Combining crowdsourcing and mapping customer behaviour in last-mile deliveries

Arianna Seghezzi*, Chiara Siragusa*, Riccardo Mangiaracina* and Angela Tumino*

** Dipartimento di Ingegneria Gestionale, Politecnico di Milano, Via Lambruschini 4B, 20156 – Milan – Italy (arianna.seghezzi@polimi.it, chiara.siragusa@polimi.it, riccardo.mangiaracina@polimi.it, angela.tumino@polimi.it)*

Abstract: In the light of the dramatic rise of online sales, last-mile deliveries (i.e., the delivery of products ordered online to the final customer) have been increasingly gaining the attention of both managers and academics. As a matter of fact, they are very critical in terms of effectiveness (as customers demand fast and accurate deliveries), and efficiency (since they imply very high costs). Henceforth, logistics players operating in the B2C e-commerce environment are striving to find and implement innovative solutions, different from the costly traditional by-van home deliveries. Among the options analysed by scholars so far, two promising ones are crowdsourcing logistics (i.e., outsourcing delivery activities to "common" people) and mapping the behaviour of customers (i.e., analysing the probability distribution of the customer presence at home and accordingly scheduling deliveries to minimise the probability of failed deliveries). In this paper, we introduce and study a combination between the two solutions, proposing a variant of the Vehicle Routing Problem, which considers both the Availability Profiles and Occasional Drivers (VRPAPOD). We model the delivery problem as a mixed-integer program and solve it with a branch-and-price algorithm. To analyse the benefit of the combined use of crowdshipping and customers availability profiles (APs), we conduct several experiments in a real context in the city of Milan, randomly extracting 100 customers in a 16 km^2 area. The combined solution is compared with two benchmarking models, namely the traditional home delivery (traditional VRP) and the crowdsourcing logistics option (Vehicle Routing Problem with Occasional Drivers (VRPOD)). Results prove that logistics players can achieve important benefits by relying on the crowd and scheduling deliveries according to clients' APs, which become more significant in case of high drivers availability.

Keywords: last-mile delivery, logistics, e-commerce, crowdsourcing

I. Introduction and Literature Background

Business-to-consumer (B2C) e-commerce has dramatically evolved over the years, switching from a simple alternative to brick-and-mortar shops, to a key direction pursued by companies to stay competitive on the market. In this context, different novel challenges emerge, especially if considering logistics. Logistics players must handle an increasing number of small parcels to be delivered to multiple destinations, which makes the last-mile delivery (LMD) segment very critical in terms of both efficiency and effectiveness (Wang et al., 2016). The desire to be competitive and gain a higher market share, in a context where customers are ever more demanding, is pushing players to reduce delivery times and offer free shipping (Chang et al., 2021). If combining this tendency with the expensive nature of the last-mile delivery, which accounts for more than half of the total delivery costs (Seghezzi et al., 2021), the step towards greater efficiency – while meeting the stringent delivery standards – becomes crucial.

Many efficiency issues can be traced back to the common problem of failed deliveries, i.e., deliveries not accomplished due to the absence of the customer at home (Pan et al., 2017), which result in higher costs for the company (as parcels are typically assigned to subsequent delivery tours for new attempts).

In this direction, an innovative last-mile delivery solution proposed by academics implies considering the probability that the client is available during the delivery. Referred to as "mapping customer behaviour", it consists in monitoring the presence of the customers at home throughout the day based on their interaction with smart-home devices, to derive profiles of the customers presence at home, and accordingly schedule deliveries (Mangiaracina et al., 2019). Among the factors facilitating the diffusion of this solution, the main ones are the growth of the market of IoT and Smart Home Devices - to track customer presence - and the evolution of data mining and analytics techniques - to extract the attendance probability profiles from available data. However, delivering according to the schedule of customers requires great flexibility and reactivity in the delivery system. As a matter of fact, the highest home-attendance time windows are concentrated during the evening (after working hours) and there is a tendency for the intervals of different customers to overlap (Archetti et al., 2020). Henceforth, to fully exploit the benefits of this innovative solution, the long 8 hour-shifts of the drivers of traditional vans are not suitable, and it should be combined with a very high degree of flexibility in managing the deliveries along the day.

With reference to this innovative solution based on mapping the presence of the customer, two major research gaps may be highlighted.

