$$

where $tc_{A,B}$ are the time independent costs between the echelons A and B and $t_{a,b}$ indicates the time needed to travel between locations $a \in A$ and $b \in B$. As before, we will denote the costs related to returns between a and b by $\tilde{tc}_{a,b}$. Note that assuming a cost component that depends only on the echelons and not on the exact origin-destination is not restrictive, as many companies use the same vehicle type between any points in two different echelons. Based on (34) and the similar relationship for returns, the time independent component of the transportation costs can be written as $TR = \sum_l TR_l$, where

$$\begin{aligned} TR_l = &(tc_{CDC,ID} + tc_{ID,L})(\sum\nolimits_{nc \in NC} (w_{nc,l} - d^{urg}_l) \\ &+ (\tilde{tc}_{ID,CDC} + \tilde{tc}_{L,ID})(\sum\nolimits_{nc \in NC} \tilde{w}_{l,nc} - r^{urg}_l) \\ &+ tc^{in}_l (\sum\nolimits_{nc \in NC} w_{nc,l} + \tilde{w}_{l,nc}) + tc_{CDC,L} y_l + \tilde{tc}_{L,CDC} \tilde{y}_l. \end{aligned}$$

The fact that both the processing and transportation costs have an important component that is only related to the delivery area, suggests that under assumptions (A14)–(A16), one could find a good approximation of the flow through existing and new channels by solving a separate optimization for each area. The main reason why this decomposition is successful is that, with the exception of the distance related travel costs, most of costs in other echelons can be taken into account. This will be done in the first phase of the optimization. Once the flow through existing and new channels in each area is estimated, in Phase 2 we formulate another global optimization, for deciding the CDCs and IDs to open. Note that in taking this decision, the exact flow through each particular new channel, or the exact location of the new channels in demand regions is not important. As the travel inside a particular area is not taken into account, this optimization is also much faster than the original one. Finally, in the third phase, we resolve the optimization in each area, to decide the capacity in new channels that needs to be installed, based on the assignment of areas to CDCs and IDs and the flows decided in Phase 2.

Next we describe in detail each phase.

5.1. Heuristic Phase 1

For each region l, we solve a separate MIP. All the variables have the same interpretation as in Section 4except of y_l and $\tilde{y}_l$, which represent the flow of direct delivery and returns through existing channels.

Objective function Phase 1

$$\sum_{c\in PH,l\in L} f_{c,l}\gamma_{c,l}^{nr} + f_{APS,l}\gamma_l^{vol} + PR_l + TR_l$$
$$+ \text{tc}_{st}(\sum_{nc\in \text{NC}} \gamma_{nc,l}^{nr} + w_{home,l} + \tilde{w}_{l,home}) + pen\sum_{c\in C}(u_{c,l} + \tilde{u}_{l,c})$$
$$+ \sum_{nc\in \text{NC}} v_{nc}(w_{nc,l} + \tilde{w}_{l,nc})$$

The first two terms of the objective represent the costs for opening/installing new channels, the third represents the processing costs related to the demand in region l, the fourth term the transportation costs related to vehicle usage, the fifth represents the transportation costs inside area l and the last two terms represent the penalty costs for unsatisfied demand and the discounts for using new channels, respectively.

Constraints Phase 1

Constraints related to demand satisfaction and channel choice

Constraints (5)–(9)

$$y_l + \sum_{nc\in NC} w_{nc,l} + \sum_{nc\in NC} u_{nc,l} + u_{ec,l} = d_l + d_l^{urg}, \tag{35}$$

$$\sum_{nc\in NC} w_{nc,l} \geqslant d_l^{urg} \tag{36}$$

$$\sum_{nc\in NC} \tilde{w}_{l,nc} \geqslant \tilde{d}_l^{urg} \tag{37}$$

Capacity constraints

Constraints (20)–(21)

$$y_l + \tilde{y}_l \leqslant cap_{l,ec} \tag{38}$$

Bounds on the number of channels

Constraints (26)–(28)

$$y_l + \tilde{y}_l \geqslant LB_l \tag{39}$$

Bounds on variables

$$w_{nc,l},\ u_{c,l},\ \gamma_{nc,l},\ \tilde{w}_{l,nc},\ \tilde{u}_{l,c},\ \gamma_l^{vol} \geqslant 0, \qquad nc\in NC,\ c\in C \tag{40}$$

$$\gamma_{c,l}^{nr} \in \mathbf{Z}_+, \qquad c\in NC \tag{41}$$

Constraint (35) ensure that the demand in area l is satisfied, while constraints (36) and (37) ensure that the urgent direct deliveries and the corresponding returns are satisfied. Observe that in the mathematical program in Section 4, these two constraints were not necessary, as they are a consequence of constraints (24) and (25). However, they are needed here, to ensure that urgent deliveries and the corresponding returns are fulfilled.

Constraints (20) and (21) ensure that the flow through new channels does not exceed the installed capacity, while constraint (38) insures that the capacity of the existing channel is not exceeded.

Constraints (39) impose a lower bound on the flow through existing channels, to ensure efficiency of these channels. Constraints (40)and (41) are the positivity and integrality constraints imposed on the variables.

The output of Phase 1 consists, for each region l, in a division of demand into demand that is satisfied through existing channels, $d_l^{EC} = y_l$ and demand that has to be satisfied through new channels, $d_l^{NC} = \sum_{nc} w_{nc,l} - d_l^{urg}$.

5.2. Heuristic Phase 2

In this phase, the heuristic decides the CDCs and IDs to be opened such that, $\sum_{l\in L} d_l^{EC}$ and $\sum_{l\in L} d_l^{NC}$ are satisfied at minimal cost. We will allow the program to reject demand if it is more profitable to pay penalties than open new *IDs* or *CDC*. The MIP program solved in this stage is described in detail below.

The following new variables will be used:

w_l^{NC}: direct flow to be satisfied through new channels in region l
$\tilde{w}_l^{NC}$: returns to be satisfied through new channels in region l
u_l^{NC}: unsatisfied demand through new channels in region l
$\tilde{u}_l^{NC}$ unsatisfied returns through new channels in region l

All the other parameters and variables have the same interpretation as in the MIP in Section 4. In particular, $y_{i,l}$ and $\tilde{y}_{l,i}$ represent the direct flow through the existing channel in region l satisfied from CDC i and the returns flow from the existing channel in region l to CDC i.

Objective function Phase 2

Fixed cost of facility operations

$$\sum_{i\in I} f_i\alpha_i + \sum_{k\in K} f_k\beta_k +$$

Processing costs at CDCs and IDs

$$p_{CDC}\sum_{i\in I}\left[\sum_{l\in L}(y_{i,l} + \tilde{y}_{l,i})\right] + (p_{CDC} + p_{ID})\sum_{k\in K}(x_{i,k} + \tilde{x}_{k,i})] +$$

Processing costs at existing and new channels

$$\sum_{l\in L} p_{nc}(w_l^{NC} + \tilde{w}_l^{NC}) + \sum_{l\in L,i\in I} p_{ec}(y_{i,l} + \tilde{y}_{l,i}) +$$

Penalty costs and channel discounts

$$pen\sum_{l\in L}(u_l^{NC} + \tilde{u}_l^{NC} + u_{ec,l} + \tilde{u}_{l,ec})$$

Transportation costs between echelons

$$\sum_{i\in I,k\in K}[tc_{i,k}x_{i,k} + \tilde{tc}_{k,i}\tilde{x}_{k,i}] + \sum_{k\in K,l\in L}(tc_{k,l}z_{k,l} + \tilde{tc}_{l,k}\tilde{z}_{l,k})$$
$$+ \sum_{i\in I,l\in L}(tc_{i,l}y_{i,l} + \tilde{tc}_{l,i}\tilde{y}_{l,i}) +$$

Transportation costs inside area l, independent on the nr. of stops

$$\sum_{l\in L} tc_l^{in}(w_l^{NC} + \tilde{w}_l^{NC}),$$

Constraints Phase 2

Demand satisfaction and returns

$$w_l^{NC} + u_l^{NC} = d_l^{NC}, \qquad l\in L \tag{42}$$

$$\sum_{i\in I} y_{i,l} + u_{ec,l} = d_l^{EC}, \qquad l\in L \tag{43}$$

$$\tilde{w}_l^{NC} + \tilde{u}_l^{NC} = \eta_l w_l^{NC}, \qquad l\in L \tag{44}$$

$$\tilde{y}_{l,i} + \tilde{u}_{l,ec} = \eta_l y_{i,l} \qquad l\in L,\ i\in I \tag{45}$$

Assignment constraints

Constraints (10)–(16)

Capacity constraints

Constraints (17)–(19)

Flow conservation constraints

Constraints (22)–(23)

$$w_l^{NC} = \sum_{k\in K} z_{k,l}, \qquad l\in L \tag{46}$$

$$\tilde{w}_l^{NC} = \sum_{k\in K} \tilde{z}_{l,k}, \qquad l\in L \tag{47}$$

Constraints (42) and (43) ensure that, in each region, the flow through new channels and existing channels equals d_l^{NC} and d_l^{EC} respectively. Constraints (44) and (45) define the returns as percentage of direct demand. Assignment constraints (10)–(16) ensure that demand in each area is either served through existing channels directly from CDC or via an ID and that each open ID is assigned to one CDC. As in MIP-OLN, constraints (16) and (10) ensure that deliveries and returns to/from an area l are processed by the same ID. Constraints (17)–(19) ensure that capacity at CDC, ID and the existing channel is not exceeded, while constraints (11) and (19) impose that the direct orders and returns through existing channels are processed at the same CDC. Constraints (22) and (23) are flow conservation constraints for direct flow and returns at IDs. Finally, constraints (46) and (47) ensure that the flow of satisfied orders and returns at area level is transported to IDs.

