Modelling and comparison of stability metrics for a re-optimisation approach of the Inventory Routing Problem under demand uncertainty

Faycal Touzout, Anne-Laure Ladier, Khaled Hadj-Hamou

Modeling and comparison of stability metrics for a re-optimisation approach of the Inventory Routing Problem under demand uncertainty

Faycal A. Touzout*, Anne-Laure Ladier, Khaled Hadj-Hamou
Univ Lyon, INSA Lyon, DISP, EA 4570, F-69100 Villeurbanne, France
{faycal.touzout,anne-laure.ladier,khaled.hadj-hamou}@insa-lyon.fr

Abstract

The inventory routing problem (IRP) is an optimisation problem that integrates transportation and inventory management decisions. When subjected to unexpected events such as demand changes, the *a posteriori* approach consists in re-optimising including the data related to this event; the challenge is to ensure that the obtained solution does not deviate too much from the original one, lest that creates important organisational issues. Therefore, a stability metric is needed when re-optimising IRP models. This article proposes a panel of stability metrics adapted from the scheduling, routing and inventory management literature to fit the requirements of the IRP and proposes mathematical formulations for the most relevant ones. A framework of comparison is proposed to validate and compare these metrics over a benchmark of 3000 instances generated from the literature. A strong correlation between the metrics is observed. Moreover, the results show that ensuring the stability of the re-optimised solutions has little impact on the initial objective, the total cost.

Keywords: inventory routing problem; transportation; inventory; re-optimisation; stability metrics; metrics comparison framework

1 Introduction

The inventory routing problem (IRP) integrates two operational problems of the supply chain: inventory management and routing. In the IRP network, a supplier is responsible for managing the inventory and the distribution of a set of clients, to satisfy their demands on a given time

*Corresponding author

horizon. The objective of the decision-maker is to decide, for each period, whether a client should be replenished, with which quantity and following which route, optimising both the inventory and transportation costs.

A common challenge faced by all supply chain operations is the management of uncertainty; this also applies to the IRP. Indeed, the multiplicity of actors and parameters of the IRP makes the range of uncertainty wide, as they can be related to both inventory management and routing components. The clients' demand may change unexpectedly; a driver may be confronted to an unexpected traffic jam, or when arriving at a client's location, the delivery parking slot might be unavailable. By drastically increasing the travelling and service time, such situations may render the plans of the decision-maker unfeasible. Andersson et al. (2010) state that "researchers should focus on development of models and methods that match the industries' need of robust and flexible plans to handle the uncertainties."

Uncertainties can either be handled in an *a priori* or an *a posteriori* fashion. The *a priori* approaches manage the uncertainties in a proactive way by making robust replenishment plans, that will be feasible even when faced with a wide range of events. Their main drawback is usually the cost of protection. As an illustration, note that the most common robust approach makes plans that resist to the worst case scenario. In the IRP, if the clients' demand is uncertain, robust plans can yield a large safety stock which can be very costly. In general with *a priori* approaches, the plans that resist uncertainties are conservative, which makes them expensive, especially when the range of variability of the uncertain parameter is wide.

Due to these limitations, we adopt a different approach by handling uncertainties *a posteriori*: the decision-maker adapts its plan to unexpected events after their occurrence is revealed. One strategy is to repair the initial solution by making small modifications to turn the solution into a feasible one. The difficulty here is the trade-off between the feasibility of the solutions and its cost-efficiency; when more changes occur, the set of possibilities increases exponentially, making repair strategies hard to develop. Another approach is a full re-optimisation of the initial solution, from the time the unexpected event is revealed until the end of the time horizon. The drawback of this approach is that the new solution can be completely different from the initial one. This creates organisational issues: for example, if the new plan requires more vehicles, and therefore more drivers, they might be difficult to find at short notice – also, if the supplier keeps changing his delivery date, the client can lose trust and look for a different supplier. Such situations lead to additional costs that are very difficult to quantify: the decision maker may therefore favour "stable" solutions when re-optimising.

But how to know a solution is stable? Herroelen and Leus (2005) define a stable plan as one that *"deviates as little as possible from the original one"*. This definition being very general, its application in the case of IRP is not straightforward. For example, let us consider a plan which replenishes five clients in a certain sequence on a given day; two solutions are proposed to deal with a demand increase from one client. The first solution keeps the same set of clients in the same sequence but replenishes them with quantities that differ from the initial solution. Another solution removes one client from the route and changes the sequence. Which of these solutions provides the smallest deviation, thus is the most stable? There is no straightforward answer.

This leads to our research question: *how to measure stability in the IRP under demand uncertainty?* Our study focuses on demand uncertainty because it can impact both routing and inventory decisions; this uncertainty is handled in an *a posteriori* approach, through re-optimisation. For this approach to make sense from a practical point of view, it is necessary to *define metrics that can adequately quantify stability for the IRP*. Since the actors and parameters of the IRP are multiple, a single stability metric can only cover a part of the full range of stability. Therefore, ensuring overall stability requires a careful choice of metrics. The correlation of the stability metrics should thus be studied in relation to each other, and to the initial objective – namely, the cost.

The main contributions of this article are threefold. First, we present an exhaustive review of stability metrics in different fields such as routing, inventory management and scheduling. We adapt seven stability metrics for the IRP and propose a mathematical formulation for three of these metrics. Finally, we conduct a comparison of the three formulated stability metrics to investigate the correlation between them and the impact of ensuring stability on the cost of the solutions.

This article is therefore organised as follows. In section 2, we review the different manners found in the IRP literature to deal with uncertainty and stability metrics applied to different operational problems. In section 3, a mathematical formulation of the IRP is proposed. Moreover, show how to re-optimise our model when faced with unexpected events, without stability concerns at this point. In section 4, the proposed stability metrics are presented and the advantages and drawbacks of adapting these measures to the IRP are discussed. Numerical experiments are performed in section 5 to test the proposed mathematical formulations of the stability metrics and compare them after defining a dominance function. The experiments are conducted over a large set of instances adapted to our re-optimisation settings, from benchmarks available in the IRP literature. Finally, section 6 concludes the study and proposes some perspectives for future research.

2 Literature Review

In this paper, we do not pretend to do an exhaustive review of the IRP literature. Therefore, we refer the interested reader to Andersson et al. (2010), Bertazzi and Speranza (2013) and Coelho et al. (2014) for IRP literature reviews that present the different variants and optimisation approaches used.

In this section, we focus on the literature of IRPs under uncertainty and stability metrics in re-optimisation approaches.

2.1 Uncertainty in the IRP literature

We identified 46 articles dealing with uncertainty in IRP, using the keywords *uncertainty, stochastic, re-optimisation, robust* combined to *inventory routing problem*; the corresponding references are summarised in Table 1. The following paragraphs aim at explaining the columns appearing in Table 1. The main sources of uncertainties for the IRP encountered in the literature are listed below.

Demand variation. The demand of the client can change : this source of uncertainty is considered in a vast majority of the articles reviewed, namely 38 out of 46 articles. It is also the case in this paper, that focuses on demand uncertainty because it can have an impact in both inventory and routing decisions.

Vehicle availability variation. A dysfunction of one of the vehicles of the fleet or the unavailability of one of the drivers can occur. Jafarian et al. (2019) and Dong et al. (2018) address this source of uncertainty.

Supplier quantity variation. The quantity available to the supplier at the beginning of each period can change. Dong et al. (2018) addresses this uncertainty.

Travelling time variations. Since routing is a part of the IRP, all the disruptions faced in urban delivery are possible, especially traffic jams or time-dependant travelling times. 11 articles take interest of this source of uncertainty.

Travelling time windows variations. Weather conditions can result in a disruption of the travelling time windows, especially when delivering liquid gases(Cho et al., 2018) and/or in the context of maritime IRP

Many scholars have tackled the different uncertainties of the IRP in an *a priori* manner, as shown in Table 1; However, although re-optimisation has been used for the IRP in a deterministic rolling horizon context (Al-Ameri et al., 2008; Rakke et al., 2011), only Dong et al. (2018) used it to cater for uncertainty issues. In Al-Ameri et al. (2008) and Rakke et al. (2011), the rolling horizon decomposition is used as a matheuristic in order to solve large-sized instances. For Dong et al. (2018) the process is slightly different. The first iteration solves the IRP under stochastic parameters using a stochastic MIP over the whole horizon. Then, for each period of the horizon, new information is revealed. If the solution of the stochastic problem is infeasible given the new information, a full re-optimisation is conducted over the whole horizon, modelling the new information as deterministic parameters while keeping the others stochastic. Once the re-optimised solution is obtained, the horizon is rolled The procedure is then iterated until the end of the horizon. It is worth noting that the re-optimised solution does not take any stability constraints into consideration and can therefore be subject to huge disruptions.

The lack of work in re-optimisation for the IRP might be explained by the fact that re-optimisation can be ill-suited to routing problems. In the case of the VRP, Salavati-Khoshghalb et al. (2019) state that from a cost perspective, re-optimising routing decisions once uncertainties are revealed is a better theoretical alternative to *a priori* approaches. However, re-optimisation yields problems that are computationally challenging to solve and might not preserve consistency in routing operations. However, although routing is an important component of the IRP, its other components provide adjustment variables that are not available in the VRP, such as its multi-period dimension, the availability of an inventory, or possible transshipment. Therefore, we believe the re-optimisation approach for the IRP is a relevant one to investigate, that fills a gap in the existing literature.