- First, very few (2) models have been presented accordingly developing variants of the VRP (Vehicle Routing Problem), and they both show some significant limitations. Pan et al. (2017) propose exploiting data about electrical energy consumption to build customer home attendance profiles. The model detects the presence or absence of the customer at home through a binary function, based on the combination of peaks or significant variations in electricity consumption. In the paper by Florio et al. (2018) customer home attendance is instead estimated based on historical data about past deliveries and preferential time windows. Nonetheless, in both cases the objective function only aims to maximise hit rate, while the travel time is introduced as a constraint, thus not considering the trade-off between the two dimensions. Moreover, these methods show unreliability issues, since the time needed to gather the necessary amount of data may be significant, and during such a long period people may change their habits.
- Second, mapping the presence of the customer is only studied in literature as a stand-alone option, and no attempts have been made to combine it with other innovative last-mile delivery solutions. A promising solution that makes flexibility its strong point is crowdsourcing logistics (Castillo et al., 2018). Based on the principles of the disruptive sharing economy phenomenon (Habibi et al., 2017), delivery activities are outsourced to "ordinary" people. These people, referred to as "the crowd", share their available capacity in terms of both time – i.e., time they devote to performing deliveries – and space - i.e., the capacity of their own vehicle. Relying on a pool of common people – who may even be temporarily engaged for one single delivery – offers the company the possibility to deliver outside the typical service hours, and to prevent the need to hire a deliveryman that works for the entire duration of a shift (Macioszek, 2017). In addition, given the fast response of occasional drivers and the increased parallelism in task execution, there are further advantages in terms of delivery speed, compliance to customer location and delivery time, and – consequently – the ability to cope with demand peaks and valleys.

II. Objectives and Methodology

In line with the identified shortcoming, the present work aims to answer the following research question:

What are the impacts of combining Crowdsourcing and Mapping Customer Behaviour on the efficiency performances of Last-Mile Delivery?

To answer the question, a mixed approach was followed, combining three major methodologies.

(i) Literature review: after a more general narrative review aimed to identify gaps and refine the research goal, a methodological-oriented review was performed to provide insights for the development of the analytical model. More specifically, two main clusters of papers were addressed. On the one hand, papers proposing VRPs that consider the customers home attendance (e.g., Florio et al. 2018). On the other hand, papers addressing crowdsourcing logistics, (e.g., Archetti et al., 2016). This analysis supported the development of the proposed model (including the variables, constraints, and objective function).

(ii) Vehicle Routing Problem with Availability Profile and Occasional Drivers (VRPAPOD) formulation: a new variant of the Vehicle Routing Problem (VRP) was developed, with the objective function including both the probability to find the customers at home (referred to as Availability Profile (AP)), and crowdsourcing (which is the outsourcing of deliveries to Occasional Drivers (OD)). The resulting variant is referred to as Vehicle Routing Problem with Availability Profile and Occasional Drivers (VRPAPOD).

(iii) Computational experiment: the analytical formulation of the VRPAPOD was followed by some computational experiments and sensitivity analyses. These were performed in order to derive numerical results concerning a specific application context. Models were coded in *Python* and solved using the *Gurobi* solver.

III. VRPAPOD

Based on the aforementioned premises, the VRPAPOD assigns deliveries to a fleet of occasional drivers providing their availability to work for a given time in exchange for economic compensation. Moreover, these deliveries are scheduled in order to meet the probability of finding customers at home as much as possible (i.e., maximizing the so-called hit rate). The VRPAPOD model has common characteristics with some of the most analysed variants of Vehicle Routing Problems. First, it is a VRP with Multiple-Time Windows (VRPMTW), since it allocates the deliveries to customers in specific time slots within the day. Second, it is a VRP with Occasional Drivers (VRPOD) as deliveries are made leveraging on the crowd. Finally, it is a VRP with Availability Profile (VRPAP) as the choice of time-windows in which customers are served depends on their home-attendance profile.

The problem is formulated as a Mixed Integer Programming model (MIP) aimed at both minimising the total compensation for ODs, and maximising the probability of successful deliveries.

The included variables and parameters are the following:

- HC: drivers' hourly fee
- y_h: binary variable, equal to 1 if driver $h \in H$ is activated for one or more deliveries, 0 otherwise.
- Shf_h: duration of the shift of driver $h \in H$

- $x_{i,j,h}$: binary variable equal to 1 when the arc (i, j) is traversed by driver h $\in H$, 0 otherwise.
- $AFDC$*(1-$prob^{p}_{i}$): expected cost of failed delivery. It is computed as the product between: the average cost of failed deliveries due to customer non-attendance (AFDC) and the probability of not finding the customer $i \in C$ at home (1-$prob^{p}_{i}$), within the time-slot [l^{p}_{i}, u^{p}_{i}] $\in Ji$ (with $p \in [1, z]$). The delivery time-horizon [T_{i}, T_{f}] has been divided in z intervals [l^{p}, u^{p}] of same length Δ. Each customer has been then associated with the set of z time-windows Ji , where each slot [l^{p}_{i}, u^{p}_{i}] $\in J_i$ defines a period in which the customer i can be served.
- $v^{p}_{i,h}$: binary variable equal to 1 when node $i \in N$ is visited by driver h$\in H$, in the time-slot p.