The output of Phase 2 consists in a set of open IDs and CDCs and an assignment of demand regions to IDs and CDCs, as well as an assignment of open IDs to open CDC.

5.3. Heuristic Phase 3

For each $l \in L$ for which an intermediary depot k exists such that $\tau_{k,l} = 1$, let $ID(l) = k$. Similarly, if there exists an $i \in I$ such that $\sigma_{l,i} = 1$ at the end of phase 2, denote i by $CDC(l)$. Finally, let $CDC(k)$ be the CDC to which an open ID k is assigned. Note that for each l, if $ID(l)$ and $CDC(l)$ exist, their uniqueness is implied by constraints (10) and (11). The uniqueness of $CDC(k)$ is implied by constraints (15).

Moreover, let $D(k) = \{l \in L: ID(l) = k \text{and} d_l^{NC} > 0 0\}$ for each open ID k. If k is not open, $D(k) = \emptyset$. Remark that in phase 2, demand regions l with d_l^{NC} can be assigned to an ID, although they will not make use of it as at the end of Phase 1, $d_l^{NC} = 0$, that is, only urgent demand, which does not pass through IDs, will be delivered through new channels.

The goal of Phase 3 is to decide how many new channels to open in each area and to redistribute the flow $x_{CDC(k)k}$, for each open ID k through the new channels in the areas k serves, namely $D(k)$. Note that this problem differs from the optimization in Phase 1 in two important aspects: first, in Phase 1, the goal was to split demand between the existing and the new channels, while in Phase 3 we only look at the new channels; second, in Phase 1, the assignment of demand regions to IDs and CDCs was not known, while in Phase 3 the algorithm can take into account the transportation costs to the assigned IDs and CDCs.

The parameters and variables used have the same meaning as in the original MIP. For each ID k that is opened in Phase 2, the following optimization problem is solved.

Objective function Phase 3

$$\sum_{c\in PH, l\in D(k)} f_{c,l}\gamma_{c,l}^{nr} + f_{APS,l}\gamma_l^{vol} + PR_l + pen\sum_{c\in NC}(u_{c,l} + \tilde{u}_{l,c}) + \sum_{nc\in NC} v_{nc}(w_{nc,l} + \tilde{w}_{l,nc}) + TR_l + tc_{st}\left(\sum_{nc\in NC}\gamma_{nc,l}^{nr} + w_{home,l} + \tilde{w}_{l,home}\right) + \sum_{l\in D(k)}(tc(t_{CDC(k),k}) + tc(t_{k,l}))\sum_{nc\in NC} w_{nc,l} - d_l^{urg}) + \sum_{l\in D(k)}\tilde{tc}(t_{k,CDC(k)}) + \tilde{tc}(t_{l,k})\left(\sum_{nc\in NC}\tilde{w}_{nc,l} - \tilde{d}_l^{urg}\right)$$

The first part of the objective function is identical to the objective in Phase 1. The last two sums represent the transportation costs to the assigned IDs and CDCs. Variables that are fixed in Phase 1 or Phase 2 will have an upper-script indicating the phase when their value is fixed, i.e., $y_{i,l}^2$ indicates the value of the variable $y_{i,l}$ at the end of phase 2.

Constraints Phase 3

Demand and return satisfaction constraints

Constraints (5)–(8)

$$\sum_{nc\in NC}(w_{nc,l} + u_{nc,l}) + y_{CDC(l),l}^2 + u_{ec,l}^2 + u_{ec,l}^1 = d_l + d_l^{urg}, \quad l \in D(k) \tag{48}$$

$$\sum_{nc\in NC} w_{nc,l} \geqslant d_l^{urg} \tag{49}$$

$$\sum_{nc\in NC}\tilde{w}_{l,nc} \geqslant \tilde{d}_l^{urg} \tag{50}$$

$$\sum_{l\in D(k), nc\in NC}(w_{nc,l} + \tilde{w}_{l,nc}) \leqslant \sum_{l\in D(k)}(z_{kl}^2 + \tilde{z}_{lk}^2 + d_l^{urg} + r_l^{urg}) \tag{51}$$

Capacity constraints

Constraints (20)–(21)

Bounds on the number of channels

Constraints (26)–(28)

$$w_{nc,l}, u_{nc,l}, \gamma_{nc,l}, \tilde{w}_{l,nc}, \tilde{u}_{l,nc}, \gamma_l^{vol} \geqslant 0, \qquad nc \in NC \tag{52}$$

$$\gamma_{c,l}^{nr} \in \mathbf{Z}_+, \qquad c \in NC \tag{53}$$

Constraints (48) ensure that the total demand in each region is covered by the flows through existing or new channels or by the flow of unsatisfied demand. Constraints (49) and (50) ensure that the urgent demand and corresponding returns are satisfied. Constraints (51) ensure that the demand and returns satisfied in the regions in $D(k)$ does not exceed the total flow through k obtained in Phase 2. As before, constraints (20) and (21) ensure that the capacity of the new channels installed is not exceeded. The bounds on the varaibles are identical to the bounds in the original MIP.

The output of Phase 3 consists in the number of new channels to be opened in each area $l \in \cup_{k\in K} D(k)$ and the flow through each channel type.

At the end of the heuristic, we fix the variables in MIP-ONL to the values obtained at the end of the phase indicated by the superscript, as indicated below. We used $\mathbf{a}$ to indicate a vector.

Variables related to opening facilities and assignment variables:

$$(\alpha, \beta, \tau, \rho, \sigma) = (\boldsymbol{\alpha}^2, \boldsymbol{\beta}^2, \boldsymbol{\tau}^2, \boldsymbol{\rho}^2, \boldsymbol{\sigma}^2) \tag{54}$$

Variables related to flow through existing channels

$$(\mathbf{y}, \tilde{\mathbf{y}}) = (\mathbf{y}^2, \tilde{\mathbf{y}}^2) \tag{55}$$

Variables related to capacity offered in new channels

$$(\gamma^{\mathbf{nr}}, \gamma^{\mathbf{vol}}) = (\gamma^{\mathbf{nr},3}, \gamma^{\mathbf{vol},3}) \tag{56}$$

Variables related to deliveries through open channels

$$w_{nc,l} = w_{nc,l}^3, \qquad l \in \cup_{k\in K} D(k), nc \in NC \tag{57}$$

$$w_{nc,l} = w_{nc,l}^1, \qquad l \in L\setminus(\cup_{k\in K} D(k)), nc \in NC \tag{58}$$

$$z_{k,l} = \sum_{nc\in NC} w_{nc,l}^3 - d_l^{urg}, \qquad k \in K, l \in D(k) \tag{59}$$

$$x_{i,k} = \sum_{l\in D(k)}\left(\sum_{nc\in NC} w_{nc,l}^3 - d_l^{urg}\right) \text{ for any open } k \in K \text{and } i = CDC(k) \tag{60}$$

Variables related to returns

$$\tilde{w}_{l,nc} = \tilde{w}_{l,nc}^3, \qquad l \in \cup_{k\in K} D(k), nc \in NC \tag{61}$$

$$\tilde{w}_{l,nc} = \tilde{w}_{l,nc}^1, \qquad l \in L\setminus(\cup_{k\in K} D(k)), nc \in NC \tag{62}$$

$$\tilde{z}_{l,k} = \sum_{nc\in NC}\tilde{w}_{l,nc}^3 - \tilde{d}_l^{urg}, \qquad k \in K, l \in D(k) \tag{63}$$

$$\tilde{x}_{k,i} = \sum_{l\in D(k)}\left(\sum_{nc\in NC}\tilde{w}_{l,nc}^3 - \tilde{d}_l^{urg}\right), \text{ for any open } k \in ID, i = CDC(k) \tag{64}$$

Variables related to unsatisfied demand and returns

$$u_{nc,l} = u_{nc,l}^3, \qquad l \in \cup_{k\in K} D(k), nc \in NC \tag{65}$$

$$u_{nc,l} = u_{nc,l}^1, \qquad l \in L\setminus(\cup_{k\in K} D(k)), nc \in NC \tag{66}$$

$$\tilde{u}_{l,nc} = \tilde{u}_{l,nc}^3, \qquad l \in \cup_{k\in K} D(k), nc \in NC \tag{67}$$

$$\tilde{u}_{l,nc} = \tilde{u}_{l,nc}^1, \qquad l \in L\setminus(\cup_{k\in K} D(k)), nc \in NC \tag{68}$$

$$u_{ec,l} = u_{ec,l}^1 + u_{ec,l}^2, \qquad l \in L \tag{69}$$

$$\tilde{u}_{l,ec} = \tilde{u}_{l,ec}^1 + u_{l,ec}^2, \qquad l \in L \tag{70}$$

All the remaining variables are set to zero.