	Uncertainties					Approaches		
	Demand	Travelling time	Travelling time windows	Vehicle availability	Supplier quantity	Stochastic optimisation	Robust optimisation	Re-optimisation
Trudeau and Dror (1992)	✓	x	x	x	x	✓	x	x
Nori and Savelsbergh (2002)	✓	x	x	x	x	✓	x	x
Kleywegt et al. (2004)	✓	x	x	x	x	✓	x	x
Christiansen and Nygreen (2005)	✓	x	x	x	x	x	✓	x
Aghezzaf (2008)	✓	✓	x	x	x	x	✓	x
Hvattum and Løkketangen (2009)	✓	x	x	x	x	✓	x	x
Chen and Lin (2009)	✓	x	x	x	x	✓	x	x
Huang and Lin (2010)	✓	x	x	x	x	✓	x	x
Yu et al. (2012)	✓	x	x	x	x	✓	x	x
Solyalı et al. (2012)	✓	x	x	x	x	x	✓	x
Karoonsoontawong and Unnikrishnan (2013)	✓	x	x	x	x	✓	x	x
Rahim et al. (2014)	✓	x	x	x	x	✓	x	x
Juan et al. (2014)	✓	x	x	x	x	✓	x	x
Nolz et al. (2014)	✓	x	x	x	x	✓	✓	x
Bertazzi et al. (2015)	✓	x	x	x	x	✓	x	x
Niakan and Rahimi (2015)	✓	✓	x	x	x	✓	x	x
Agra et al. (2015)	x	✓	✓	x	x	✓	x	x
Malekly (2015)	✓	✓	x	x	x	✓	x	x
Zhang et al. (2015)	x	✓	x	x	x	x	✓	x
Soysal et al. (2015)	✓	x	x	x	x	✓	x	x
Soysal (2016)	✓	x	x	x	x	✓	x	x
Chow (2016)	✓	x	x	x	x	✓	x	x
Bertsimas et al. (2016)	✓	x	x	x	x	x	✓	x
Yang et al. (2016)	✓	x	x	x	x	x	✓	x
Li et al. (2016b)	✓	x	x	x	x	✓	x	x
Li et al. (2016a)	✓	x	x	x	x	x	✓	x
Ercan and Cinar (2017)	✓	x	x	x	x	✓	x	x
Kazemi et al. (2017)	✓	x	x	x	x	x	✓	x
Rahbari et al. (2017)	✓	✓	x	x	x	x	✓	x
Rahimi et al. (2017)	✓	✓	x	x	x	✓	x	x
Agra et al. (2018)	x	x	✓	x	x	x	✓	x
Cho et al. (2018)	x	x	✓	x	x	✓	x	x
Hu et al. (2018)	✓	x	x	x	x	✓	x	x
Micheli and Mantella (2018)	✓	x	x	x	x	✓	x	x
Gruler et al. (2018)	✓	x	x	x	x	✓	x	x
Ghasemi and Bashiri (2018)	✓	x	x	x	x	✓	x	x
Soysal et al. (2018)	✓	x	x	x	x	✓	x	x
Lefever (2018a)	x	✓	x	x	x	✓	✓	x
Dong et al. (2018)	✓	✓	✓	✓	✓	✓	x	✓
Jafarian et al. (2019)	x	x	x	✓	x	✓	x	x
Nikzad et al. (2019)	✓	x	x	x	x	✓	x	x
Rodrigues et al. (2019)	x	✓	x	x	x	✓	✓	x
Fardi et al. (2019)	✓	x	x	x	x	x	✓	x
Diz et al. (2019)	x	x	✓	x	x	x	✓	x
Mirzapour Al-e-hashem et al. (2019)	✓	x	x	x	x	✓	x	x
Lefever et al. (2019)	x	✓	x	x	x	x	✓	x

Table 1: IRP under uncertainty literature overview

2.2 Stability as a re-optimisation criterion

The definition of stability in re-optimisation problems has been tackled for different fields, but not for the IRP, as shown in subsection 2.1. However, since the problem under consideration is an integration of two sub-problems, respectively, inventory management and routing, we can look at the definition of stability in these component problems. The following sections present the related works dedicated to stability in these two fields, in addition to scheduling problems, which are commonly used as reductions for routing problems. Within these fields, a special care is given to problems modelled over a rolling horizon: while they do not carry out re-optimisation as such, they often deal with uncertainty with a real concern for stability.

2.2.1 Stability concerns in scheduling

Stability when re-optimising was first tackled in the scheduling field by Wu et al. (1993). In this work, heuristics are proposed to re-schedule jobs on one machine when a disruption occurs, e.g. a machine failure, with two objectives: efficiency (i.e. makespan) and stability. Two strategies are conducted: first, a full re-optimisation of the unfinished tasks before the disruption. Second, a "right-shift" strategy, where the exact sequence that was to be performed in the original schedule is performed after the disruption, absorbing the idle times if possible in the process. In this context, a stable schedule is one that minimises the sum of the absolute difference of starting times for all tasks as well as preserving the sequence between the re-optimised and the original schedule.

Several articles extend the work done by Wu et al. (1993). Cowling and Johansson (2002) present a bi-objective study that takes utility and stability as optimisation criteria. The utility is defined as the deviation in completion time between the original schedule and the re-optimised one. Stability takes, in addition to the deviation of the starting time, the deviation of the completion time of each task separately. This is because the disruption in this case is not only a machine dysfunction, but can be, among others, a change in the processing time as well. Rangsaritratsamee et al. (2004) present a dynamic job shop rescheduling. policy called "discrete event driven rescheduling" At each rescheduling point, the jobs to be scheduled are the ones that are not scheduled at the previous rescheduling point or the ones that arrived afterwards. The multi-objective efficiency/stability is reconsidered, and in addition to the starting time deviation, a new stability metriccalled "the total deviation penalty". It associates a penalty, restrictively, to jobs rescheduled earlier: the earlier the job is rescheduled, the bigger the penalty.

Curry and Peters (2005) present a simulation study that investigates the trade-off between two conflicting objectives, respectively, a step-wise increasing tardiness cost function and a metric of stability, defined by the proportion of rescheduled jobs that change machine assignment during rescheduling.

Pfeiffer et al. (2007) present a simulation-based stability evaluation of different rescheduling policies, such as: right-shift schedule, complete re-scheduling. Both single and multi-machine job-shop problems are investigated and an industrial application is presented. In this case, the stability metrics considered are the ones presented in Rangsaritratsamee et al. (2004), i.e. starting time deviation and total deviation penalty.

2.2.2 Stability concerns in routing

Sörensen (2006) proposes a bi-objective approach to ensure the stability of the route in a vehicle routing problem. The stability of a route is assessed by the adaptation of the "edit distance" approach introduced by Levenshtein (1966). The idea is to minimise the number of steps to transform a string into another string by a set of operators: addition, deletion and substitution. In the case of the VRP, the strings are the set of routes of the solution.

Schönberger and Kopfer (2008) tackle the problem of "nervousness", defined as "the symptom appearing during the transition from the so far followed schedule to an updated schedule after additional requests appeared". Nervousness problems are classified into two types: external nervousness which affects the client (e.g. a change in the arrival time) and internal nervousness which does not matter for the client (e.g. a re-assignment to another vehicle) Three metrics are used to assess the nervousness, two external and one internal. Mode Selection Nervousness represents the ratio of the clients visited with a different transport mode compared to the original schedule; Resource Assignment Nervousness represents the ratio of the clients that are re-assigned to a different vehicle. The internal metric, Arrival Time Nervousness, represents the ratio of clients for which the arrival time has been modified.

Zhang and Tang (2007) and Wang et al. (2011) take interest in the VRP with time windows under uncertainty. Disruptions for Zhang and Tang (2007) are the unavailability of a vehicle for an interval of time due to vehicle failure or traffic conditions. In Wang et al. (2011), possible disruptions are a modification of the delivery address of a client, a deviation in its time window, a perturbation of its demand, a deletion of a request or any combination of these disruptions. To ensure the stability of the re-optimised solution, both articles use a metric on customer service time where a penalty is applied if the client is served in a time that is outside of its time window. Wang et al. (2011) extend the stability metrics with two other criteria. First, a metric on driving paths where a penalty is applied if an arc that did not exist in the original solution appears in the re-optimised one and vice versa. Second, a metric on delivery costs is the deviation between the costs of the original and the re-optimised solution.

Dettenbach and Ubber (2015) present a mathematical model to re-optimise tours in last mile distribution with electric vehicles fleet. The network is in this case decomposed into multiple districts and all clients of one district are assigned to a vehicle. In case of a failure, which is presented as a dysfunction of one of the electric vehicles or the absence of a driver, the model is supposed to select a back-up district whose clients are to be dispatched and added to the other districts. The stability is modelled as constraints. Each time a district is dispatched, it is decomposed into a certain number of paths, this number being smaller or equal to the number of operational vehicles, i.e. the number of non-dispatched districts. These paths are then inserted in other districts, in the place of one and only one edge, keeping the sequence the same as in the original route.

2.2.3 Stability concerns in inventory management

Uncertainty is a key component in inventory management. Usually, it is tackled using stochastic formulations and/or by sizing a safety stock, which are *a priori* methods.

Re-optimisation is not a field of work in inventory management, because the problem of when and how much to order when demand changes is rather easy : a typical example is Wilson's model (Harris, 1919) whose complexity is in $O(1)$.

Many scholars take interest on problems of inventory management by assessing the impact on the cost while comparing different replenishment strategies or when considering a bad evaluation of the holding cost, referring to it as a "robustness study". Inderfurth (1994) is the first to tackle the problem of "nervousness" in inventory management. In this paper, the nervousness in the context of a rolling planning horizon is shown to be heavily affected by the choice of the inventory policy. A comparison between the (s, S) and (s, nQ) policies is presented. In this context, a robust/stable solution is one that minimises the deviation in the solution cost.

However, what interests us is a solution that minimises the deviation in the solution structure, i.e. the order quantity and order frequency. This definition of stability is implicitly implied in well known replenishment strategies (such as fixed order quantity, economic order quantity, or order up-to-levels) that set replenishment at fixed intervals and/or in fixed quantity. Here the stability is ensured by the solution structure itself, i.e. the fact that we are looking for a fixed interval and/or a fixed quantity. Therefore a stability metric does not seem necessary.

However, in the context of production planning with problems such as Lot Sizing, Master Production Schedule (MPS), Material Requirements Planning (MRP) or Capacity Expansion Planning (CEP), where the notion of inventory management is a key element, stability in the sense of solution structure has been of interest. The work in Sridharan et al. (1988) is the first to tackle the problem of quantifying stability under rolling horizon planning In this paper, stability is quantified by the difference between the quantity planned originally and the re-optimised quantity for the MPS. They propose to weight the metric in order to prioritise the stability of closer periods. This metric has been adapted to other problems (Herrera et al., 2016; Kadipasaoglu and Sridharan, 1997; Kimms, 1998; Narayanan and Robinson, 2010; Sahin et al., 2008).

Kadipasaoglu and Sridharan (1997) adapt the same metric for the MRP and propose a new one based on the deviation in frequency of replenishments. Kimms (1998) adapts the metric to CEP and Lot Sizing. Sahin et al. (2008) extend the work of Kadipasaoglu and Sridharan (1997) and propose a new metric that quantifies the number of deviations in replenishments: a deviation occurs if a client is replenished at a period when it was not supposed to be in the initial plan, and vice versa. Narayanan and Robinson (2010) combine the metrics of Sridharan et al. (1988) and Sahin et al. (2008) for the joint replenishment lot sizing problem. The authors investigate the trade-off between stability and the cost of the obtained solutions. Finally, Herrera et al. (2016) extend the work of Sridharan et al. (1988) for the MPS by proposing sub-metrics of stability that quantify the deviation of the quantity for each period and not only for the whole horizon. Their experiments show that an improvement in stability does not lead to a significant cost increase.

2.2.4 Literature review synthesis

A summary of the conducted literature review of the stability metrics is presented in Table 2. The rows of this table use the names of the stability metrics proposed for the IRP in section 4, and link them to their counterpart in the literature review of the different fields – namely, scheduling, routing and inventory management. The names given in the table cells are those used by the authors, when they differ from the name we use. This shows the terminology appears to depend on the fields, although the metrics are quite similar: hence the need for us to propose names for the IRP metrics.