In line with most of the reviewed papers involving the crowd (e.g., Archetti et al. 2016, Macrina et al. 2017), the objective function is defined as a cost minimisation. It consists of two main components.

$$\min(\sum_{h\in H}(HC * y_h * Shf_h)\ (1) + \sum_{h\in H}\sum_{i\in N}\sum_{j\in C}\sum_{p=1}^{z}(x_{i,j,h} * (1 - prob_i^p) * AFDC. * v_{i,h}^p)\ (2)$$

The first element (1) computes the total cost incurred by the company to reward ODs (Occasional Drivers). For each activated OD, a fixed hourly fee is multiplied by the duration of the related working time-slot. The second element (2) of the formula includes the total expected cost of failed deliveries incurred by the company. It is computed as the sum of the failure probabilities due to customers non-attendance, multiplied by the Average Cost of Failed Deliveries (AFDC). Depending on the different ODs distributions, the system selects those time-slots that maximise the total percentage of successful deliveries.

To model the VRPAPOD, different constraints were defined. The main ones are the following:

Spatial constraints. They ensure that the spatial dimension does not lead to infeasible solutions when matching OD and goods to be delivered. The category includes those constraints that regulate the movement of the occasional drivers between the different nodes of the graph: flow balancing and route continuity; sub-tours elimination; closed-circuit; visiting each customer once; ODs leaving the depot at most once; preventing departure from the endpoint.

Temporal constraints. They guarantee that the time windows of drivers and customers are respected. They include: non-negativity of service time; compliance with customers and drivers time windows; sequentiality of the service times of customers in the tour; service for each node in one and only one time slot.

Matching and delivery constraints. They govern the behaviour of occasional drivers. The category includes those constraints related to the demand - total customers demand satisfaction - and capacity - depot and vehicle capacity, and maximum number of drivers.

The variables and parameters, as well as the complete formulation of the model, are presented in Appendix A.

A. Benchmarking models

To validate the proposed combination and to understand the convenience of combining the mapping of customers' behaviour with crowdshipping, two benchmark models were considered: the traditional VRP and the VRPOD. Comparing the VRPAPOD with a VRPOD allows to assess the contributions brought by the inclusion of customer availability profiles within the delivery plan. On the other hand, evaluating the VRPODAP model against a VRP provides an overall view of the potential and limitations of the proposed combination compared to the traditional van-based system. Unlike the VRPAPOD, the benchmark models do not take customer home-attendance profiles into account. Nonetheless, the objective functions are similar, and are defined as a cost minimisation.

In the traditional VRP, a fleet of trucks is deployed by a company to deliver parcels to a set of customers C. However, since vans are owned by the transport company, the objective becomes the minimisation of transport cost, which depends on the distance travelled. Regarding the constraints, apart from the changes related to the use of the van, the main difference lies in the waiting time. The waiting time, which is defined as the time during which the driver is neither travelling nor serving customers, is set to 0 in the traditional VRP.

In the VRPOD, deliveries are performed by a fleet of occasional drivers who are available to make deliveries in exchange for an economic reward (Archetti et al., 2016). The system assigns delivery requests and routings to ODs in order to minimise the total compensations borne by the company. As in the case of the VRPODAP, the total costs depend on the length of the shifts of the ODs selected for deliveries. Concerning the constraints, also in the VRPOD the waiting time is set to 0. As the customers home-attendance profiles are not taken into account, there are no reasons for drivers to spend time other than travelling or serving customers.

B. Resolution algorithm and tuning phase

Concerning the resolution process, models were solved both heuristically and optimally through variants of the branch-and-cut (B&C).

Within the algorithm, a search tree is constructed. Each node of the graph consists of a relaxed mixed-integer linear program that includes: the constraints of the original problem; its integer variables relaxed to be continuous; and restrictions added on the bounds of the integer variables to incentivise them to take integer values. In the sub-problems solution, if integer variables take on fractional values, cuts are introduced. The latter are restrictions that cut away parts of the feasible region of the relaxation containing fractional solutions.

Periodically, in the various stages of the B&C algorithm, different heuristics are applied to the relaxed subproblems to find feasible integer solutions. In line with previous literature, the resolution algorithm was followed by a testing phase for early validation and models improvement. It was organised in two main sub-processes: solver choice and tuning model performance.