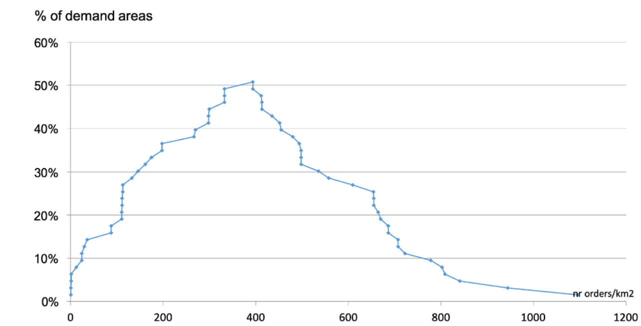


Fig. 2. Distribution of demand density.

Proposition The MIP-heuristic returns a feasible solution to MIP-OLN.

Proof We will show that the vector defined by (57)–(70), denoted farther by SOL, gives a feasible solution to MIP-OLN.

Constraints related to demand/returns satisfaction and channel choice

By constraints (48), for every $l \in \cup_{k\in K} D(k)$,

$$y^2_{CDC(l),l} + \sum\nolimits_{nc\in NC} (w^3_{nc,l} + u^3_{nc,l}) + u^2_{ec,l} + u^1_{ec,l} = d_l + d_l^{urg}.$$

It is easy to see that SOL satisfies constraint (4) by noting that by its definition, $y_{i,l} = y^2_{i,l}$, $w_{nc,l} = w^3_{nc,l}$, $u_{nc,l} = u^3_{nc,l}$ and $u_{ec,l} = u^2_{ec,l} + u^1_{ec,l}$ (see (55), (57), (65), and (69)).

For $l \in L\backslash(\cup_{k\in K} D(k))$, constraint (35) implies

$$y^1_l + \sum\nolimits_{nc\in NC} w^1_{nc,l} + \sum\nolimits_{nc\in NC} u^1_{nc,l} + u^1_{ec,l} = d_l + d_l^{urg}. \tag{71}$$

Constraints (43) and the definition of d_l^{EC} imply that

$$\sum_{i\in I} y^2_{i,l} + u^2_{ec,l} = y^1_l. \tag{72}$$

Combining (71) and (72), we obtain

$$y^2_{i,l} + u^2_{ec,l} + \sum\nolimits_{nc\in NC} w^1_{nc,l} + \sum\nolimits_{nc\in NC} u^1_{nc,l} + u^1_{ec,l} = d_l + d_l^{urg}. \tag{73}$$

It is now easy to see that SOL also satisfies (4) for $l \in \cup_{k\in K} D(k)$, by observing that $y_{i,l} = y^2_{i,l}$, $w_{nc,l} = w^1_{nc,l}$, $u_{nc,l} = u^1_{nc,l}$ and $u_{ec,l} = u^2_{ec,l} + u^1_{ec,l}$ by (55), (58), (66), and (69).

As $w_{nc,l}$ and $u_{nc,l}$ are fixed in Phase 3 for $l \in \cup_{k\in K} D(k)$ and in Phase 1 for $l \in L\backslash \cup_{k\in K} D(k)$, they satisfy constraints (5)–(8), which are imposed in the MIP solved in both phases. Similarly, as $\widetilde{y}_{l,i}$ and $\widetilde{u}_{l,ec}$ are fixed in Phase 2, they satisfy constraint (9).

Assignment constraints

To verify the assignment constraints, there is no need to distinguish between the regions in $\cup_{k\in K} D(k)$ and the others, as all the regions are included in Phase 2.

As $(\alpha, \beta, \tau, \rho, \sigma)$ are fixed at the end of Phase 2, it must satisfy (10)–(15) which are imposed in this phase.

To verify (16), note that $\tau_{l,k} = \tau^2_{l,k} = 0$ for any $k \neq ID(l)$. For all $l \notin D(k)$, $z_{k,l} = 0$ and $\widetilde{z}_{l,k} = 0$, by the definition of these variables (see (59) and (63)). As $\{l: ID(l) = k\} \subseteq D(k)$, it follows that for $l \in L$ and $k \in K$ such that $k \neq ID(l)$, $z_{k,l} = 0$ and $\widetilde{z}_{l,k} = 0$ and (16) is satisfied. For $l \in D(k)$, $z_{k,l} \leqslant d_l$, since $z_{k,l} = \sum_{nc\in NC} w^3_{nc,l} - d_l^{urg}$ by definition and $\sum_{nc\in NC} w^3_{nc,l} \leqslant d_l + d_l^{urg}$ by (48). Moreover, by constraints (8), $\widetilde{w}^3_{l,nc} \leqslant \eta w^3_{nc,l}$. By combining this relation with the definition of $\widetilde{z}_{l,k}$,

$$\widetilde{z}_{l,k} \leqslant \eta z_{k,l} + \eta d_l^{urg} - r_l^{urg} \leqslant \eta d_l,$$

where we have used that $\widetilde{d}_l^{urg} = \eta d_l^{urg}$. Hence, $z_{k,l} + \widetilde{z}_{l,k} \leqslant (1+\eta) d_l \tau_{l,k}$.

Capacity constraints

To verify (17), for each $i \in CDC$, denote by $ID(i) = \{k \in K: CDC(k) = i\}$ and by $L(i) = \{l \in L: CDC(l) = i\}$. For $k \notin ID(i)$, $x_{ik} = \widetilde{x}_{k,i} = 0$ and for $l \notin L(i)$, $y_{i,l} = \widetilde{y}_{l,i} = 0$. Recall that by (60)–(64), for each $k \in ID(i)$, we have that

$$\begin{aligned} x_{i,k} + \widetilde{x}_{k,i} &= \sum\nolimits_{l\in D(k)} \left(\sum\nolimits_{nc\in NC} w^3_{nc,l} - d_l^{urg}\right) \\ &\quad + \sum\nolimits_{l\in D(k)} \left(\sum\nolimits_{nc\in NC} \widetilde{w}^3_{l,nc} - r_l^{urg}\right) \\ &\leqslant \sum\nolimits_{l\in D(k)} (z^2_{kl} + \widetilde{z}^2_{lk}) \leqslant \sum\nolimits_{i\in I} x^2_{i,k} + \sum\nolimits_{i\in I} \widetilde{x}^2_{k,i} \end{aligned} \tag{74}$$

where for the second inequality we used constraint (51) in Phase 3 and for the last inequality we used constraints (22) and (23) in Phase 2. Finally, by using (17) in Phase 2, we conclude that

$$\begin{aligned} \sum\nolimits_{k\in ID(i)} (x_{i,k} + \widetilde{x}_{k,i}) + \sum\nolimits_{l\in L(i)} (y_{i,l} + \widetilde{y}_{l,i}) &\leqslant \sum\nolimits_{i\in I} (x^2_{i,k} + \widetilde{x}^2_{k,i}) + \sum\nolimits_{l\in L(i)} (y_{i,l} + \widetilde{y}_{l,i}) \\ &\leqslant cap_i \alpha_i. \end{aligned}$$

This implies that the constructed solution satisfies (17). Moreover, (74) and constraints (18) imposed in Phase 2, imply that $x_{i,k} + \widetilde{x}_{k,i} \leqslant cap_k \rho_{k,i}$. Hence, SOL also satisfies (18).

As the variables $y_{i,l}$, $\widetilde{y}_{l,i}$ and $\sigma_{l,i}$ are fixed in Phase 2, where constraint (19) is imposed, the final solution of the heuristic also satisfies this constraint. Similarly, constraints (20) and (21) are satisfied, as the variables $w_{nc,l}$, $\widetilde{w}_{l,nc}$, $\gamma^{nr}_{nc,l}$ and γ^{vol}_l are fixed Phase 1 for regions $l \in L\backslash(\cup_{k\in K} D(k))$ and in Phase 3 for $l \in \cup_{k\in K} D(k)$, where these constraints are also imposed.

Flow conservation constraints

Definitions (59) and (60) combined with the fact that, for all $l \in L\backslash(\cup_{k\in K} D(k))$, $z_{k,l} = 0$ and for all $k \in K$ and $i \in I$, with $i \neq CDC(k)$, $x_{i,k} = 0$, imply that SOL satisfies (22). Furthermore, by the definition of $z_{k,l}$ and $\widetilde{z}_{l,k}$, SOL also satisfies constraints (24) and (25) for $l \in \cup_{k\in K} D(k)$. By the definition of $D(k)$ in Phase 3, $d_l^{NC} = 0$, for $l \in L\backslash(\cup_{k\in K} D(k))$. Equivalently, $\sum_{nc} w^1_{nc,l} = d_l^{urg}$. As $z_{k,l} = 0$ for $l \in L\backslash(\cup_{k\in K} D(k))$, it follows that (24) is also satisfied by $l \in L\backslash(\cup_{k\in K} D(k))$. Similar arguments can be used to show that SOL also satisfies (25) for $l \in L\backslash(\cup_{k\in K} D(k))$.

Bounds on the variables The bounds imposed in Phase 2 and Phase 3 ensure that the vector defined by (54)–(70) satisfies (26)–(33). Note that constraints (49) and (50) in Phase 3 ensure that the variables $z_{k,l}$, $\widetilde{z}_{l,k}$, $x_{i,k}$ and $\widetilde{x}_{k,i}$ are positive. ■

6. Case study

In this section we apply the model in Section 4, illustrated in Fig. 1, on a case study related to a parcel carrier in Madrid, Spain. The city is divided into 63 demand areas, each corresponding to a postal code (see Fig. 6). The demand areas differ in demand density (defined as nr.orders/km^2) as shown in Fig. 2. There are 9 areas with very low demand density (below 40 orders/km^2), 16 areas with low density (between 40 and 290 orders/km^2), 19 areas with medium density (between 290 and 500 orders/km^2), 12 areas with high density (between 500 and 710 orders/km^2) and 7 areas with high density (between 710 and 1000 orders/km^2). The density of demand increases with the proximity to the city center.