	Scheduling	Routing	Inventory management
Sequence preservation		Dettenbach and Ubber (2015); Wang et al. (2011)	
Visit deviation			Replenishment deviation (Narayanan and Robinson, 2010; Sahin et al., 2008)
Quantity deviation			Herrera et al. (2016); Kadipasaoglu and Sridharan (1997); Kimms (1998); Narayanan and Robinson (2010); Sahin et al. (2008); Sridharan et al. (1988)
Edit distance		Sörensen (2006)	
Client re-allocation	Machine re-assignment (Curry and Peters, 2005)	Resource re-assignment nervousness (Schönberger and Kopfer, 2008)	
Delivery system modification		Mode selection nervousness (Schönberger and Kopfer, 2008)	
Visiting time deviation	Starting time deviation (Cowling and Johansson, 2002; Pfeiffer et al., 2007; Rangsaritratsamee et al., 2004; Wu et al., 1993)	Arrival time deviation (Schönberger and Kopfer, 2008; Wang et al., 2011; Zhang and Tang, 2007)	

Table 2: A summary of the stability metrics in the literature

3 Mathematical formulation

Prior to the re-optimisation problem, a mathematical formulation of the IRP is presented, based on the work of Archetti et al. (2007).

3.1 Notations

The following notations are used for the formulation of the IRP and throughout the rest of this article:

Definition sets:

- $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a graph where vertex $0 \in \mathcal{V}$ represents the supplier, $\mathcal{V}\backslash\{0\}$ the set of clients and $\mathcal{E}$ the set of edges.
- $\mathcal{H} = \{0, 1, ..., |\mathcal{H}|\}$ is a time horizon where $t \in \mathcal{H}$ represent the index of the period. Note that $t = 0$ represent the initial state.

Client data:

- D_i^t is the demand of client $i \in \mathcal{V}\backslash\{0\}$ at period $t \in \mathcal{H}$.
- s_i^0 represents the initial inventory (at period $t = 0$) of client $i \in \mathcal{V}\backslash\{0\}$.
- $S_i^{\max}$ is the maximum inventory level for client $i \in \mathcal{V}\backslash\{0\}$.

Supplier data:

- R^t is the quantity of products available or produced at supplier $0 \in \mathcal{V}$ for period $t \in \mathcal{H}$.
- s_0^0 represents the initial inventory of the supplier.
- C represents the capacity of the vehicle.

Costs:

- h_i is the holding cost paid for each product in the inventory of the client/supplier $i \in \mathcal{V}$ at the end of period $t \in \mathcal{H}$.
- $c_{i,j}$ is the cost of travelling edge $(i, j) \in \mathcal{E}$.

The mixed integer linear programs that follow use the decision variables listed below:

- $x_{i,j}^t = 1$ if edge $(i, j) \in \mathcal{E}$ is travelled by vehicle at period $t \in \mathcal{H}$, 0 otherwise.
- $y_i^t = 1$ if client $i \in \mathcal{V}\backslash\{0\}$ is visited at period $t \in \mathcal{H}$, 0 otherwise.
- $s_i^t \in \mathbb{R}$ is the inventory level of client $i \in \mathcal{V}\backslash\{0\}$ at the end of period $t \in \mathcal{H}$.
- $q_i^t \in \mathbb{R}$ is the quantity sent to client $i \in \mathcal{V}\backslash\{0\}$ at period $t \in \mathcal{H}$.

IRP					
min	$obj^{\mathrm{IRP}} = c_{i,j} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}, i<j} \sum_{t \in \mathcal{H} \setminus \{0\}} x_{i,j}^t + h_i \sum_{i \in \mathcal{V}} \sum_{t \in \mathcal{H}} s_i^t$				
s.t	s_0^t	$=$	$s_0^{t-1} - \sum_{i \in \mathcal{V} \setminus \{0\}} q_i^t + R^t$	$\forall t \in \mathcal{H} \setminus \{0\}$	(1)
	s_i^t	$=$	$s_i^{t-1} + q_i^t - D_i^t$	$\forall i \in \mathcal{V} \setminus \{0\}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(2)
	s_i^t	$\leq$	$S_i^{\max}$	$\forall i \in \mathcal{V} \setminus \{0\}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(3)
	$q_i^t + s_i^{t-1}$	$\leq$	$S_i^{\max}$	$\forall i \in \mathcal{V} \setminus \{0\}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(4)
	q_i^t	$\leq$	$S_i^{\max} \times y_i^t$	$\forall i \in \mathcal{V} \setminus \{0\}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(5)
	q_0^t	$\leq$	$C \times y_0^t$	$\forall t \in \mathcal{H} \setminus \{0\}$	(6)
	$\sum_{j \in \mathcal{V} \setminus \{0\}} x_{i,j}^t + \sum_{j \in \mathcal{V} \setminus \{0\}} x_{j,i}^t$	$=$	$2 \times y_i^t$	$\forall i \in \mathcal{V}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(7)
	$\sum_{i \in \mathscr{S}} \sum_{j \in \mathscr{S}, i<j} x_{i,j}^t$	$\leq$	$\lvert \mathscr{S} \rvert - 1$	$\forall \mathscr{S} \subseteq \mathcal{V} \setminus \{0\}$, $t \in \mathcal{H} \setminus \{0\}$	(8)
	$x_{i,j}^t$	$\in$	$\{0, 1\}$	$\forall i, j \in \mathcal{V}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(9)
	y_i^t	$\in$	$\{0, 1\}$	$\forall i \in \mathcal{V}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(10)
	q_i^t	$\geq$	0	$\forall i \in \mathcal{V} \setminus \{0\}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(11)
	s_i^t	$\geq$	0	$\forall i \in \mathcal{V}$, $\forall t \in \mathcal{H} \setminus \{0\}$	(12)

3.2 Mathematical formulation of the IRP

The objective computes the total travelling cost and the total holding cost for $i \in \mathcal{V}$ for the whole time horizon $\mathcal{H}$.

Constraints (1) are flow conservation constraints that compute the inventory level of the supplier at each period $t \in \mathcal{H} \setminus \{0\}$ from its previous inventory level, the quantity produced at t and the quantities sent to the clients at t. Similarly, Constraints (2) state the flow conservation constraints regarding the clients. They compute the inventory level of each client $i \in \mathcal{V} \setminus \{0\}$ for each period $t \in \mathcal{H} \setminus \{0\}$ from its previous inventory level, the quantity received from the supplier and its demand for period t.

The inventory capacity is enforced through several constraints: Constraints (3) state that the inventory level of client $i \in \mathcal{V} \setminus \{0\}$ at any period $t \in \mathcal{H}$ must be lower than $S_i^{\max}$, and Constraints (4) state that a replenishment of this client at period $t \in \mathcal{H} \setminus \{0\}$ cannot exceed its maximal inventory level.

Constraints (5) link variables y_i^t with q_i^t, stating that a client $i \in \mathcal{V}\backslash\{0\}$ which receives a quantity at period $t \in \mathcal{H}\backslash\{0\}$, is necessarily visited at t. $S^{\max}$ is used here as an upper bound for q_i^t. Constraints (6) works similarly for the supplier, stating that the quantity leaving supplier 0 at period $t \in \mathcal{H}\backslash\{0\}$ is limited by the vehicle capacity C.

Constraints (7) state that if a location is visited, it is entered and left once. Constraints (8) eliminates sub-tours. Finally, constraints (9) to (12) enforce integrality and non-negativity conditions on the variables.

In order to solve the IRP and the different re-optimisation models of the IRP that will be proposed in the next sections, a branch-and-cut algorithm is used. The mathematical formulation presented previously is strengthened with additional valid inequalities presented in Archetti et al. (2007), Coelho and Laporte (2014) and Desaulniers et al. (2016). The bounds are improved and the routing component of the IRP re-formulated according to the work of Lefever (2018b). Finally, the subtour elimination constraints are added dynamically into a branch-and-cut procedure.

It is worth noting that our contribution does not lie in the solving method, as the branch-and-cut procedure used only combines state-of-the-art cuts and reformulations of the IRP in order to solve the problem efficiently.

3.3 An illustrative example

Let us consider a small instance w of the IRP where $\mathcal{V} = \{0, 1, 2, ..., 5\}$ is constituted of five clients besides supplier 0, time horizon $|\mathcal{H}| = 3$ and vehicle capacity $C = 150$. Table 3 lists all data related to instance w. The columns represent, respectively, the indices i of the supplier/clients, the coordinates $(x_i; y_i)$, the initial inventory s_i^0, the maximum inventory $S_i^{\max}$, the quantity R^t available to supplier 0 at each time period of horizon $\mathcal{H}$, the demand d_i^t of clients $\{1, 2, ..., 5\}$ for each time period of horizon $\mathcal{H}$, and finally the holding costs. The length of each edge approximates the cost of travelling the edge. This Euclidean distance is computed from the coordinates of the locations.

Table 4 and Figure 1 present an optimal solution for instance w. Table 4 presents the inventory levels at the end of each period of each location and the quantities sent from the supplier to each client for the whole time horizon, and Figure 1 gives a graphical overview of the same

i	$(x_i; y_i)$	s_i^0	S_i^{max}	R^t			D_i^t			h_i
				$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$	
0	(154;417)	510	$+\infty$	193	193	193				0.03
1	(172;334)	40	130				65	65	65	0.02
2	(267;87)	20	70				35	35	35	0.03
3	(148;433)	58	116				58	58	58	0.03
4	(355;444)	24	48				24	24	24	0.02
5	(38;152)	11	22				11	11	11	0.02

Table 3: A representation of the data of instance w

solution. In period $t = 1$, deliveries are made to clients 1, 2 and 5 in this order. For $t = 2$, the supplier replenishes clients 1, 3 and 4. Finally, for period $t = 3$, clients 1, 2, 3 and 5 are visited.

Note that in Table 4 the inventory level is given at the end of the period, whereas Figure 1 shows it at the beginning of the period. This choice was made to improve the figure readability; the inventory level at the end of the period in Figure 1 can be computed by adding the quantity received to the inventory level at the beginning of the period, minus the demand of the client for that period.

i	s_i^t			q_i^t		
	$t=1$	$t=2$	$t=3$	$t=1$	$t=2$	$t=3$
0	553	596	639			
1	64	43	24	89	44	46
2	35	0	0	50	0	35
3	0	0	0	0	58	58
4	0	24	0	0	48	0
5	11	0	0	11	0	11

Table 4: An optimal solution of instance w

This example will be used in the rest of the article to illustrate the applicability of some of the stability metrics proposed.

3.4 Cost-based re-optimisation

Let us now consider that the demand of the clients are subject to unexpected changes. Each of such changes is modelled as a deterministic event $E = \{p, \mathcal{D}^E\}$ where $p \in \mathcal{H}$ represents the period for which the event happens and $\mathcal{D}^E$ represents a set of modified demands. Note that the event always occurs at the beginning of period p.