IV. COMPUTATIONAL EXPERIMENT

A. Base case

Several experiments were conducted with different data sets to evaluate the effects of the proposed VRPAPOD on the efficiency performance in LMD. The model was applied on an area of 16,2 $km2$ within the city of Milan.

Considering the probability profiles of customers home-attendance, a sociological analysis was performed. The study focused on the characteristics of the households living in the examined area, paying attention to family composition, age of the members, and employment status. Then, such data was used to define a distribution of customers organised into five classes. In the light of the scientific studies on occupancy detection through inhabitants' electricity consumption (Monacchi et al., 2014) and on the analysis of the impact of family composition on home occupancy habits (Bison et al., 2018), a probability profile was defined for each class (with C1 to C5 being the classes of customers), and is presented in Fig.1.

# Timeslots	Timeslots (min)	C1	C2	C3	C4	C5.a	C5.b
1	[0, 60]	95%	5%	5%	40%	20%	80%
2	(60, 120]	95%	10%	10%	35%	20%	80%
3	(120, 180]	95%	10%	10%	35%	20%	80%
4	(180, 240]	95%	10%	10%	35%	35%	65%
5	(240, 300]	95%	10%	10%	40%	30%	70%
6	(300, 360]	95%	25%	25%	50%	25%	75%
7	(360, 420]	95%	20%	40%	45%	25%	75%
8	(420, 480]	95%	10%	40%	40%	40%	60%
9	(480, 540]	95%	10%	50%	40%	30%	55%
10	(540, 600]	95%	25%	70%	50%	50%	50%
11	(600, 660]	95%	80%	85%	80%	60%	50%
12	(660, 720]	95%	85%	90%	85%	60%	55%
13	(720, **780**]	95%	95%	95%	95%	50%	60%

Fig.1. Availability profiles

The computational experiments were run using *Gurobi* solver accessed via *Python* API. Despite the model tuning phase, the experiments still required significant computing power. For the VRPAPOD and VRPOD, to get around this problem and solve large-size instances (i.e., 100,90,80 clients), the customers extracted from the analysed area were randomly divided into two subgroups of equal dimensions (S1, S2). The results of the instances were obtained by aggregating the outputs of each sub-part. For each instance, the divergence, given by the random selection of the samples (e.g., S1, S2) from the original distribution of the customers' classes, was mitigated by repeating the extraction process 3 times. For sake of example, the outcomes of an instance of 100 customers were derived from the average results of 2x3 subgroups of size 50.

Table I summarises the main input values.

TABLE I
INPUT VALUES FOR THE BASE CASE

Variable/ parameter	Value
Delivery time horizon	0-780 min (8-21); 0-540 min (8-17) (VRP)
Length of timeslots	60 min
Number of customers (NC)	100
Service time	2 min/customer
Customer demand	1 parcel/customer
Depot capacity	100 parcels
Driver hourly fee	15 €/hour
Average failed delivery cost	16.9 €
Number of ODs	16
OD capacity	15 parcels
ODs distribution	2x60 min; 6x120 min; 8x180 min
Variable cost	33km/h*13l/100km*1.5€/l =0.11 €/min
Fixed cost	170€
Truck capacity	100 parcels

Table II shows the results of the three models in the presented base case. In line with the research question, the emphasis was placed on those outputs having a closer relationship with the efficiency performances: the total delivery costs, the percentage of failed deliveries, and the average delivery cost per parcel (CP). Although the VRPAPOD has the highest expenses in terms of compensation paid to ODs (240 €), it implies the lowest cost per parcel (2.54€). This result can be entirely attributed to the improvements in the percentage of failed deliveries (as the delivery cost per parcels is derived allocating the total cost to successfully accomplished deliveries).

TABLE II
RESULTS FOR THE BASE CASE

Model	Total delivery cost (€)	% failed deliveries	Delivery cost per parcel (€/parcel)
VRP	200.98	36.2%	3.15
VRPOD	240	31.86%	3.12
VRPAPOD	240	5.85%	2.54

The higher total cost is mainly due to the availability of ODs under consideration. In fact, the system activates drivers serving customers in their best time windows (highest attendance probability), including those working for longer shifts. Conversely, the reduction of failed deliveries is due to the possibility to schedule deliveries according to the attendance profiles of customers.

Focusing on the VRP, the further worsening of the failure rate (36.2%) compared to the other models is mainly related to the narrow horizon (8:00-17:00). In fact, besides not taking into account the probability of attendance, the van does not benefit from the delivery-time extension characterising the crowd.