Table 1
Main parameters related to facilities.

	CDC	ID	Existing Channel	APS	Store	Kiosk	Home
Fixed cost/day(€)	4247	439	158	5.10	4.6	1.78	0
Processing cost/order(€)	0.12	0.05	0.11	0.11	0.11	0.11	0
Capacity (orders or returns)	8000	7000	300	20	12	7	1
Discount(€/order or €/return)				0.1	0.1	0.1	0
Penalty (€/order or €/return)			3	3	3	3	3

Table 2
Parameters related to transportation costs.

	truck (CDC-ID)	van (ID-area l)	van (inside area l)
capacity of a vehicle (m3)	16.32	8.448	8.448
$t_{operations}$(min)	30	30	30
tc_v (€/min)	0.60	0.446	0.446
wage costs(Euro/min)	0.2	0.2	0.2

In each area, there is one *existing channel*, namely the parcel office corresponding to the postal code. We consider the possibility of installing 4 *new channel* types in each area: Automatic Parcel Station (APS), stores, kiosks and homes.

The carrier uses trucks to pick up online orders from distribution centers in the outskirts of Madrid (outside the ringroad). The distribution centers are the sites where retailers or their associated 3PLs fulfill online orders. Inside CDCs, parcels are sorted and consolidated by destination. Orders that have to be collected by consumers in the existing 63 parcel offices, are delivered directly by trucks. Orders that have to be delivered to Automated Parcel Stations (APS), stores, kiosks or consumers homes are sent by truck to intermediary depots (ID), inside the ring road of the city. Each intermediary depot covers a group of demand areas in the city. The transportation between ID's and demand areas is performed by vans. Urgent orders, that require immediate delivery, are fulfilled in retailer's urban stores, are picked up by the vans that travel between the IDs and demand areas and are delivered via the closest route. Commercial returns follow the same route in reverse direction.

The carrier is interested where to open CDC's, IDs and new channels such that the total costs of opening facilities and transportation costs, as given in Section 4.1 is minimized.

We next present the parameters of the basic scenario.

Parameters related to facilities in the basic scenario

All facilities of the same type are considered identical. The maximum number of CDCs that can be opened is 2 while the maximum number of intermediary depots is 8. The parameters related to the facilities and new channels are given in Table 1:

For an exact description on how the facility costs were obtained, we refer to the Appendix A. For stores and kiosks, information on the maximum number of channels to open has been collected from each postal area (see Appendix A). For APSs and home deliveries, the maximum capacity that can be opened in an area is not restricted (i.e., $e_{nc,l} = \infty$).

Parameters related to orders and demand in the basic scenario

The total demand is assumed to be equal to $D = 75000$ packages per day. The demand in area l is calculated as $\alpha_l * D$, where alpha is the percentage of population of Madrid living in area l. The percentages α_l were calculated based on the data regarding the population between 25 and 64 years old obtained from the Institute of Statistics of Madrid Community (Estadistica de Población de la Comunidad de Madrid, 1996). The value used for the demand per area can be found in the Appendix A.

Urgent demand in area l is calculated as $d_l^{urg} = 2\% d_l$, while the percentage returns in area l are equal to $\eta_l = 6.5\%$ of the demand. A similar relationship holds for urgent deliveries.

For each new channel, in the basic scenario, $m_{nc} = 0.935 * cap_{nc}$. The distance a customer is willing to walk to a new channel was estimated to be 420 m.

Parameters related to transportation costs in the basic scenario

We assume 2 type of vehicles: trucks, that transport order between CDC-ID and vans, for the transportation between ID and demand area, and inside demand area. The parameters related to the 2 types of vehicles are given in Table 2.

Next we explain how these parameters have been used to calculate the transportation costs.

All the transportation costs between two facilities in different echelons can be split into a component that is independent on the time travelled and one that depends on the time (distance) travelled between the specific locations. We use time instead of distance to calculate costs, due to the different speeds allowed in a city. For two echelons A and B, where $\{A, B\} \subset \{CDC, ID, L\}$, the travel time independent component is defined as $tc_{A,B} = q_{veh(A,B)} * t_{operations} * \text{wage costs}$, where $q_{veh(A,B)} = \frac{\text{average volume of an item}}{\text{capacity vehicle (A, B)}}$ represents the percentage of a vehicle used, $veh(A, B)$ is the vehicle used between echelons A and B and $t_{operations}$ is the time needed to load/unload a vehicle.

The time(distance) dependent component between two echelons is equal to $tc(t_{a,b}) = q_{veh} * t_{trip(a,b)} * tc_v$, where $t_{trip(a,b)}$ is the time needed for the trip between facilities a and b and tc_v is the cost per time unit related to the vehicle used. The costs in the reverse direction are calculated in the same way.

The transportation costs inside area have a component that is independent on the time travelled (in this case, on the number of stops made) and a component that depends on the number of stops. The transportation costs inside a region l, that are independent on the number of stops, are equal to $tc_l^{in} = q_{veh(A,B)} * t_{operations} * \text{wage costs}$.

The estimation of the transportation costs between stops was given in (3). A piecewise linear approximation of this function, with segments defined by 36 points between [0, 2300] was constructed via Gurobi. We chose for a large interval in order to capture the function more accurately – however, smaller intervals can be chosen depending on what is found a reasonable maximum number of stops in an area. The quality of the piecewise linear approximation is discussed in Appendix B. The operation costs per stop were calculated as $tc_{\text{opstop}} = \text{ctt}_{nc}\text{wage cost}$, where ctt_{nc} represents the average time needed to reach a channel once the van is stopped.

In calculating the time needed to travel between two points, we extracted the speed from Google (see Appendix A).

All the MIP programs have been solved with GUROBI 8.1.1 and all the experiments were run on an Intel Core i7-4770CPU, 3.4 GHz, 16 GB RAM.

We will focus our analysis on the direct flows, as the most predominant flows in the network. All the flows are rounded to the nearest integer.

Table 3
Total cost structure.

Costs	Cost Existing channels (€)	Costs New channels (€)	Percentage cost existing channels	Percentage cost new channels	Percentage rightarrow tal cost
Facility	18 152	16 063	24.46%	21.64%	46.10%
Processing	12 706	8 877	17.12%	11.96%	29.08%
Line haul transportation	1 057	7 351	1.42%	9.90%	11.33%
Transportation inside demad area	0	3 636	0	8.6%	4.9%
Penalty	10	83	0.01%	0.11%	0.12%
Discounts	0	6 291	0	8.47%	8.47%
Total costs	31925	42301	43%	57%	100%

Table 4
Structure of the optimal solution.

	Parcel offices	Home delivery	APS	Stores	Kiosks
New Facilities installed	0		156	40	39
Orders	16711	758	58405	484	177
Returns	1086	49	3796	31	12
Total flow	17797	807	62201	515	189
Total flow (%)	21.83%	0.99%	76.31%	0.63%	0.23%

6.1. Base case analysis

The total costs in the base case scenario amounts to 74225 €/day. Table 3 shows the distribution of the cost components in the optimal solution. The main component of the cost is the facility costs (46% of the total costs). As expected in an omnichannel environment, a large part of the costs (29.08%) are due to processing orders. The transportation costs are 16.23% of the total costs, 11.33% being incurred by long-haul transportation and 4.9% by transportation inside demand areas. New channels, mostly APS, account for 57% of the total costs and their total associated transportation costs are 14.8% of the total costs.

The structure of the network given by the optimal solution is given in Table 4. All the 63 existing channels are used as much as possible, as being the most efficient with respect to transportation costs, capacity and facility costs. The flow through the existing channels amounts at 21.83% of the total flow. From the new channels, APSs are preferred, due to the fact that more facilities can be installed at one location. APSs serve the largest percentage of orders and returns, namely 76.3%. As seen in Table 4, home deliveries account for a small percentage of the flows of orders and returns, namely only 1%. This is due to the higher transportation costs associated to home deliveries and the economy of scale offered by other channels. The small number of stores (40) and kiosks (39) used is due to small existing number of such facilities in demand areas. These results confirm the carrier's strategy to focus on installing APSs in Madrid and whole Spain.

One of the main factors influencing the flow structure per channel is the density of demand. The impact of order density on the usage of the different channels is shown in Fig. 3. As expected, when the number of orders per demand area is small (below 500/km^2), the existing channel is preferred. When the total number of orders exceeds the capacity of the existing channels, other channels start being used, hereby decreasing the percentage of flow through the existing channel.

Table 5
Parameter variation.

Parameter	Range multiplication factors	Increment
Demand	1–1.5	0.1
fixed costs	1–1.5	0.1
tc_v	1–1.5	0.1
WD	0.2–1.2	0.2
Speed	0.2–1	0.2
(m_{APS}, m_{stores}, m_{kiosks}, m_{home})	(0.2, 0.2, 0.2,1)	
	(0.2, 0.2, 0.8, 1)	
	(0.2, 0.8, 0.2, 1)	
	(0.8, 0.2, 0.2, 1)	
	(0.8, 0.2, 0.8, 1)	
	(0.8, 0.8, 0.2, 1)	
	(0.2, 0.8, 0.8, 1)	
	(0.8, 0.8, 0.8, 1)	
(m_{APS}, m_{stores}, m_{kiosks})[a]	1.2–1.6	0.2

[a] m_{home} = 1.