An example of an event for instance w is presented in Table 5. Event E in this case occurs

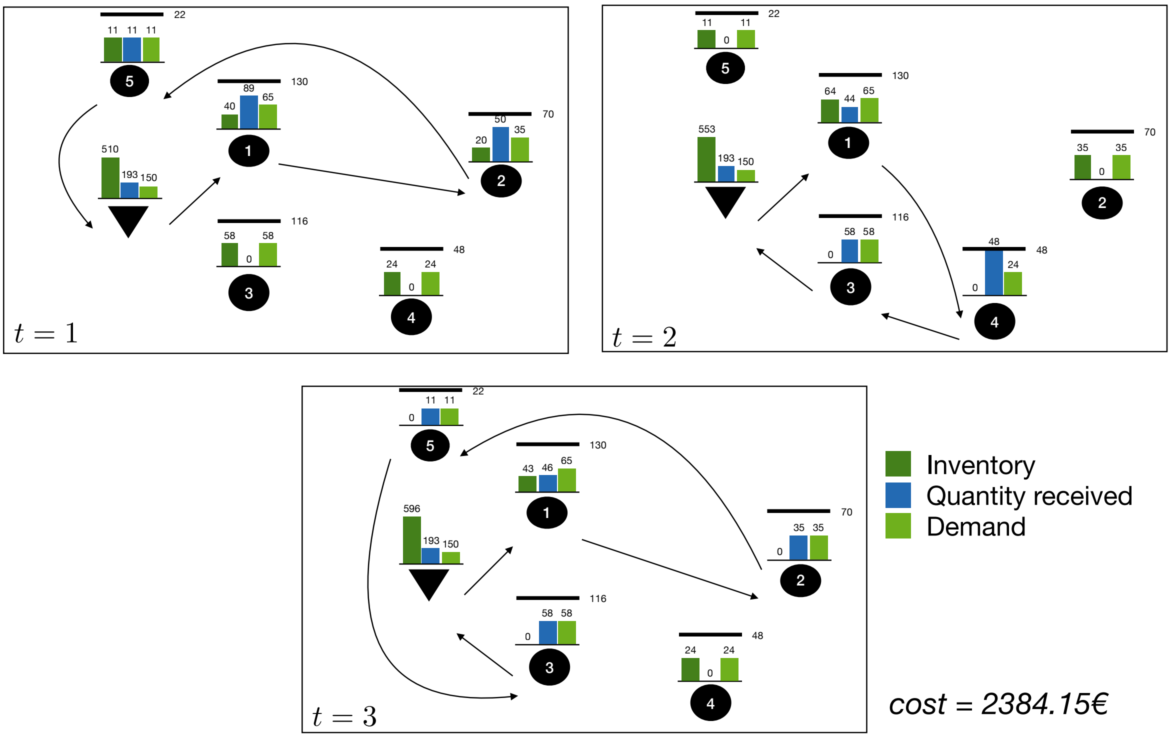


Figure 1: A representation of an optimal solution of instance w

at period $p = 2$. Therefore, a re-optimisation is needed for the second and third periods. The demands that are modified compared to Table 3 are shown with bold characters. For example, for client 4, demands in periods $t = 2$ and $t = 3$ increased to 48 instead of 24, whereas for client 1, demand for period $t = 2$ turned to 0 instead of 65.

i	D_i^t		
	$t = 1$	$t = 2$	$t = 3$
1	65	**0**	65
2	35	35	**70**
3	58	**0**	58
4	24	**48**	**48**
5	11	**0**	11

Table 5: A representation of an event for instance w

In order to adapt to these new demands, a re-optimisation is needed for horizon $\mathcal{H}^{\text{new}} = \{p, p+1, ..., |\mathcal{H}|\}$. To mathematically formulate the re-optimisation problem, let $\mathcal{S} = \{X_{i,j}^t, Y_i^t, Q_i^t, S_i^t\}$ and $\mathcal{S}^{\text{new}} = \{x_{i,j}^t, y_i^t, q_i^t, s_i^t\}$ be, respectively, a set of data representing the original solution for the deterministic problem and a set of decision variables for the re-optimisation problem.

Compared to the original model, the cost-based re-optimisation problem time horizon $\mathcal{H}$ is

Cost-based re-optimisation (IRPR)

$$\min \quad obj^{\text{IRP}} = c_{i,j} \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{V}, i<j} \sum_{t\in\mathcal{H}\setminus\{0\}} x_{i,j}^t + h_i \sum_{i\in\mathcal{V}} \sum_{t\in\mathcal{H}} s_i^t$$

$$\begin{array}{llcll}
\text{s.t} & s_0^t & = & s_0^{t-1} - \sum_{i\in\mathcal{V}\setminus\{0\}} q_i^t + R^t & \forall t \in \mathcal{H}^{\text{new}} \quad (13)\\
& s_i^t & = & s_i^{t-1} + q_i^t - D_i^t & \forall i \in \mathcal{V}\setminus\{0\}\ ,\ \forall t \in \mathcal{H}^{\text{new}} \quad (14)\\
& s_i^t & \leq & S_i^{\max} & \forall i \in \mathcal{V}\setminus\{0\}\ ,\ \forall t \in \mathcal{H}^{\text{new}} \quad (15)\\
& q_i^t + s_i^{t-1} & \leq & S_i^{\max} & \forall i \in \mathcal{V}\setminus\{0\}\ ,\ \forall t \in \mathcal{H}^{\text{new}} \quad (16)\\
& q_i^t & \leq & S_i^{\max} \times y_i^t & \forall i \in \mathcal{V}\setminus\{0\}\ ,\ \forall t \in \mathcal{H}^{\text{new}} \quad (17)\\
& q_0^t & \leq & C \times y_0^t & \forall t \in \mathcal{H}^{\text{new}} \quad (18)\\
& \sum_{j\in\mathcal{V}\setminus\{0\}} x_{i,j}^t + \sum_{j\in\mathcal{V}\setminus\{0\}} x_{j,i}^t & = & 2 \times y_i^t & \forall i \in \mathcal{V}\ ,\ \forall t \in \mathcal{H}^{\text{new}} \quad (19)\\
& \sum_{i\in\mathscr{S}} \sum_{j\in\mathscr{S}, i<j} x_{i,j}^t & \leq & |\mathscr{S}| - 1 & \forall \mathscr{S} \subseteq \mathcal{V}\setminus\{0\}\ ,\ t \in \mathcal{H}^{\text{new}} \quad (20)\\
& \text{(9) to (12)} & & & \\
& x_{i,j}^t & = & X_{i,j}^t & \forall (i,j) \in \mathcal{E}\ ,\ \forall t \in \mathcal{H}^{\text{fixed}} \quad (21)\\
& y_i^t & = & Y_i^t & \forall i \in \mathcal{V}\ ,\ \forall t \in \mathcal{H}^{\text{fixed}} \quad (22)\\
& s_i^t & = & S_i^t & \forall i \in \mathcal{V}\ ,\ \forall t \in \mathcal{H}^{\text{fixed}} \quad (23)\\
& q_i^t & = & Q_i^t & \forall i \in \mathcal{V}\ ,\ \forall t \in \mathcal{H}^{\text{fixed}} \quad (24)
\end{array}$$

decomposed in two parts: $\mathcal{H}^{\text{fixed}} = \{0, 1, \ldots, p-1\}$ and $\mathcal{H}^{\text{new}} = \{p, p+1, ..., |\mathcal{H}|\}$. In the first part of this time horizon, i.e. before the occurrence of the event, constraints (21) to (24) are added. They fix the decision variables of $\mathcal{S}^{\text{new}}$ to the values taken by the decision variables in $\mathcal{S}$. For the second part of the time horizon, i.e. after the event, the constraints are kept the same as in the IRP model.

Figure 2 shows an optimal solution for instance w faced with the event presented in Table 5. This solution is obtained using the model formulated above. The variables regarding period 1 in $\mathcal{S}^{\text{new}}$ are fixed to the values determined in $\mathcal{S}$, since the event occurs at period 2. During $t = 2$, client 1 is no longer visited in comparison to the initial solution $\mathcal{S}$. Finally, for $t = 3$, clients 1, 2 and 4 are visited in this order, instead of 1, 2, 5 and 3.

The problem with this new solution $\mathcal{S}^{\text{new}}$ is that it can be perceived as too different from solution $\mathcal{S}$ in terms of solution structure. Therefore, stability metrics are needed to reduce this

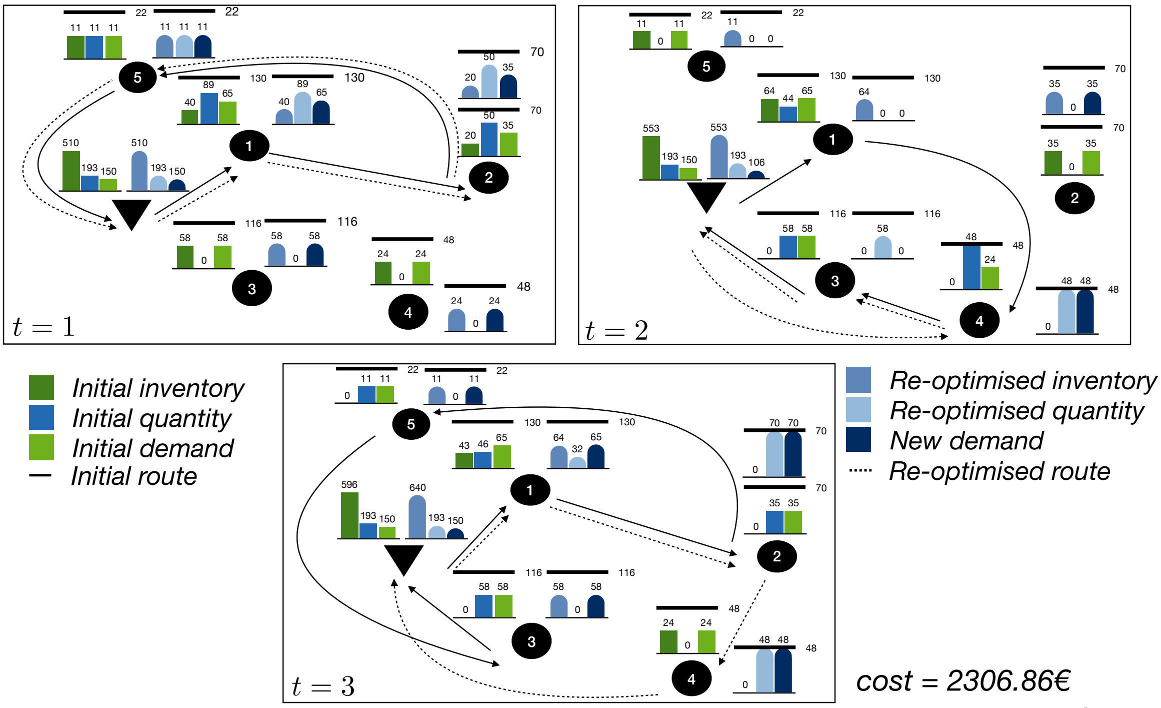


Figure 2: A representation of an optimal solution for the re-optimised instance w

difference.

4 Adaptation and mathematical formulation of stability metrics

In this section, we discuss the relevance for the IRP of the stability metrics found in the literature review carried out in subsection 2.2. The advantages and drawbacks of these each are analysed. Only those that are judged relevant for the IRP are mathematically formulated and illustrated using the example of subsection 3.3.