B. Sensitivity analysis

Some sensitivity analyses were run to test the robustness of the results. The parameters under scrutiny were: OD capacity; number of customers; length of delivery time horizon.

OD capacity The considered variation (Table III) includes: 8, 10, 15 (base case 15). The reduction in the capacity leads to a worsening of the percentage of failed deliveries caused by the low number of available ODs, especially of those working in the evening (when there is a higher density of optimal time windows). The same applies to the cost per order, which tends to increase as the capacity decreases. This trend is explained by the growth in both the ODs compensation and the percentage of failed deliveries. Comparing the two variants of the crowd-based VRP (i.e., VRPOD, VRPAPOD), there is a general cost reduction when introducing the availability profile. In fact, although ODs compensations are generally not lower than those of the VRPOD, maximising successful deliveries results in an overall reduction in the cost per parcel. The situation is different when comparing the outcomes with those of the VRP (with the van capacity staying the same). Given the base case distribution of ODs, decreasing their capacity in the VRPAPOD leads to higher costs.

TABLE III
RESULTS OF SENSITIVITY ANALYSIS – OD CAPACITY

Model	Delivery cost (€)			% failed deliveries			CP (€)		
	8	10	15	8	10	15	8	10	15
VRPAPOD	480	305	240	7.7	6.2	5.9	5.2	3.3	2.6
VRPOD	480	300	210	31.8	32.8	31.9	7.1	4.5	3.1
VRP		201			36.2			3.2	

Number of customers. The considered variation (Table IV) includes: 80, 90, 100 customers (base case 100). Decreasing the number of customers to be served leads to a significant reduction in the ODs compensation and costs per parcel. While the former is a consequence of the lower number of activated drivers, the latter is caused by their shorter average shift duration (with a corresponding decrease in the average compensation). The lower cost per parcel of the VRPAPOD, compared to both the VRPOD and VRP models, is explained by the improved percentage of successful deliveries. With respect to the truck-based system, the cost gap getting wider as the number of clients decreases reflects the opposite trends of the two models, given the influence of van fixed costs on a smaller customer base.

TABLE IV
RESULTS OF SENSITIVITY ANALYSIS – NUMBER OF CUSTOMERS

Model	Delivery cost (€)			% failed deliveries			CP (€)		
	80	90	100	80	90	100	80	90	100
VRPAPOD	165	210	240	5.9	5.8	5.9	2.2	2.5	2.6
VRPOD	150	180	210	33.4	28.3	31.9	2.8	2.8	3.1
VRP	197	199	201	37.6	37.9	36.2	3.9	3.6	3.2

OD capacity and Number of customers. After varying the number of customers and the ODs capacity individually, the analysis was completed addressing the two parameters together. Figure 2 shows the delta cost per parcel obtained for different dimensions, considering the cost of the base case as reference value. It is computed as:

$$\Delta\, cost_{i,j} = cost\ per\ parcel_{i,j} - cost\ per\ parcel_{15,100}\ \ \forall\, i\, \in \{15,10,8\}, \forall\, j\, \in \{100,90,80\}$$

Reading the matrix by column, it is manifest how reducing the ODs capacity leads to an increase in the cost per parcel. On the other hand, when examining the values per line, the reduction in the number of customers with the base-case distribution of ODs leads to a decrease in costs. This improvement is mainly attributable to the shorter average shift lengths of the activated ODs.

ODs capacity \ Number of customers	80	90	100
15	-0,40	-0,12	2,59
10	+0,58	+0,93	+0,66
8	+1,51	+2,05	+2,61

Fig. 2. Δ cost VRPAPOD, 16ODs

An exception to this trend is represented by the instance "Number of customers: 90, capacity ODs: 10", in which, due to the sub-division in half instances for complexity reduction, the system activated the same drivers as the larger-size problem "Number of customers: 100, capacity ODs: 10". The outlier testifies that, in the case of equal average shift duration, the impact of the ODs capacity saturation plays a significant role in cost variation.

Additional analyses were performed on the percentage variation of the cost per parcel for the three VRP variants. Considering VRPOD vs. VRPAPOD, integrating the attendance profiles of customers into crowd-delivery models (under the same conditions of capacity and customers’ size) leads to an overall increase in efficiency In fact, for each capacity value and number of customers, the VRPOD costs more than the VRPAPOD. The result is due to the higher percentage of successful deliveries. More complex considerations emerge when comparing the VRPAPOD and the VRP. For this specific case, with this base-case distribution of ODs, the behaviour cannot be generalised but should be analysed case by case. VRPAPOD is more cost-efficient for higher ODs capacities, with the performance gap increasing as the customers size decreases. The situation is different for lower ODs capacities, where the van system appears to have a lower cost per parcel due to the higher compensation paid to occasional drivers.