If the remaining demand to be delivered through other channels is very small, home delivery is preferred, until opening other new channels becomes efficient (this is the case for demand areas with less than 500/km^2). When the number of orders/demand area increases above 500/km^2, the usage of APSs increases. The areas with higher number of home deliveries (between 30% and 40%) are characterized by low demand density. In these cases, opening new channels does not result in economy of scale in deliveries.

6.2. Impact of problem parameters

To analyze the impact of different parameters, we have constructed 36 scenarios, by using the multiplication factors in Table 5.

Impact of changes in demand

To study the impact of demand variations, we have increased the average number of orders in the base scenario by a multiplication factors varying from 1 to 1.5 in steps of 0.1 (from 76,566 orders to 114,849 orders in steps of 7656 orders). For each scenario, the number of each type of facility opened and the flow through each facility type is given in Table 6.

As seen in Table 6, up to a demand multiplying factor of 1.2, the flow through all new channels increases. This is to be expected, as the existing channels are used at maximum capacity. For higher factors, the capacity of all the IDs is reached, hence demand cannot be satisfied anymore. Most of the demand will be served via APSs, except in regions

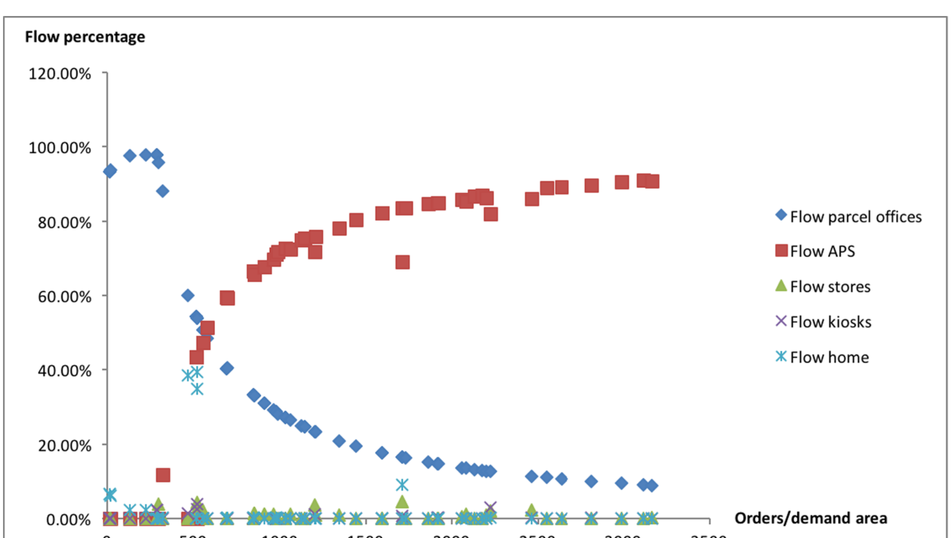


Fig. 3. Flow percentages in the basic scenario.

Table 6
Impact of demand increase on the number of facilities opened.

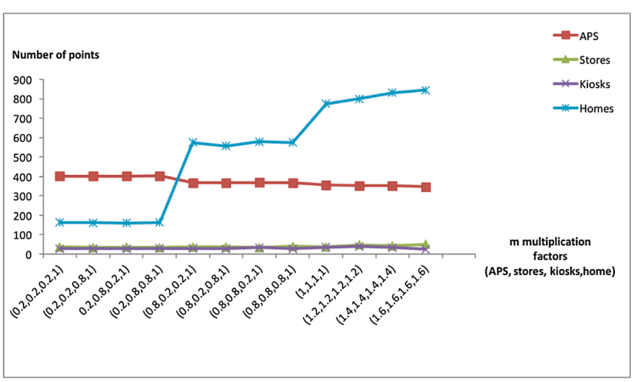
Demand factor	Nr. APS locations	Nr. stores	Nr. kiosks	Home deliv.	Flow APS	Flow stores	Flow kiosks
1	356	43	27	758	58 405	484	177
1.1	367	45	45	835	65 658	507	296
1.2	388	50	57	989	73 169	563	374
1.3	390	43	25	14	80 085	483	163
1.4	309	11	2	18	80 727	117	14
1.5	283	0	1	13	81 015	0	7

where demand is very low. For these regions, as demand is scattered, paying the fixed costs cannot be justified by the low usage. Note that the number of APS locations decreases for large enough demand factors, when demand equals or exceeds total capacity. The reason is that the demand per unit area around each channel increases, hence, the amount of demand that can be satisfied through the total ID capacity can be delivered though a lower number of channels.

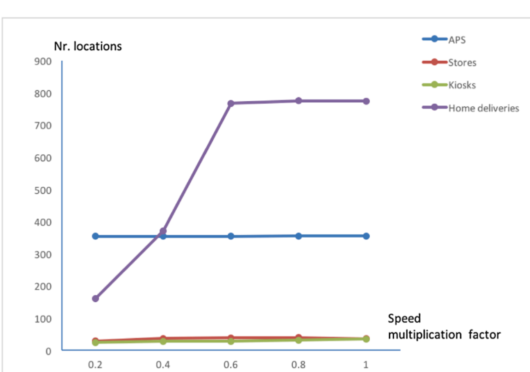
Impact of customer preferences

To illustrate the impact of customer preference for a specific channel, we will use the combinations in the last 7 rows of Table 5. The results are presented in Fig. 4a. The results indicate that while m_{APS} is kept constant ($m = 0.2$ or $m = 0.8$) and the other m-parameters are varied, the solution remains relatively constant. Changing m_{APS} has the largest impact on the solution. As a consequence of the increased utilization, the number of APS locations decreases by 7%. Due to the maximum walking distance imposed, the number of home deliveries increases. For $m > 1.2$, which means that the demand per new channel type actually exceeds its capacity, (recall that in the base scenario $m = 0.935 * cap$), the number of home deliveries increases due to insufficient capacity in new channels.

Impact of speed. We have analyzed the impact of vehicle speed



(a) Regression constant m



(b) Vehicle speed

Fig. 4. Impact of parameters on the optimal solution.

(congestion) by varying the speed by a factor k ranging from 0.2 to 1, in steps of 0.2. The results are shown in Fig. 4b. Recall that the speed impacts only the transportation costs inside an area; the lower the speed, the higher the transportation costs. For very low speed ($k = 0.2$), it is preferable to skip home delivery and pay penalties, as delivering would be more costly. As the speed increases and the transportation costs inside area decrease, the number of home deliveries increases. Similar effects can be observed when the fixed vehicle costs are varied.

Impact of the maximum walking distance

To measure the impact of the maximum walking distance a consumer is willing to walk to pick up her order, we multiplied $WD_{nc,l}$ in the base scenario by a factor ranging from 0.2 to 1.2 in steps of 0.2 (this gave a walking distance varying between 84 to 504 m in steps of 84 m). The percentage of flow through a specific channel in different scenarios is depicted in Fig. 4b, while the number of facilities of each type is presented in Table 7.

Results in Fig. 4b show that in all scenarios, the flow through existing offices remains constant, around 21.84%. This is due to the fact that for the existing offices, we did not pose a walking distance constraint, as these are traditional channels with which people are more used. In our setting, due to their efficiency, they are used at maximum capacity. The flow through stores and kiosks remain small, due to the restrictions imposed by their availability in an area. When customers are not willing to walk a long distance to pick up their orders, home delivery is the most popular channel, encompassing 72.44% of the orders. For walking distances over 168 m, the usage of APSs grows significatively, with values over 60%, as lower distribution costs compensate for the fixed APS costs. The longer customers are willing to walk (above 336 m), the lower the number of locations where APSs will be installed, despite the fact that the flow through APSs decreases only slightly (see Fig. 5). For higher walking distances, the flow through the APS is replaced by flow through stores, kiosks and home delivery (see Fig. 5). Recall that in our model, demand that can be satisfied through a channel is proportional with the distance customers are willing to walk to that channel. For larger distances, stores and kiosks become more efficient, as they can serve a larger demand. Between kiosks and stores, kiosks are preferred for small distances, while for larger distances, above 168 m, a higher percentage of the flow will go through stores. This can be explained by the fact that in our case study, stores have larger capacity, while the fixed costs per capacity unit is equal to the ones for kiosks.

Impact of facility costs

We have studied the impact of facility costs, by varying them by a factor k, ranging from 1.1 to 1.5, in steps of 0.1. In all cases, the

Table 7
Impact of walking distance on solution.