There are several ways to include these metrics in the mathematical formulation of the cost-based re-optimisation presented in subsection 3.4: either as a single objective, as hard constraints or as soft constraints, i.e. integrated to the objective as penalties to be paid each time the metric is violated. In this section, the former option is used: the objectives of the mathematical formulations proposed aim at minimising the violations of the metrics.

4.1 Sequence preservation

Description: Sequence preservation is an important stability metric used in both scheduling (Wu et al., 1993) and routing problems (Dettenbach and Ubber, 2015). In the literature, it is used as an additional constraint where the sequence (or part of the sequence) of an original solution must be preserved in the re-optimised solution (Dettenbach and Ubber, 2015).

Let us consider an example where clients $\{a, b, c\}$ are visited in the original solution $\mathcal{S}$ at time t with route $R = \{a - b - c\}$. The event adds two more clients that should be visited at time t, thus the set of clients to visit in this route becomes $\{a, b, c, d, e\}$. A subset of possible solutions for the re-optimised case would be routes: $R_1^{\text{new}} = \{a-b-c-d-e\}$, $R_2^{\text{new}} = \{a-b-c-e-d\}$ and $R_3^{\text{new}} = \{a - d - b - e - c\}$. A stable solution in this case can be defined by one that minimises the number of violations of the sequence. To compute this number, the edges taken in solution $\mathcal{S}$ are compared with the ones taken in solution $\mathcal{S}^{\text{new}}$. For example, for R_1^{new} and R_2^{new} the number of violations of the sequence would be equal to 0 since the edges between a and b and between b and c exist in $\mathcal{S}^{\text{new}}$. However, for R_3^{new} it would be equal to 2 since both edges are missing.

Advantages: In urban delivery the products to deliver are generally stored inside the truck following the sequence of the solution so that the products of the first client to visit are the most accessible. Disturbing the original sequence when re-optimising would result in an increase of the visiting time and therefore the duration of the route. Preserving the sequence avoids such issues. However, this is only relevant for the period of occurrence of event p. Another advantage is that, as drivers are given the initial plans, they may plan other operations accordingly. Changing the sequence may have an impact on these operations. Finally, the other advantage is implicit, as preserving the sequence can preserve the visiting times of the clients as well.

Drawbacks: A sequence changes only when clients are added to or removed from a route, compared to the original solution. Re-optimising the IRP with sequence preservation therefore implies solving a travelling salesman problem (TSP) when a vertex is inserted, removed or substituted. Because it is well known that an efficient solution of such a modified TSP can be completely different from a solution of the original one (Archetti et al., 2003), preserving the sequence can have a huge impact on the cost of the solution. Another drawback of this metric is that it is mainly external since it has no effect whatsoever on the client.

Mathematical formulation: Let us now mathematically formulate the sequence preservation metric for the IRP. Let $z_{i,j}^t$ be a binary variable that is equal to 1 if there is a sequence violation, i.e. if edge $(i, j) \in \mathcal{E}$ is used in a route at time t in solution $\mathcal{S}$ but not in $\mathcal{S}^{\mathbf{new}}$, or if it is used in $\mathcal{S}^{\mathbf{new}}$ but not in $\mathcal{S}$.

Sequence preservation-based re-optimisation (IRPR-SP)				
min	$\sum_{t\in\mathcal{H}^{\text{new}}}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V},i<j} z_{i,j}^t$			
s.t.	(9) to (24)			
	$z_{i,j}^t = \lvert X_{i,j}^t - x_{i,j}^t\rvert$	$\forall t \in \mathcal{H}^{\text{new}}$, $\forall i \in \mathcal{V}$, $\forall j \in \mathcal{V}, i < j$		(25)
	$z_{i,j}^t \in \{0,1\}$	$\forall t \in \mathcal{H}^{\text{new}}$, $\forall i \in \mathcal{V}$, $\forall j \in \mathcal{V}, i < j$		(26)

The IRPR-SP model modifies the formulation of the IRPR by changing the objective function (which becomes the minimisation of the number of sequence violations) and extending it with constraints (25) and (26). Constraints (25) count one violation of sequence if an edge (i, j) is taken in $\mathcal{S}$ and not in $\mathcal{S}^{\text{new}}$ and vice versa; they are not linear, but we note that a constraint $z = |x - y|$ can be linearized as follows if z appears in the minimization objective:

$$z = |x - y| \Leftrightarrow \begin{cases} z \geq x - y \\ z \geq y - x \end{cases}$$

Constraints (26) ensure the binarity of variables $z_{i,j}^t$.

Figure 3 represents an optimal solution using the model IRPR-SP. As in Figure 2, the re-optimisation process starts at the second period. In period $t = 2$, the same sequence is preserved. However, for $t = 3$, there is one sequence violation, due to the addition of client 4 to the tour. Indeed, in this case, the arc $1 - 2$ is no longer taken and is replaced by the sequence $1 - 4 - 2$.

4.2 Visit deviation

Description: Visit deviation is a metric we adapt for the IRP and based on the metric presented in Sahin et al. (2008). This metric is not applicable to purely routing problems such as TSP and VRP. Indeed, due to the time dimension of the IRP, it is possible not to visit a client for a certain period, as long as its demand is satisfied, which is not the case for the problems cited previously. The visit deviation quantifies the number of clients that are visited in the re-optimised solution whereas they were not supposed to be in the original one and vice-versa. This metric is mainly external according to Schönberger and Kopfer (2008)'s classification, i.e. it is designed to cater for clients satisfaction.

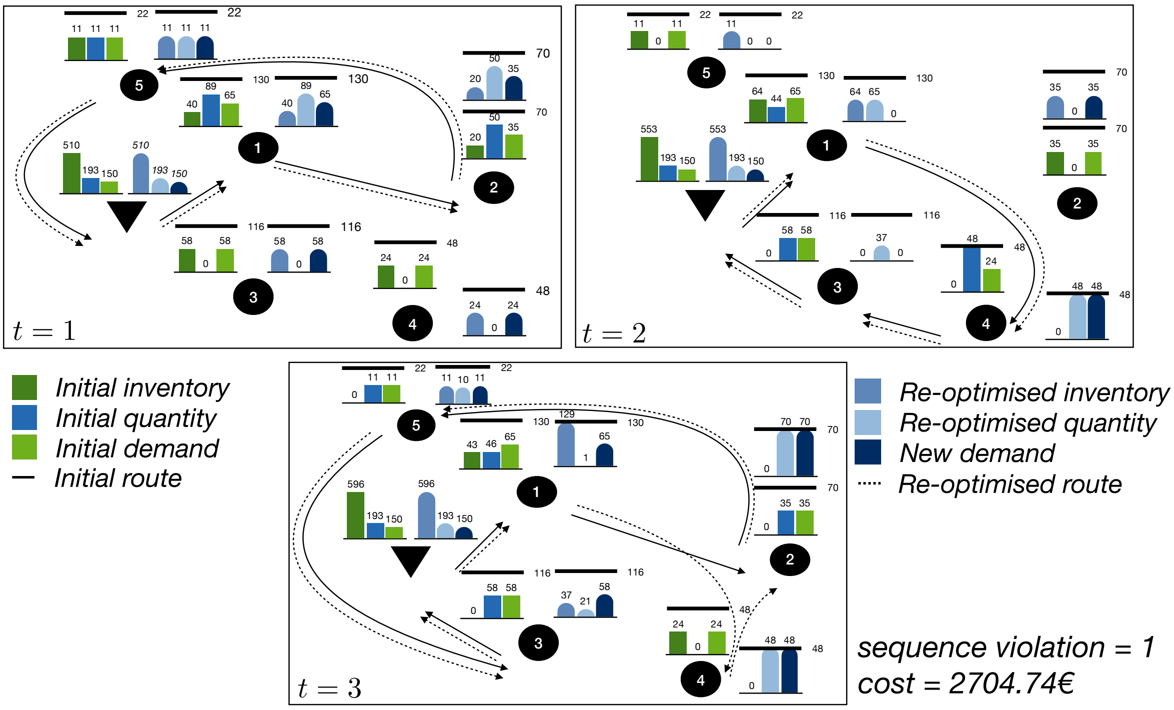


Figure 3: An optimal solution for instance w with minimum sequence violations

Advantages: A client which is visited when it was not supposed to be may face some planning issues related to the unavailability of resources such as the human workforce, machinery, parking slots... Conversely, expecting a visit that does not actually occur causes a waste of time and resources. Therefore, minimising the visits violation minimises the disruptions for the client and increases the reliability of the supplier from the clients' point of view.

Drawbacks: Because the visit deviation metric is mainly external, i.e. favours the client's point of view, it does not explicitly ensure any internal stability. Therefore, it cannot guarantee an efficient routing. However, it is worth noting that the visit deviation metric does ensure internal and external stability in inventory management in an implicit way, since not visiting a client that was supposed to be resupplied, or visiting a client that was not supposed to be resupplied, has a negative impact on the stability of the inventory of both the supplier and the client.

Another drawback of the metric is that, in some cases, it can be counterproductive: minimising the visit deviation can lead to a client being visited for no reason (no product being delivered) other than to keep it visited.

Mathematical formulation: Let z_i^t be a binary variable that is equal to 1 if there is a visit deviation for client i at period t, i.e. if it is visited while it was not supposed to be, or if it is not visited while it was supposed to be.

Visit deviation-based re-optimisation (IRPR-VD)

$$\min \sum_{t\in\mathcal{H}^{\text{new}}} \sum_{i\in\mathcal{V}\setminus\{0\}} z_i^t$$

s.t. (9) to (24)

$$z_i^t = |Y_i^t - y_i^t| \quad \forall t \in \mathcal{H}^{\text{new}}, \forall i \in \mathcal{V}\setminus\{0\} \tag{27}$$

$$z_i^t \in \{0,1\} \quad \forall t \in \mathcal{H}^{\text{new}}, \forall i \in \mathcal{V}\setminus\{0\} \tag{28}$$

The objective of the IRPR-VD is the minimisation of the number of visit deviations. It extends the IRPR with constraints (27) and (28). Constraints (27) define a deviation for a client i in period t as happening if it is visited in $\mathcal{S}$ and not visited in $\mathcal{S}^{\text{new}}$ or vice-versa. It can be linearized as explained before. Constraints (28) ensure the binarity of variables z_i^t.

An optimal solution of the IRPR-VD on our example instance w is presented in Figure 4. In period 2 the same clients are visited as in $\mathcal{S}$. For period 3, client 4 is visited in addition to clients 1, 2, 3 and 5. Therefore, there is one visit deviation in total, the objective is 1.