Delivery time horizon

This last analysis evaluates the impact of the proposed innovative solution applied in the same delivery time horizon as the traditional VRP, i.e., from 8 to 17 (and not from 8 to 21). With such a shorter delivery horizon, the positive effect of combining Crowdsourcing and

Availability Profiles is significantly reduced. In fact, compared to the VRPAPOD assessed over the 8-21 horizon, there is a significant increase in failure rate (average failure rate of 31.1% against 6.8% of the VRPAPOD case 8-21,with 28ODs having a capacity equal to 8). This results in higher costs per parcel, regardless of the number of customers considered. If compared to the traditional van-based system, at full customer size (i.e., 100), the cost per parcel is slightly higher. However, when the number of customers to be served decreases, the situation is reversed; this is due to both the lower failure rates of the VRPAPOD and the higher influence of van fixed costs (Fig.3.). This analysis shows that delivering with the same working horizon of the traditional system, which is usually shorter, prevents the full exploitation of the potential of the configuration proposed. In fact, drivers cannot visit customers in those time slots that generally correspond to a higher probability of attendance (evening hours).

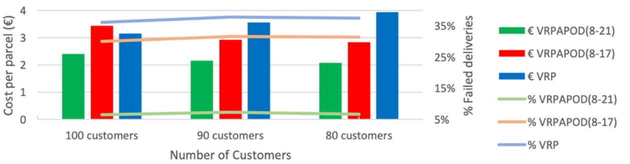

Fig. 3. Cost per parcel and % failed deliveries VRPAPOD (8-21) vs. (8-17)

V. CONCLUSIONS

With the introduction of the new VRP variant (i.e., the VRAPOD model, including both Availability Profiles and Occasional Drivers), it has been possible to answer the research question guiding the work. The computational experiments show that the simultaneous application of crowd-delivery and customer presence profiles leads to benefits in terms of reduced probability of failed deliveries, with a consequent increase in efficiency. These results have been confirmed by comparing the proposed model with the most common delivery configuration (i.e., van-based traditional VRP) and the application of a crowd-based model (i.e., VRPOD).

This research has both academic and managerial implications. Considering academia, it enriches extant literature in the field of last-mile deliveries in a twofold direction (both for the combination of different innovative last-mile solutions, and for the single Mapping Customer Behaviour components). As a matter of fact, on the one hand, it proposes a joint use of solutions not yet approached by academic research, providing a mathematical formulation for a new family of VRPs, i.e., VRPAPOD. On the other hand, in developing the model, the work addresses some topics literature that have been less discussed so far, introducing, for instance, the probability distribution of home-attendance in the objective function and the multiple-parcels delivery system for the crowd. Considering industry, beside confirming and quantifying the potential of the proposed delivery option from a theoretical side, the computational experiment highlighted the stemming opportunities for managers. The analyses show the critical parameters to be monitored to effectively implement the combination (e.g., adequate availability of occasional drivers).

The results demonstrated how integrating the attendance probability of customers into the scheduling of deliveries can lead to benefits, in terms of both efficiency and effectiveness. The combination gives the possibility to deliver outside traditional working hours while reducing the probability that ODs delivery attempts will fail. Moreover, it ensures better scheduling of deliveries in case customers' home attendance is concentrated in specific periods of the day.

Despite the highlighted contribution, the work also presents some limitations, which could be solved through further developments. Among the main reasons behind such shortcomings, there is the well-known difficulty in solving Vehicle Routing Problems (which are NP-hard problems), combined with the introduction of innovative solutions. First, the occasional drivers were distributed homogeneously in different slots within the day, with the only precaution to cover the entire delivery time horizon. However, in real cases, a higher number of ODs is generally available in the afternoon or evening hours. Still, this change could lead to an improvement in the VRPAPOD performances (and thus to a greater advantage compared to the traditional VRP). Future works could be aimed at assessing the impact of new distributions of drivers on the cost outcomes. Second, the point of view of the logistics company was considered to assess the behaviour of the delivery configuration under scrutiny, with reference to efficiency goals. Future studies could be performed to evaluate the perspective of occasional drivers, or testing the outcome in terms of sustainability performances (i.e., social and environmental). Third, like any other data-driven model, the approach was limited by computational constraints. The optimal resolution method showed its inefficiency when the problem size increased, and it took an extremely long time to provide a solution. The resolution of the instances in an acceptable timeframe led to simplifications in computational experiments. First of all, for models involving the crowd, the partition of the customers to be served into sub-problems of the same size was implemented. Furthermore, some assumptions were introduced when deriving the availability profiles: clients were clustered into five classes, each one characterized by its attendance profile, extracted as the daily average of the weekdays during the period of data collection. It would be interesting for future works to consider each client's presence profile taking into account the daily variation, including weekends. In line with these considerations, an adapted heuristic approach is necessary to investigate the problem on a more complex and larger scale. Thus, as future work, an ad-hoc heuristic method could be developed to give near-optimal solutions for more realistic and larger instances in a reasonable time.