Walking dist. (m)	Nr. APS locations	Nr. stores	Nr. kiosks	Home deliv.
84	102	77	180	55 454
168	1058	43	72	12 429
252	885	27	17	1 684
336	540	23	16	652
420	356	43	27	758
504	278	62	72	1236

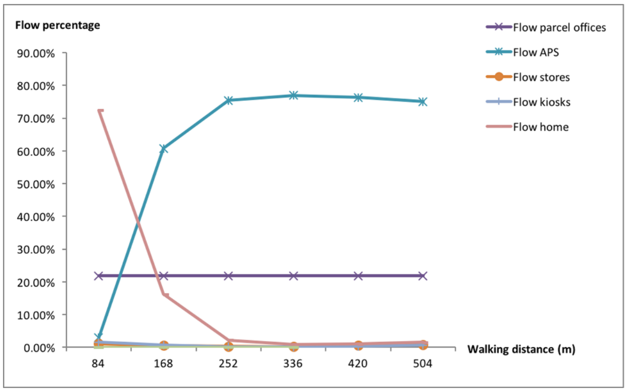


Fig. 5. Impact of walking distance on flow percentage.

percentage of flow through each channel remained the same as in Fig. 3, while the costs increased with the increase of facility costs. This robustness can be explained by the fact that multiplying all the facility costs by the same factor does not change the ordering of channels w.r.t the fixed costs/unit of capacity, which seems to be an important characteristic in choosing for a certain type of facility. Recall that the existing channel, which has the lowest cost/capacity unit is the most preferred, and APS, which has the second lowest cost/capacity unit, is the second preferred channel.

7. Quality of the heuristic

For all the scenarios in Section 6, we have compared the solution of the heuristic with the optimal solution of the piecewise linear approximation proposed in Section 4.1.

The detailed results regarding computation time and quality of the solution can be found in Appendix C, Table 10. On the above scenarios, the heuristic has an average relative error of 0.07%, with the highest errors for demand factors 1.3 and 1.4 (0.96% and 1.01% respectively). In our experiments, the relative error in all cost types but the transportation costs inside demand areas (facility, processing, penalty -costs, transportation costs CDC-demand area, transportation costs CDC-ID and transportation costs ID-area) is less than 0.65%. For the transportation costs inside demand areas, the relative average error is 2.92%, with the highest differences, 46.26% and 69.23% registered for demand multiplication factors of 1.3 (Scenario 4) and 1.4 (Scenario 5) respectively. In both scenarios, the difference in transportation costs is mainly caused by differences in 2 areas: one with low demand density (Demand Area 28042 in the outskirts, with 24 orders/km^2) and one with very high demand density (Demand Area 28015 in the center of the city, with 1096 order/km^2). In both areas, while the optimal solution chooses to leave demand unsatisfied, the heuristic chooses to satisfy it already in Phase 1. The main reason for this behaviour is that in Phase 1, the heuristic decides the percentage of satisfied/unsatisfied demand without knowing the exact location of the ID's and CDC's. As a result, the heuristic cannot capture in all detail the trade-off between penalty costs and the transportation costs between ID and areas. A remedy is attempted in Phase 3, where, given the open IDs, the optimal reallocation of demand among channels is found. However, by fixing the IDs, only a part of the trade-off is captured. Despite the difference in transportation costs in these areas, the transportation costs in the other 61 areas are similar to the costs in the piecewise linear MIP. Therefore, the impact on the total costs is small (the final cost differences are 0.96% in Scenario 4 and 1.01% in Scenario 5).

Table 10 also shows that the on average, the heuristic and the optimal solution satisfy the same quantity of demand.

The computation times of both methods are given in Table 10. On average, the heuristic was 42.5 times faster than the MIP (see Table 10). The average time for solving the MIP formulation was 320 s, with a minimum of 14sec. for a demand multiplication factor of 1.5, and a maximum of 3600 s. for demand multiplication factors 1.1. and 1.3. The heuristic spent the same time for the first case, while it solved the other case in 16Section (355 times faster). It seems the problem is harder to solve when the total demand is close to the total capacities of the ID's. For these cases, the heuristic spent most computing time in Phase 2, indicating that the choice of CDCCs and IDs is the bottleneck. When demand exceeds the capacity considerably (demand factor 1.5), the problems becomes easy, as a large fraction of demand can be disregarded.

These results suggest that the heuristic performs well when demand is lower than the total available capacity. If demand is higher than total capacity, the heuristic has higher transportation costs inside the area. However, if these costs are a small percentage of the total costs, this deviation does not affect the quality of the solution.

8. Conclusions and discussion

This paper proposed a MIP model for the network design problem of a parcel carrier that manages online orders from omnichannel retailers. The novelty of this paper is that it integrates customer preferences for deliveries and returns, the maximum walking distance to a delivery option and the impact of channel availability on customers' channel choice. We also used a detailed last mile transportation cost function, that is approximated via a piecewise linear function. We applied the model to the network design problem of a parcel carrier that serves online orders and the associated returns in the city of Madrid.

The model and the results show the importance of the right positioning and dimensioning of the capacity of new channels, when consumer preferences are taken into account. Our experiments brought forward other factors that play a role in network design: demand density, the maximum distance customers are willing to walk and speed (congestion). The model supports the idea that APSs are suitable for demand areas with moderate and high demand density. In areas with low demand density and high distribution times, home deliveries are more efficient, as installing facilities within a reasonable distance from customers is too costly in these regions. Very high traffic congestion (low vehicle speed) may result in the carrier refusing orders and paying penalty costs.

The influence of the above mentioned factors has been simplified in this model and could be further refined and improved. The demand is assumed to be estimated by expected values, and a simple linear relation is assumed between the number/volume of new channels and demand. Further research could focus on embedding more sofisticated customer preference models into the MIP, such as discrete choice models. Note that these models are highly non-linear and new algorithms need to be developed to deal with this challenge.For a review of some exact methods, heuristics and meta-heuristics for solving nonlinear mixed integer problems, we refer to Belotti et al. (2013) and Gendreau and Potvin (2010). Another very interesting extension would be to incorporate more detailed information on the demand process, such as seasonality of demand.

For a special case, often encountered in practice, the paper also describes a very fast MIP- based heuristic that finds solutions of good quality. In our experiments, the heuristic gave accurate approximations for all the costs involved (within 1%), with the exception of the situation when demand exceeds the total available capacity at intermediary depots. In these cases, although the heuristic has a good overall performance (below 1%), the transportation costs inside an area are overestimated. Further research could focus on methods that improve this aspect.

CRediT authorship contribution statement

Javier Guerrero-Lorente: Conceptualization, Data curation, Methodology, Validation, Writing - original draft, Writing - review &

editing. **Adriana F. Gabor:** Funding acquisition, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Eva Ponce-Cueto:** Conceptualization, Data curation, Methodology, Writing - review & editing, Supervision, Project administration.

Acknowledgements

We thank Guillaume Crognier from École Polytechnique Paris, France, for his help with the computational experiments and insightful comments on the paper and to dr. Edgar Blanco for many insightful discussions on the topic. This publication is based upon work supported by the Khalifa University of Science and Technology under Award No. RC2 DSO and Grant 874000218, and United Arab Emirates University under Grant 31B065.

Appendix A. Detailed parameter calculations

A.1. Facilities related parameters

The fixed costs per day for CDC, ID and parcel offices were calculated as $\frac{\text{Area} * \text{Rent per year}}{\text{Nr. working days}}$ and for the new channels was calculated by $\frac{\text{cap}_{nc} * \text{Rent per year} + \text{Equipment cost}}{\text{Nr. working days}}$, assuming $\text{Nr. working days} = 260$. The values for the areas and price are given in Table 8.

Table 8
Detailed parameters related to facilities.

	CDC	ID	Existing Channel	APS	Store	Kiosk	Home
Area (m^2)	20000	500	180	1.43	0.86	0.5	0
Rent per year (€/m^2)	55.2	228	228	228	228	228	0
Equipment costs (€)	1000	1000	1000	0	0	0	0
Fixed cost/day (€)	4247	439	158	5.10	4.6	1.78	0
Capacity (orders or returns)	8000	7000	300	20	12	7	1

A.2. Demand characteristics

A map of the demand areas can be found in Fig. 6.
The characteristics of demand areas can be found in Table 9.

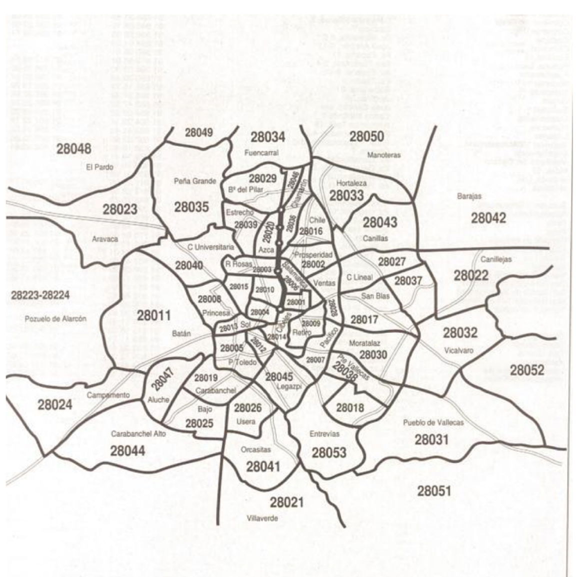


Fig. 6. Demand Areas.

Table 9
Parameters related to demand areas.