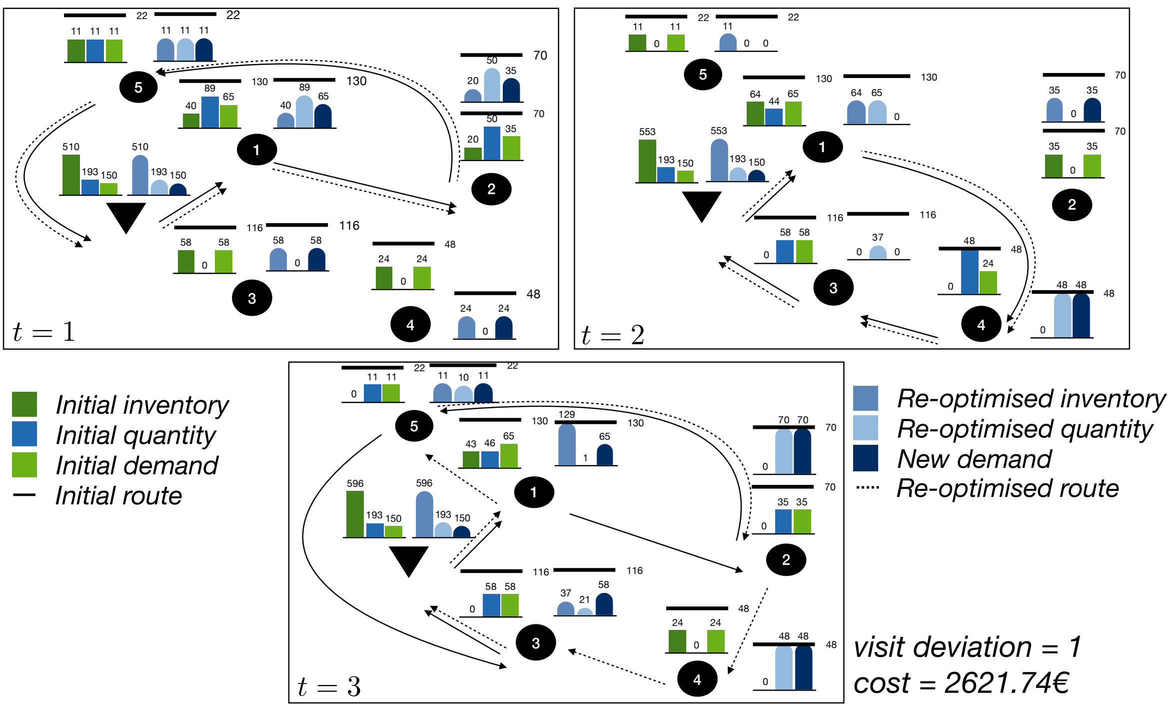


Figure 4: An optimal solution for instance w with minimum visit deviations

4.3 Quantity received deviation

Description: We propose the quantity received deviation as a stability metric for the IRP, in order to address the lack of stability metrics in inventory management previously emphasised in the literature review. This metric computes the difference between the quantity that was supposed to be sent in the original solution and the quantity received in the re-optimised one. This metric is meant to improve the service quality, i.e. it is mainly external.

Advantages: A good point is that when the quantity received deviation is kept at a minimum, fewer planning issues are faced. Indeed, a client receiving less quantity than planned has uselessly mobilised costly resources for this process. Conversely, if the client receives more products than planned, there might be a shortage of resources which may increase the service time, thus disrupting the entire delivery plan.

This metric is the only one that handles explicitly the inventory management component of the IRP.

Drawbacks: The logic behind VMI is to let the decision maker, who has the best overview of the network, decide whom to serve and with which quantity. However, when limiting the value of the quantity deviation metric, such flexibility is somewhat constrained. Note that, although this drawback exists for all the metrics proposed, we believe that it is much more significant in the case of quantity deviation.

Mathematical formulation: let $\tilde{q}_i^t \in \mathbb{R}$ be the quantity received deviation, i.e. the difference of quantity received by client i at time t, between the original solution and the re-optimised one.

Quantity deviation-based re-optimisation (IRPR-QD)

$$\min \sum_{t \in \mathcal{H}^{\text{new}}} \sum_{i \in \mathcal{V} \setminus \{0\}} \tilde{q}_i^t$$

s.t. (9) to (24)

$$\tilde{q}_i^t = |Q_i^t - q_i^t| \quad \forall t \in \mathcal{H}^{\text{new}}, \forall i \in \mathcal{V} \setminus \{0\} \tag{29}$$

$$\tilde{q}_i^t \geq 0 \quad \forall t \in \mathcal{H}^{\text{new}}, \forall i \in \mathcal{V} \setminus \{0\} \tag{30}$$

The objective of the IRPR-QD is to minimise the total quantity deviations. Compared to the IRPR, it adds constraints (29) and (30). Constraints (29) compute the difference between the quantity planned in solution $\mathcal{S}$ and received in $\mathcal{S}^{\text{new}}$. These constraints can easily be linearised as mentioned before. Constraints (30) ensure the non-negativity of variables $\tilde{q}_i^t$.

An optimal solution of the IRP-QD on instance w is presented in Figure 5. The differences in quantity occur in period $t = 3$ just as in SP and VD. For example, client 1 receives 22 instead of 46. The total amount of the differences over all the clients sums up to 166.

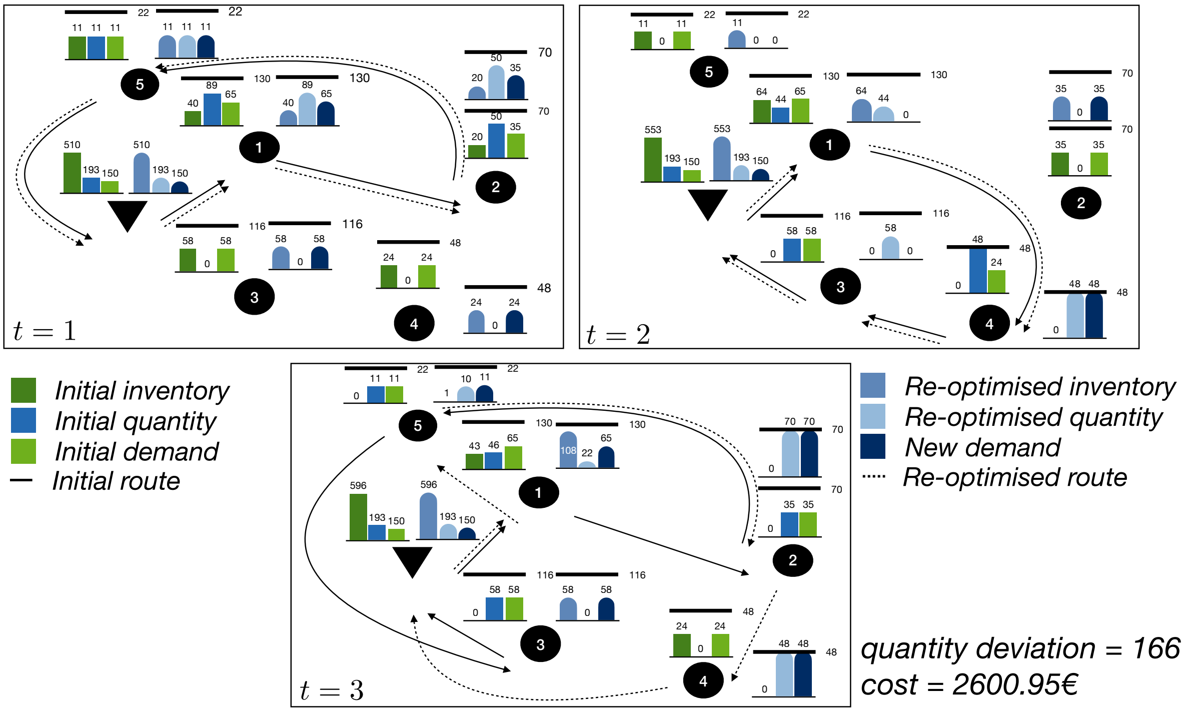


Figure 5: An optimal solution for instance w with minimum quantity deviation

4.4 Clients re-allocation

Description: When re-optimising, a way to achieve stability is to minimise the re-allocations of clients to drivers from other areas. The computation of the number of re-allocations is as follows: one occurs for each visited client in $\mathcal{S}$, if the driver is no longer the same in $\mathcal{S}^{\text{new}}$. This metric is an adaptation of the "Resource Assignment Nervousness" presented in Schönberger and Kopfer (2008).

Advantages: When the clients re-allocation metric is kept at a minimum, the driver is less disturbed by the events, which enhances his/her performance as well as the satisfaction of the client.

Drawbacks: In the IRP, clients that were not visited in the original plan can be added to a route in the re-optimised solution, or clients that were previously visited can be removed from the route. This will not affect the clients re-allocation metric at all, since the clients re-allocation metric changes only when a client is visited in both $\mathcal{S}$ and $\mathcal{S}^{\text{new}}$ but with different vehicles. For example, let $\mathcal{K}_1 = \{a, b, c, d\}$ and $\mathcal{K}_2 = \{e, f, g\}$ be the sets of clients visited by vehicle 1 and 2, respectively, in $\mathcal{S}$. If in $\mathcal{S}^{\text{new}}$, $\mathcal{K}_1^{\text{new}} = \{a, b, c\}$ and $\mathcal{K}_2^{\text{new}} = \{e, f, g, h\}$, the number of visited clients that are allocated to different vehicles is equal to zero, even if client d has been removed and client h added. This example shows the limitations of this metric, that does not detect some big differences between $\mathcal{S}$ and $\mathcal{S}^{\text{new}}$.

We do not propose a mathematical formulation for the client re-allocation metric because we chose to focus on a variant of the IRP where only one vehicle is used (see model in subsection 3.4), whereas the client re-allocation metric is only compatible with multi-vehicle variants.

4.5 Edit distance

Description: As mentioned in subsection 2.2, the idea behind the edit distance in Sörensen (2006) is to minimise the number of operations (addition, deletion, substitution) that transform the original solution into the re-optimised one. This metric is easily adapted to the case of the IRP.

Advantages: The advantage of this stability metric is mostly computational: the operators used to compute the edit distance (addition, deletion and substitution) could be used within a local search procedure. Local search algorithms are known to be very efficient heuristics for sequencing problems, therefore they can help solving routing component of the IRP.

Drawbacks: Let us take the example of a time period in which to visit client $\{a, b, c\}$ should be visited, with route $R = \{a - b - c\}$ as initial solution. Considering the addition of two clients $\{d, e\}$, routes $R_1^{\text{new}} = \{a - b - c - d - e\}$ and $R_2^{\text{new}} = \{a - d - b - e - c\}$ can both be obtained by performing two addition operations. For R_1^{new}, d and e are inserted to the end of the sequence, whereas for R_2^{new}, d and e are inserted, respectively, between $a - b$ and $b - c$. One could argue that R_1^{new} represents a more stable solution than R_2^{new} as it does not change the initial sequence. However, the edit distance metric cannot differentiate between them.

Another drawback of the edit distance metric is that it is generally considered as an indicator of stability and not an active optimisation criterion. This is due to the NP-hardness of defining the optimal edit distance for $\mathcal{S}^{\text{new}}$ in relation to $\mathcal{S}$ (Sörensen, 2006). Therefore, we do not see how to propose a linear model that optimises the edit distance and do not propose a mathematical formulation for the edit distance metric.