REFERENCES

[1] Archetti, C and Bertazzi, L. (2021), "Recent challenges in Routing and Inventory Routing: E- commerce and last-mile delivery", Networks., 77: 255– 268,

[2] Castillo, V. E., Bell, J. E., Rose, W. J. and Rodrigues, A.M. (2018), "Crowdsourcing Last Mile Delivery: Strategic Implications and Future Research Directions", Journal of Business Logistics, 39: 7-25,

[3] Chang, S., Dong, Y. and Wang, X. (2021), "Optimal shipping policy in retail competition and its effect on customers", Electronic Commerce Research and Applications, Vol. 45, 101020

[4] Florio, A. M., Feillet, D. and Hartl, R. F. (2018), "The delivery problem: Optimizing hit rates in e- commerce deliveries", Transportation Research Part B: Methodological, Vol. 117, Part A, pp. 455-472

[5] Habibi, M. R., Davidson, A. and Laroche, M. (2017), "What managers should know about the sharing economy", Business Horizons, Vol. 60, Issue 1, pp. 113-121,

[6] Macioszek, E. (2017), "First and Last Mile Delivery–Problems and Issues.", Scientific and Technical Conference Transport Systems Theory And Practice, 147–154. Springer, Cham.

[7] Macrina, G., Di Puglia Pugliese, L., Guerriero, F. and Laganà, D. (2017), "The Vehicle Routing Problem with Occasional Drivers and Time Windows". In: Sforza A., Sterle C. (eds) Optimization and Decision Science: Methodologies and Applications, ODS 2017, Springer Proceedings in Mathematics & Statistics, Vol. 217, Springer, Cham

[8] Mangiaracina, R., Perego, A., Seghezzi, A. and Tumino, A. (2019), "Innovative solutions to increase last-mile delivery efficiency in B2C e-commerce: a literature review", International Journal of Physical Distribution & Logistics Management, Vol. 49 No. 9, pp. 901-920.

[9] Pan, S., Giannikas, V., Han, Y., Grover-Silva, E. and Qiao, B. (2017), "Using customer-related data to enhance e-grocery home delivery", Industrial Management & Data Systems, Vol. 117 No. 9, pp. 1917-1933,

[10] Seghezzi, A. and Mangiaracina, R. (2021), "Investigating multi-parcel crowdsourcing logistics for B2C e-commerce last-mile deliveries", International Journal of Logistics Research and Applications

[11] Wang, Y., Zhang, D., Liu, Q., Shen, F. and Lee, L. H. (2016), "Towards enhancing the last-mile delivery: An effective crowd-tasking model with scalable solutions", Transportation Research Part E: Logistics and Transportation Review, Vol. 93, pp. 279-293.

Appendix A VARIABLES AND PARAMETERS

Sets

$\boldsymbol{C}$	Set of customers served by ODs
$\boldsymbol{H}$	Set of ODs available
$\boldsymbol{o, f}$	*o* is the point from which ODs depart and *f* the same node but taken as point of arrival
$\boldsymbol{N}$	Set of total nodes (*o U C U f*)
$\boldsymbol{J_i}$	Set of available time-windows of customer $i \in C$
$\boldsymbol{J_o}$	Set of available time-windows for node o

Variables

$\boldsymbol{x_{i,j,h}}$	$\begin{cases} 1 \ if \ the \ driver \ h \in H \ uses \ the \ arch \ connecting \ node \ i \ with \ node \ j \\ 0 \ otherwise \end{cases}$
$\boldsymbol{y_h}$	$\begin{cases} 1 \ if \ the \ driver \ h \in H \ is \ activated \\ 0 \ \ otherwise \end{cases}$
$\boldsymbol{v_{i,h}^p}$	$\begin{cases} 1 \ if \ node \ i \in N \ is \ served \ by \ driver \ h \in H \ in \ the \ time \ window \ [l_i^p, u_i^p] \\ 0 \ otherwise \end{cases}$
$\boldsymbol{s_{i,h}}$	Decision variables specifying the time at which driver h $\in H$ arrives at node $i \in N$