Code Area l	d_l	d_l^{urg}	Nr.stores	Nr.kiosks	$speed_l$	Code Area l	d_l	d_l^{urg}	Nr. stores	Nr.kiosks	$speed_l$
PO28001	969	20	2	2	16.09	PO28027	2015	41	5	10	15
PO28002A	892	18	2	4	26.89	PO28029	2759	56	6	10	11.21
PO28002B	892	18	2	4	26.89	PO28030	3056	62	5	14	17.92
PO28003	1318	27	1	4	20.12	PO28031A	830	17	4	5	20.65
PO28004	947	19	3	2	12.88	PO28031B	830	17	4	4	25.37
PO28005	2041	41	6	5	17.49	PO28032A	545	11	2	4	31.64
PO28006A	281	6	2	4	16.92	PO28032B	545	11	2	4	31.64
PO28006B	281	6	2	4	12.9	PO28033	2588	52	5	4	12.33
PO28006C	281	6	1	3	16.84	PO28034A	1012	21	2	3	15.83
PO28007	2178	44	4	10	11.01	PO28034B	1012	21	1	4	13.72
PO28008A	507	11	2	2	21.83	PO28035	1824	37	6	6	14.35
PO28008B	507	11	1	2	21.83	PO28036	682	14	0	1	28.05
PO28009	1187	24	1	9	15.72	PO28037	1696	34	2	11	26.66
PO28010	1178	24	4	9	12.47	PO28038A	1560	32	2	3	41.8
PO28011A	958	20	5	3	14.4	PO28038B	1560	32	1	4	36.62
PO28011B	958	20	5	2	14.4	PO28039	2091	42	7	4	12.45
PO28012	1103	23	3	6	15.18	PO28040	288	6	2	1	21.06
PO28013	313	7	4	6	11.35	PO28041A	837	17	2	4	18.84
PO28014	569	12	1	3	20.4	PO28041B	837	17	1	4	22.59
PO28015	1677	34	7	2	14.75	PO28041C	837	17	1	5	18.31
PO28016	1039	21	1	5	22.92	PO28042A	512	11	2	2	20.37
PO28017	3103	63	6	9	24.36	PO28042B	512	11	1	3	15.31
PO28018	2929	59	5	7	20.16	PO28043A	1122	23	4	6	13.78
PO28019	2502	51	13	6	44.37	PO28043B	1122	23	3	7	17.15
PO28020	1412	29	7	6	40.3	PO28044	2156	44	7	12	48.14
PO28021	2132	43	6	6	27.88	PO28045	1676	34	5	3	48.14
PO28022A	676	14	2	3	15.19	PO28048	125	3	0	2	51.26
PO28022B	676	14	3	4	18.78	PO28049A	14	1	0	1	47.33
PO28023	459	10	0	1	16.25	PO28049B	14	1	0	1	51.06
PO28024	1880	38	2	4	35.83	PO28050	218	5	3	1	67.36
PO28025	2413	49	6	11	17.14	PO28051	15	1	0	1	64.21
PO28026	1884	38	5	8	15.57						

Appendix B. Quality of the piecewise approximation

Fig. 7 shows the squareroot function and the piecewise linear approximation used in the MIP. The two functions are very close to each other, the highest error being in the interval [0,60]. Fig. 8 presents a zoom of the functions on this region showing that the maximum error is less than 0.2. Recall that in the cost function, $\sqrt{x}$ is multiplied by $\sqrt{Area_l} * tc_v * t_l$ (see 1). In our base scenario, this factor takes an average value of 6.65, with a maximum of 18 and minimum of 2. Assuming the worst case, in which in each demand region the approximation error of the squareroot function is 0.2, the total error in our base scenario is less then 0.2% of the optimal value.

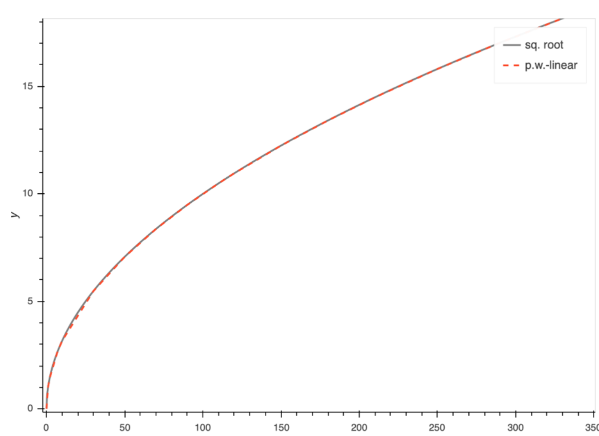


Fig. 7. Approximation of the square root function by a piecewise linear function for up to 350 stops.

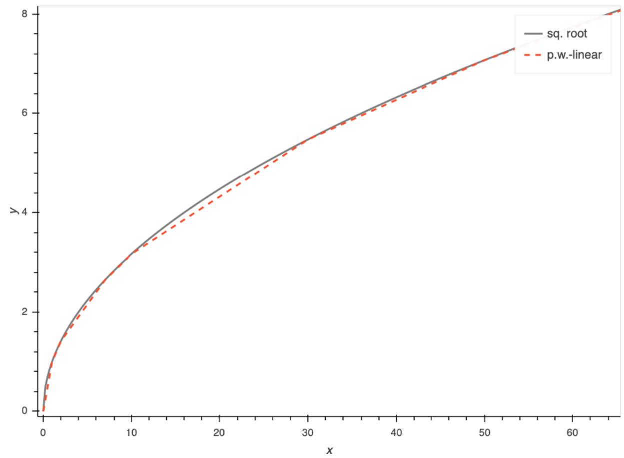


Fig. 8. Approximation of the square root function by a piecewise linear function for up to 60 stops.

Appendix C. Quality of the heuristic

See Table 10.

Table 10
Quality of heuristic

Scen.o	Factor	Time MIP-OLN (sec.)	Time Improv.[a]	Total cost difference[b]	Diff. APS loc.[c]	Diff. Stores	Diff. Kiosks	Diff. Homes	Diff APS.	Diff.unsat. demand
1	D1.0	70	25.8	0.03%	0	4	−9	−4	4	3
2	D1.1	3600	1	0.07%	1	0	−18	67	7	−23
3	D1.2	619	451.9	0.04%	0	−2	−2	2	4	3
4	D1.3	3600	355.4	0.96%	−1	0	8	678	−35	17
5	D1.4	124	11.3	1.01%	45	24	34	875	−71	21
6	D1.5	14	43.3	0.14%	23	12	15	9	−7	31
7	FC1.1	63	16.2	0.02%	−2	1	−4	15	4	−13
8	FC1.2	121	14.3	0.01%	−1	0	−3	8	3	−5
9	FC1.3	71	21.1	0.02%	0	−6	−3	−5	9	1
10	FC1.4	112	19.7	0.00%	0	−5	−2	0	7	1
11	FC1.5	105	9.6	−0.01%	−1	3	−15	−1	7	2
12	VC1.1	145	23.2	0.04%	−1	3	−6	−5	4	2
13	VC1.2	163	2.7	0.01%	−1	2	−9	8	7	−13
14	VC1.3	89	24.1	0.01%	−1	3	−7	−3	5	1
15	VC1.4	107	36.5	0.02%	0	1	−3	−4	4	1
16	VC1.5	71	27.2	0.03%	0	4	−5	−5	3	1
17	WD0.2	37	6	−0.06%	0	0	0	47	2	−47
18	WD0.4	54	9.7	0.00%	0	−2	−2	0	5	0
19	WD0.6	93	10.1	0.01%	1	2	−9	−3	6	0
20	WD0.8	100	25.9	0.02%	−1	2	−2	−10	5	1
21	WD1.2	84	7.6	0.00%	2	3	−8	−2	3	0
22	S0.2	107	16.6	0.08%	0	−1	−4	0	6	1
23	S0.4	75	5.7	0.04%	0	0	−3	−2	5	0
24	S0.6	66	37.3	0.00%	0	0	−2	−2	4	1
25	S0.8	74	10.6	0.02%	−1	0	−7	−5	7	3
26	M(0.2, 0.2, 0.2, 1)	63	21.7	0.01%	−2	4	−8	−3	4	1
27	M(0.2, 0.2, 0.8, 1)	91	7.9	0.01%	1	4	−8	−2	5	−5
28	M(0.2, 0.8, 0.2, 1)	233	13.9	0.00%	−1	1	−8	0	7	0
29	M(0.2, 0.8, 0.8, 1)	137	5.8	0.00%	−2	7	−9	−5	2	2
30	M(0.8, 0.2, 0.2, 1)	176	18	0.02%	0	0	−8	−3	8	1
31	M(0.8, 0.2, 0.8, 1)	158	3.5	0.01%	−1	0	−4	17	4	−15
32	M(0.8, 0.8, 0.2, 1)	76	10.8	0.01%	−1	7	−11	−8	4	0
33	M(0.8, 0.8, 0.8, 1)	104	8.8	0.00%	−2	0	−5	0	5	0
34	M(1.2, 1.2, 1.2, 1.2, 1)	199	70.2	0.02%	3	−8	−4	−3	8	7
35	M(1.4, 1.4, 1.4, 1.4, 1)	260	138.4	0.05%	2	−5	−6	−15	9	24
36	M(1.6, 1.6, 1.6, 1.6, 1)	778	65.9	0.06%	6	−16	8	30	9	−4
	Average	320.5	42.5	0.07%	1.55	1.34	−3.82	43.72	1.87	0.24

a. Time improvement = Time Heuristic/Time MIP-OLN. b. Cost difference = (Cost Heuristic- Optimum)/Optimum.
c. Difference value = Value Heuristic-Value optimum.

References

Acimovic, J., & Graves, S. C. (2014). Making better fulfillment decisions on the fly in an online retail environment. *Manufacturing & Service Operations Management, 17*(1), 34–51.

Agatz, N. A., Fleischmann, M., & Van Nunen, J. A. (2008). E-fulfillment and multi-channel distribution–a review. *European Journal of Operational Research, 187*(2), 339–356.