4.6 Delivery system modification

Description: Delivery system modification as a metric can be considered only when deliveries can be subcontracted, e.g. for the IRP with transshipment. It is adapted from the "Mode Selection Nervousness" metric presented in Schönberger and Kopfer (2008). It quantifies variations in the mode of delivery between the original solution and the re-optimised one by counting the number of clients that are visited by the supplier's own fleet in $\mathcal{S}^{\text{new}}$ rather than a subcontractor, or that are visited by a subcontractor instead of the supplier's fleet.

Advantages: This metric helps limiting the changes in resources between the original and the re-optimised solutions. Indeed, the decision maker has to determine how many drivers/owned vehicles he/she will need to perform the routes of the solution, as well as the number of transshipments. Minimising the modifications in the delivery system should avoid contractual issues with the transportation provider if deliveries that were supposed to be outsourced are actually managed internally, or vice versa.

Drawbacks: The main drawback of the delivery system modification metric is that it can only be considered within a variant of the IRP, i.e. when transshipment is possible. In addition, it does not ensure any external stability. Indeed, it does not usually matter for a client whether the delivery is made by a subcontractor or an owned vehicle, as long as the right quantity is received.

The delivery system modification metric is compatible only with variants of the IRP where transshipment is considered. Therefore, no mathematical formulation is proposed for this metric.

4.7 Visiting time deviation

Description: The visiting time deviation metric can be considered when time-related considerations, such as travel times, arrival times, etc, are explicitly modelled in the IRP. This metric sums the differences, for each client, between the time of visit in the initial solution and the re-optimised one.

Advantages: The advantage of this metric is twofold. First, when the visiting time deviations are at a minimum, the client is more satisfied, since the resulting plan does not disturb too much its schedule regarding closing time, other deliveries, unavailability of resources... Second, it ensures, indirectly, some internal stability. Indeed, in order to minimise the deviations in the time of visit, the paths of the drivers must not undergo heavy changes.

Drawbacks: The only drawback of this metric is that it does not ensure any stability in terms of inventory management. However, it is very efficient for the routing component of the IRP since it ensures both internal and external stability.

The visiting time deviation as a metric is only relevant when the time aspect is considered for the routing component. Therefore, we do not propose a mathematical formulation due to its incompatibility with our variant of the IRP.

4.8 A qualitative discussion

The panel of stability metrics presented in this section ensure a large range of stability for the IRP and are compatible with different variants. As a summary, Table 6 shows, for each stability metric proposed, if an internal or external stability in both routing and inventory management is ensured, if it is compatible with the three most commonly used variants of the IRP in the literature, respectively, IRP (multi-vehicle included), IRP with transshipment (IRPT) and IRP with time windows (IRPTW) and finally, if a Mixed Integer Linear Programm (MILP) is proposed for the metric. For example, visit deviation ensures external stability in both routing and inventory management and internal stability in inventory management. It is also compatible with every variant of the IRP. For the sequence preservation on the other hand, although it can be compatible with all the variants of the IRP, it only ensures internal routing stability. The last column shows that MILP models are proposed for three metrics only, namely, sequence preservation, visit deviation and quantity deviation. This is due to the incompatibility of some of the metrics with the variant of the IRP proposed in this paper (client re-allocation, delivery system modification and visiting time deviation) or to their non linearity (edit distance).

	Routing		Inv. management		Problem variants			Proposed
	Internal	External	Internal	External	IRP	IRPT	IRPTW	MILP model
Sequence preservation	✓	x	x	x	✓	✓	✓	✓
Visit deviation	x	✓	✓	✓	✓	✓	✓	✓
Quantity received deviation	x	x	✓	✓	✓	✓	✓	✓
Clients re-allocation	x	✓	x	x	✓	✓	✓	x
Edit distance	✓	x	x	x	✓	✓	✓	x
Delivery system modification	✓	x	x	x	x	✓	x	x
Visiting time deviation	✓	✓	x	x	x	x	✓	x

Table 6: Categorisation of the IRP stability metrics

It appears from Table 6 that no stability metric is able to ensure internal and external stability in both routing and inventory management. Therefore, ensuring the full range of stability for a certain variant of the IRP can be a tedious task. For example, combining sequence preservation, visit deviation and quantity deviation does ensure the full range of stability. However, including all of them to an optimisation model comes with its challenges. If the metrics are considered as hard constraints, it is very complicated to determine a thresh-hold of stability for the metrics that ensures that a feasible solution can be obtained. On the other hand, if they are considered as soft constraints, it is difficult to define an appropriate weight for each metric. Therefore, in the next section, we propose a way to compare the stability metrics by investigating their behaviour in relation to each other and to the initial objective function, the cost. The objective of this comparison is to determine if there is a dominance relationship between the metrics in order to eliminate the dominated or the redundant ones.

5 Comparison of the stability metrics

In this section, experiments are conducted in order to compare the stability metrics. To that purpose, the benchmark used is described in subsection 5.1, a dominance function is introduced in subsection 5.2, then experimental results are presented and discussed in subsection 5.3.

All experiments are conducted on a CPU Intel Xeon E5-1620 v3 @3.5Ghz with 64GB RAM with a 600 seconds time limit. The subtour elimination constraints are added dynamically using Gurobi 9.0.0 as a solver with the lazyConstraints parameter. All models are implemented with Java in Eclipse IDE.

Note that only three metrics here are compared, namely, sequence preservation, visit deviation and quantity deviation. As explained in section 4 and shown in Table 6, these metrics are the only ones judged both relevant and compatible with the IRP variant considered in this paper.

5.1 Benchmark

To model the initial problem before unexpected events happen, we use the benchmark proposed by Archetti et al. (2007), which is the most commonly used in the literature. Table 7 describes the characteristics of the instances used. There are two different time horizons ($|\mathcal{H}| = 3$ and 6 periods), and different possible numbers of clients $|\mathcal{V}|$ for each time horizon. The holding cost h is significantly higher in half of the instances. For each combination ($|\mathcal{H}|$, $|\mathcal{V}|$, h), five (#instances) instances exist, on which ten (#events) events are generated for three different sets of scenarios. For example, for $|\mathcal{H}| = 3$ and $|\mathcal{V}| = 5$, there are 5 different instances when the holding cost h is low and 5 others when it is high. For each one of these 10 different instances, 10 events for each scenario are generated, which makes the total number of instances for $|\mathcal{H}| = 3$ and $|\mathcal{V}| = 5$ equal to $10 \times 10 \times 3 = 300$. With all combinations considered, the total number of instances is equal to 3000.

$\lvert\mathcal{H}\rvert$	$\lvert\mathcal{V}\rvert$	h	#instances	Scenarios	#events per scenario
3 6	{5, 10, 20, 30, 40, 50} {5, 10, 20, 30}	{high, low}	5	{1, 2, 3}	10

Table 7: Benchmark for the experimental results

The three different scenarios considered to generate the events are presented in Table 8. The table gives the percentage of clients that change their demand in each scenario; the columns shows the demand variation in each case. For example, in scenario 1, the demand of each client have a probability of 0.1 to become 0, a probability of 0.8 to stay the same and a probability of 0.1 to double. These variations are rather drastic to ensure that significant changes can be observed in the structure of the re-optimised solution. The procedure to generate an event is presented in algorithm 1.

Scenario	Demand				
	$= 0$	$\times 0.5$	$==$	$\times 1.5$	$\times 2$
1	10%	-	80%	-	10%
2	15%	10%	50%	10%	15%
3	30%	-	40%	-	30%

Table 8: Event scenarios

Algorithm 1: Generation of an event

input: Instance w, u index of the scenario
randomly generate a period p where $p \in \mathcal{H}$
for $t \in \{p, p+1, ..., |\mathcal{H}|\}$ **do**
  for $i \in \mathcal{V} \backslash \{0\}$ **do**
    Depending on the scenario u, D_i^t is modified according to Table 8
  end for
end for
return

5.2 A dominance function

Let w be an instance of the problem, $\mathcal{W}$ a set of instances, and $\mathcal{F} = \{f_1, f_2, ..., f_{|\mathcal{F}|}\}$ a set of metrics. An optimal solution of instance w with the metric $f_v \in \mathcal{F}$ as an objective function is denoted by $f_v^*(w)$. We denote by $f_{v'}^*(w|f_v^*(w))$ the optimal solution of instance w with metric $f_{v'}$ as objective function, metric f_v being kept at its optimal value $f_v^*(w)$ through an added constraint.

The dominance of metric f_v over metric $f_{v'}$ for set $\mathcal{W}$ is denoted as $f_v \succ^{\mathcal{W}} f_{v'}$. A metric f_v dominates a metric $f_{v'}$ for set $\mathcal{W}$ if and only if for all instances $w \in \mathcal{W}$ it is possible to find the optimal solution for metric $f_{v'}$ when metric f_v is fixed to its optimal value, i.e. $f_{v'}^*(w) = f_{v'}^*(w|f_v^*(w))$.

Let us now set $|\mathcal{F}| = 4$ and $f_1 = \hat{C}$, $f_2 = SP$, $f_3 = VD$ and $f_4 = QD$ where $\hat{C}$ is the cost, SP the sequence preservation, VD the visit deviation and QD the quantity deviation. In order to compare these metrics, the instances of the benchmark need to be solved to optimality for each metric, and then for each metric fixed to its optimal value through an added constraint, the other metrics are solved to optimality. The procedure for one instance and one event is presented in algorithm 2 and Table 9: it starts by solving the initial problem and obtaining the initial solution $\mathcal{S}$. Afterwards, for each metric $f_v \in \mathcal{F}$ the re-optimisation solution $\mathcal{S}_{f_v}^{\text{new}}$ is obtained from the initial solution S and the event E^w and the optimal value of the metric f_v^* is obtained. For the remaining metrics $f_{v'}$, a new re-optimisation problem is solved, where the value of metric f_v^{new} of index v is fixed to its optimal value f_v^*. The results of this procedure is a matrix $\mathcal{M}$ of dimensions $|\mathcal{F}| \times |\mathcal{F}|$ (Table 9).

As an illustration, in Table 9, cell SP^* contains the optimal solution when the objective is to minimise the sequence preservation, cell $SP^*(VD^*)$ contains the optimal solution when the objective is to minimise the sequence preservation while keeping the visit deviation at its optimal value VD^*, and finally, $VD^*(SP^*)$ contains the optimal solution of the visit deviation when the sequence preservation is kept at its optimal value.