Parameters

$\boldsymbol{cap_h}$	Capacity of driver $h \in H$
$\boldsymbol{Q}$	Capacity of the depot/store
$\boldsymbol{PC_i^p}$	Expected cost for possible failed deliveries associated to customer $i \in C$ in the time-windows p $\in [1, z]$
$\boldsymbol{Shf_h}$	Shift's length of driver $h \in H$
$\boldsymbol{(l_h, u_h)}$	Time windows of driver $h \in H$
$\boldsymbol{t_{i,j}}$	Travel time from node i to node j
$\boldsymbol{q_i}$	Quantity to be delivered to customers i
$\boldsymbol{st_i}$	Service time at node $i \in N$
$\boldsymbol{HC}$	Drivers' hourly fee.
$\boldsymbol{[l_i^p, u_i^p]}$	Available time windows of node $i \in N, with \ i \neq f$, with p $\in [1, z]$
$\boldsymbol{\Delta}$	Length of the time-windows
$\boldsymbol{[T_i, T_f]}$	Delivery time horizon
$\boldsymbol{Dtot}$	Total demand to be delivered
$\boldsymbol{NC}$	Number of customers to be served
$\boldsymbol{OD_{max}}$	Maximum number of drivers available

Appendix B. OBJECTIVE FUNCTION AND CONSTRAINTS

Objective function

$$\min(\sum_{h \in H}(HC * y_h * Shf_h) + \sum_{h \in H} \sum_{i \in N} \sum_{j \in C} \sum_{p=1}^{z} (x_{i,j,h} * PC_i^p * v_{i,h}^p)$$

Constraints

$$\sum_{i \in N} \sum_{j \in C} (x_{i,j,h} * q_j) \leq cap_h y_h \ \ \forall h \in H$$

$$\sum_{i \in N} \sum_{j \in C} \sum_{h \in H} (x_{i,j,h} * q_j) \leq Q$$

$$\sum_{i \in N} \sum_{j \in C} \sum_{h \in H} (x_{i,j,h} * q_j) = Dtot$$

$$\sum_{j \in C} x_{o,j,h} = y_h \ \ \forall h \in H$$

$$\sum_{i \in C} x_{i,f,h} = y_h \ \forall h \in H$$

$$\sum_{i \in N} \sum_{h \in H} x_{i,j,h} = 1 \ \ \forall j \in C$$

$$x_{i,i,h} = 0 \ \ \forall i \in N, \ \forall h \in H$$

$$x_{o,f,h} = 0 \ \ \forall h \in H$$

$$\sum_{j \in N} x_{f,j,h} = 0 \ \ \forall h \in H$$

$$\sum_{i \in N} x_{i,j,h} = \sum_{b \in N, b \neq i} x_{j,b,h} \quad \forall j \in C, \forall h \in H$$

$$\sum_{p=1}^{z} (v_{i,h}^p * l_i^p) \leq \sum_{j \in N} x_{i,j,h} * (\sum_{p=1}^{z} (v_{j,h}^p * l_j^p)) \ \ \forall i \in N \ with \ i \neq f, \forall h \in H$$

$$\sum_{p=1}^{z} v_{i,h}^p = \sum_{j \in N} x_{i,j,h} \ \ \forall i \in N \ with \ i \neq f, \forall h \in H$$

$$s_{i,h} + \sum_{j \in N} (x_{i,j,h} * (t_{i,j} + st_i)) \leq \sum_{j \in N} (s_{j,h} * x_{i,j,h}) \ \ \forall i \in N \ with \ i \neq f, \forall h \in H$$

$$\sum_{p=1}^{z} (v_{i,h}^p * l_i^p) \leq s_{i,h} \leq \sum_{p=1}^{z} (v_{i,h}^p * u_i^p) \ \ \forall h \in H, \forall i \in N \ with \ i \neq f$$

$$\sum_{h \in H} y_h \leq OD_{max}$$

$$l_h \leq s_{o,h} \ \forall h \in H$$

$$\sum_{i \in C} ((s_{i,h} + st_i) * x_{i,f,h}) \leq u_h \ \ \forall h \in H$$

$$x_{i,j,k} \in \{0,1\} \ \forall (i,j) \in N$$

$$y_h \in \{0,1\} \ \forall h \in H$$

$$v_{i,h}^p \in \{0,1\} \ \forall i \in N, \forall p \in \{1, z\}, \forall h \in H$$

$$s_{i,h} \geq 0 \ \forall i \in N, \forall h \in H$$