Alumur, S., Nickel, S., Saldanha-da Gama, F., & Verter, V. (2012). Multi-period reverse logistics network design. *European Journal of Operational Research, 220*(1), 67–78.

Ansari, S., Başdere, M., Li, X., Ouyang, Y., & Smilowitz, K. (2018). Advancements in continuous approximation models for logistics and transportation systems: 1996–2016. *Transportation Research, Series B: Methodological, 107*, 229–252.

Asdemir, K., Jacob, V. S., & Krishnan, R. (2009). Dynamic pricing of multiple home delivery options. *European Journal of Operational Research, 196*(1), 246–257.

Belotti, P., Kirches, C., Leyffer, S., Linderoth, J., Luedtke, J., & Mahajan, A. (2013). Mixed-integer nonlinear optimization. *Acta Numerica, 22*, 1–131.

Bretthauer, K. M., Mahar, S., & Venakataramanan, M. (2010). Inventory and distribution strategies for retail/e-tail organizations. *Computers & Industrial Engineering, 58*(1), 119–132.

Cachon, G. P. (2014). Retail store density and the cost of greenhouse gas emissions. *Management Science, 60*(8), 1907–1925.

Campbell, A. M., & Savelsbergh, M. (2006). Incentive schemes for attended home delivery services. *Transportation Science, 40*(3), 327–341.

Côté, J.-F., & Potvin, J.-Y. (2009). A tabu search heuristic for the vehicle routing problem with private fleet and common carrier. *European Journal of Operational Research, 198*(2), 464–469.

Daganzo, C. F. (1984). The distance traveled to visitnpoints with a maximum ofcstops per vehicle: An analytic model and an application. *Transportation Science, 18*(4), 331–350.

Daganzo, C. F. (1988). A comparison of in-vehicle and out-of- vehicle freight consolidation strategies. *Transportation Research Part B: Methodological, 22*(3), 173–180.

Daganzo, C. F. (1999). *Logistics system analysis.* Springer.

de Koster, M. (2002). The logistics behind the enter click. In A. Klose, & Luk Van Wassenhove (Eds.). *Quantitative approaches to distribution logistics and supply chain management* (pp. 131–148). Springer.

de Oliveira, L. K., Morganti, E., Dablanc, L., & de Oliveira, R. L. M. (2017). Analysis of the potential demand of automated delivery stations for e-commerce deliveries in belo horizonte, brazil. *Research in Transportation Economics, 65*, 34–43.

Deutsch, Y., & Golany, B. (2018). A parcel locker network as a solution to the logistics last mile problem. *International Journal of Production Research, 56*(1–2), 251–261.

Drexl, M., & Schneider, M. (2015). A survey of variants and extensions of the location-routing problem. *European Journal of Operational Research, 241*(2), 283–308.

Estadistica de Población de la Comunidad de Madrid (1996). Tabla 4001. Superficie, densidad y población clasificada por sexo y grupos de edad, volume 4. Instituto de Estadística de la Comunidad de Madrid.

Farahani, R. Z., Hekmatfar, M., Fahimnia, B., & Kazemzadeh, N. (2014). Hierarchical facility location problem: Models, classifications, techniques, and applications. *Computers and Industrial Engineering, 68*, 104–117.

Fleischmann, M., Beullens, P., Bloemhof-ruwaard, J. M., & Van Wassenhove, L. N. (2001). The impact of product recovery on logistics network design. *Production and Operations Management, 10*(2), 156–173.

Gendreau, M., & Potvin, J.-Y. (Eds.). (2010). *Handbook of metaheuristics*. US: Springer.

Govindan, K., Soleimani, H., & Kannan, D. (2015). Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operational Research, 240*(3), 603–626.

Govindarajan, A., Sinha, A., & Uichanco, J. (2018). *Joint inventory and ful fillment decisions for omnichannel retail networks.* Technical report.

Guerrero-Lorente, J., Ponce-Cueto, E., & Blanco, E. (2017). A model that integrates direct and reverse flows in omnichannel logistics networks. In Amorim, Marlene, F. C. V. M. J. & Prado, C. (Eds.), *Engineering Systems and Networks. Lecture Notes in Management and Industrial Engineering* (Vol. 1, pp. 89–97). Springer.

Hillier, F. S., & Lieberman, G. J. (1995). *Introduction to operations research.* McGraw-Hill Science, Engineering & Mathematics.

Hübner, A., Kuhn, H., & Wollenburg, J. (2016). Last mile fulfilment and distribution in omni-channel grocery retailing: A strategic planning framework. *International Journal of Retail & Distribution Management, 44*(3), 228–247.

Iwan, S., Kijewska, K., & Lemke, J. (2016). Analysis of parcel lockers' efficiency as the last mile delivery solution–the results of the research in poland. *Transportation Research Procedia, 12*, 644–655.

Jayaraman, V., Patterson, R. A., & Rolland, E. (2003). The design of reverse distribution networks: Models and solution procedures. *European Journal of Operational Research, 150*(1), 128–149.

Khatami, M., Mahootchi, M., & Farahani, R. Z. (2015). Benders' decomposition for concurrent redesign of forward and closed-loop supply chain network with demand and return uncertainties. *Transportation Research Part E: Logistics and Transportation Review, 79*, 1–21.

Krikke, H., Bloemhof-Ruwaard, J., & Van Wassenhove, L. N. (2003). Concurrent product and closed-loop supply chain design with an application to refrigerators. *International Journal of Production Research, 41*(16), 3689–3719.

Lang, G., & Bressolles, G. (2013). *Economic performance and customer expectation in e-fulfillment systems: A multi-channel retailer perspective. Supply Chain Forum: An International JournalVol. 14*. *Supply Chain Forum: An International Journal* Taylor & Francis16–26.

Lim, H., & Shiode, N. (2011). The impact of online shopping demand on physical distribution networks: a simulation approach. *International Journal of Physical Distribution & Logistics Management, 41*(8), 732–749.

Lim, S. F. W. T., Rabinovich, E., Rogers, D. S., & Lasester, T. M. (2016). Last-mile supply network distribution in omnichannel retailing: A configuration-based typology. *Foundations and Trends in Technology, Information and Operations Management, 10*(1), 1–87.

Lipsman, A. (2019). Global e-commerce 2019.

Listeş, O., & Dekker, R. (2005). A stochastic approach to a case study for product recovery network design. *European Journal of Operational Research, 160*(1), 268–287.

Mahar, S., & Wright, P. (2009). The value of postponing online fulfillment decisions in multi-channel retail/e-tail organizations. *Computers & Operations Research, 36*(11), 3061–3072.

Marchet, G., Melacini, M., Perotti, S., Rasini, M., & Tappia, E. (2018). Business logistics models in omni-channel: A classification framework and empirical analysis. *International Journal of Physical Distribution & Logistics Management, 48*(4), 439–464.

McLeod, F., Cherret, T., & Song, L. (2006). Transport impacts of local collection/delivery points. *International Journal of Logistics: Research and Applications, 9*(3), 307–317.

Min, H., Ko, H. J., & Ko, C. S. (2006). A genetic algorithm approach to developing the multi-echelon reverse logistics network for product returns. *Omega, 34*(1), 56–69.

Nagy, G., & Salhi, S. (2007). Location-routing: Issues, models and methods. *European Journal of Operational Research, 177*(2), 649–672.

Newell, G. (1973). Scheduling, location, transportation, and continuum mechanics: Some simple approximations to optimization problems. *SIAM Journal on Applied Mathematics, 25*(3), 346–360.

Newell, G. F. (1971). Dispatching policies for a transportation route. *Transportation Science, 5*(1), 91–105.

Prodhon, C., & Prins, C. (2014). A survey of recent research on location-routing problems. *European Journal of Operational Research, 238*(1), 1–17.

Rigby, D. K. (2011). The future of shopping. *Harvard Business Review*.

Şahin, G., & Süral, H. (2007). A review of hierarchical facility location models. *Computers and Operations Research, 34*(8), 2310–2331.

Salema, M. I. G., Póvoa, A. P. B., & Novais, A. Q. (2009). A strategic and tactical model for closed-loop supply chains. *OR Spectrum, 31*(3), 573–599.

Taha, H. (2007). *Operations research: An introduction.* New Jersey: Pearson Education Inc Publisher.

Winkenbach, M., Kleindorfer, P. R., & Spinler, S. (2015). Enabling urban logistics services at la poste through multi-echelon location-routing. *Transportation Science, 50*(2), 520–540.

Yadav, V. S., Tripathi, S., & Singh, A. (2017). Exploring omnichannel and network design in omni environment. *Cogent Engineering, 4*(1).

Yadav, V. S., Tripathi, S., & Singh, A. (2019). Bi-objective optimization for sustainable supply chain network design in omnichannel. *Journal of Manufacturing Technology Management, 30*(6), 972–986.

Yang, X., Strauss, A. K., Currie, C. S. M., & Eglese, R. (2016). Choice-based demand management and vehicle routing in e-fulfillment. *Transportation Science, 50*(2), 473–488.

Zhang, S., Lee, C. K. M., Wu, K., & Choy, K. L. (2016). Multi-objective optimization for sustainable supply chain network design considering multiple distribution channels. *Expert Systems with Applications, 65*, 87–99.

Zhang, S., Zhu, H., Li, X., & Wang, Y. (2019). Omni-channel product distribution network design by using the improved particle swarm optimization algorithm. *Discrete Dynamics in Nature and Society, 2019*, 1–15.