Algorithm 2: Solving procedure for an instance $w \in \mathcal{W}$

input: An instance $w \in \mathcal{W}$, one of its events E^w, an empty matrix $\mathcal{M}$ of $|\mathcal{F}| \times |\mathcal{F}|$ dimension
solve the initial problem to get $\mathcal{S}$ as optimal solution
for $f_v \in |\mathcal{F}|$ **do**
  solve to get $\mathcal{S}^{\text{new}}_{f_v}(E^w, \mathcal{S})$ and set $\mathcal{M}_{f_v,f_v}$ to $f^*_v(w)$
  for $f_{v'} \in |\mathcal{F}|$, $f_{v'} \neq f_v$ **do**
    add the constraint $f^{\text{new}}_v(w) = f^*_v(w)$
    solve to get $\mathcal{S}^{\text{new}}_{f_{v'}}(E^w, \mathcal{S}|f^*_v(w))$ and set $\mathcal{M}_{f_v,f_{v'}}$ to $f^*_{v'}(w)|f^*_v$
  end for
end for
return $\mathcal{M}$

	$\hat{C}$	SP	VD	QD
$\hat{C}$	$\hat{C}^*$	$SP^*(\hat{C}^*)$	$VD^*(\hat{C}^*)$	$QD^*(\hat{C}^*)$
SP	$\hat{C}^*(SP^*)$	SP^*	$VD^*(SP^*)$	$QD^*(SP^*)$
VD	$\hat{C}^*(VD^*)$	$SP^*(VD^*)$	VD^*	$QD^*(VD^*)$
QD	$\hat{C}^*(QD^*)$	$SP^*(QD^*)$	$VD^*(QD^*)$	QD^*

Table 9: Structure of matrix $\mathcal{M}$ for one instance and one of its events

An example for an instance where $|\mathcal{H}| = 6$ and $|\mathcal{V}| = 30$, subjected to an event drawn from the probability distribution of scenario 1, is presented in Table 10. It shows that $VD \succ^w SP$, since $SP^* = SP^*(VD^*) = 0$ and that $SP \succ^w VD$, since $VD^* = VD^*(SP^*) = 3$. Therefore, a dominance relation exists between SP and VD in both ways, thus it is not a strict one. These two metrics can be said to be equivalent, i.e. optimising the sequence preservation first does not keep us from finding the optimal solution for visit deviation and vice-versa. Furthermore, observation of metrics QD and VD shows these metrics to be divergent, i.e. optimising one deteriorates the other.

	$\hat{C}$	SP	VD	QD
$\hat{C}$	8328.03	7	8	1334
SP	8456.58	0	3	468
VD	9016.52	0	3	764
QD	9399.88	0	6	468

Table 10: Results for an instance where $|\mathcal{H}| = 6$ and $|\mathcal{V}| = 30$

5.3 Results and discussion

Comparison of SP, VD and QD. All 3000 instances of set $\mathcal{W}$ are subjected to the procedure described in algorithm 2, solving $|\mathcal{F}| \times |\mathcal{F}| = 16$ MILPs for each instance. The results are presented in Table 11 where each cell represents the ratio of instances where f_v dominates $f_{v'}$.

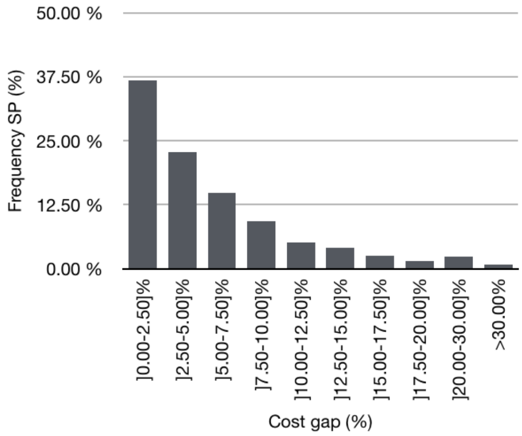
$\succ^{\mathcal{W}}$	$\hat{C}$	SP	VD	QD
$\hat{C}$	-	31.77 %	20.87 %	13.12 %
SP	31.77 %	-	95.33 %	95.19 %
VD	20.87 %	95.33 %	-	89.04 %
QD	13.12 %	95.19 %	89.04 %	-

Table 11: Dominance results

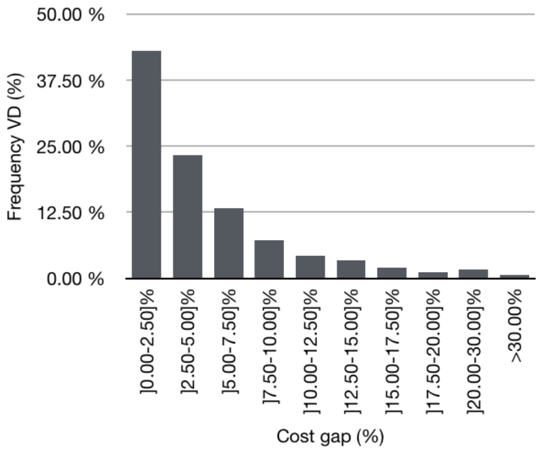
Table 11 shows a high rate of equivalence between the stability metrics, namely, sequence preservation, visit deviation and quantity deviation. Indeed, for sequence preservation, there is an equivalence rate of 95% with both visit deviation and quantity deviation. This means that divergences happen in only 5% of the cases: in these cases, optimising sequence preservation deteriorates visit deviation and/or quantity deviation (and vice-versa). Moreover, although the rate of equivalence between visit deviation and quantity deviation drops to 89% it is still very strong. This amount of equivalence was intuitively expected due to the interrelationship of the mathematical formulations of the stability metrics.

Since keeping a plan stable is supposed to be costly, low equivalence rates were intuitively expected between the cost and the stability metrics. Yet, the results show the cost to be equivalent for, respectively, 31.77%, 20.87% and 13.12% of the cases for sequence preservation, visit deviation and quantity deviation. To deepen this analysis, we thus take a closer look at the evolution of the cost in comparison to the stability metrics when they are divergent. For each instance $w \in \mathcal{W}$, the gap between the optimal cost and the optimal cost when a stability metric is kept at its optimal value $g^{\hat{C}}_{f_v}$ is computed. For example, for the case of cost and sequence preservation: $g^{\hat{C}}_{SP} = \frac{\hat{C}^*(SP^*)-\hat{C}^*}{\hat{C}^*}$. Figure 6 represent the distribution of these gaps for, respectively, sequence preservation, visit deviation and quantity deviation when the cost and these metrics are divergent, i.e. 68.23% of the instances for SP, 79.13% for VD and 86.88% for QD.

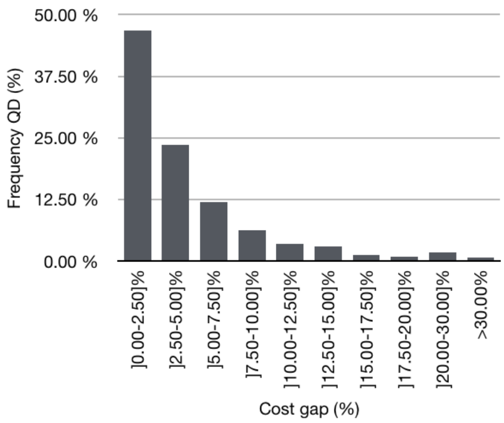
Figure 6a shows the distribution of the gap in the cost regarding SP. We can see that almost 37% of the instances have a gap in the cost inferior to 2.5% and almost 25% when the gap is between 2.5% and 5%. Therefore, for 60% of the instances, the cost is deteriorated only up to



(a) Sequence preservation SP



(b) Visit deviation VD



(c) Quantity deviation QD

Figure 6: A representation of the distribution of the gap in the cost when the cost and the stability metrics SP, VD and QD are divergent

5% when SP is optimal. Furthermore, almost 20% of the instances have a gap within 5% and 10%. On the other hand, less than 1% of the instances have a gap superior to 30%. A similar distribution is observed for both VD and QD for the cost gap interval $]5, +\infty[$. However, for the interval $]0, 10]$ it goes up to 66% for VD and up to 70% for QD. It is worth noting that the choice was made not to include the ratio of the instances where there is an equivalence between the cost and the stability metrics to show that the histograms have the same shape when there is divergence. Table 12 shows the distribution of the gap in cost in relation to the stability metrics for all the instances, including when there is an equivalence, i.e. $g_{f_v}^{\hat{C}} = 0$.

	$g_{f_v}^{\hat{C}}\%$										
	$= 0$	$]0, 2.5]$	$]2.5, 5]$	$]5, 7.5]$	$]7.5, 10]$	$]10, 12.5]$	$]12.5, 15]$	$]15, 17.5]$	$]17.5, 20]$	$]20, 30]$	≥ 30
SP	31.77	25.16	15.60	10.13	6.32	3.49	2.73	1.69	1.04	1.55	0.57
VD	20.87	34.07	18.43	10.53	5.75	3.43	2.75	1.57	0.89	1.29	0.46
QD	13.12	40.59	20.53	10.46	5.51	3.10	2.56	1.14	0.82	1.57	0.64

Table 12: Distribution of the gap in the cost

Two conclusions can be drawn from Table 12. First, the difference in equivalence rate between the stability metrics and the cost is due to very small deviations in the cost. Indeed, the distribution of the gap in the cost for interval $]5, +\infty[$ is the same for all three metrics. The only difference is observed in interval $[0, 5]$ where the difference in equivalence rate between SP and VD is shifted to a gap smaller than 5%. The other conclusion is that the performance of re-optimising with the stability metrics depend on the decision maker. If the decision maker accepts the deterioration of the cost of the solution up to a certain threshold in order to ensure the stability of the solutions, the results show that if the threshold is fixed to 5%, in almost 75% of the instances optimising stability yields an acceptable cost. This rate goes up to almost 90% when the threshold is fixed to 10%.

To ensure that these results can be generalised, at least for set $\mathcal{W}$, a decomposition of the results is performed for the different parameters of the benchmark: number of clients $|\mathcal{V}|$, event scenario (1, 2 or 3) and the size of horizon $|\mathcal{H}|$. For the equivalence rates presented in Table 11, the results stay the same no matter which event scenario is performed. On the other hand, the number of clients and the horizon size have an impact on the equivalence rates between the cost and the stability metrics. For example, when $|\mathcal{V}| = 5$, for $|\mathcal{H}| = 3$ the equivalence rates between cost and SP, VD and QD are, respectively, 52%, 53% and 40% and drop to 31%, 37% and 27% for $|\mathcal{H}| = 6$. When $|\mathcal{V}| = 30$, for $|\mathcal{H}| = 3$, the equivalence rates are 28%, 11% and

4% and drop to 20%, 5% and 2% for $|\mathcal{H}| = 6$. This means that the larger the instances get, in terms of number of clients or size of the time horizon, the smaller the equivalence rate becomes. However, these two parameters do not have any impact on the distribution of the cost, as 90% of the solutions lead to a solution with a gap in the cost inferior to 10%.

On the hardness of the re-optimisation problem. Another observation that is worth mentioning is the difficulty of solving the re-optimisation problems. Because the re-optimisation problems lead to models that are smaller than the original ones, part of the variables being fixed, we intuitively expected them to be easier to solve. However, Table 13 shows that, contrary to the original MILPs, some re-optimisation MILPs are not solved to optimality in the given time limit (600 seconds). For example, when the objective of the re-optimisation is to minimise cost $\hat{C}$, 2884 instances out of 3000 are solved to optimality.

	$\hat{C}$	SP	VD	QD
$\hat{C}$	2884	2841	2841	2832
SP	2830	2941	2941	2940
VD	2945	2945	2945	2945
QD	2945	2945	2945	2945

Table 13: Number of optimal solutions (out of 3000 instances)

We believe this difficulty of solving the re-optimisation problem comes from the demand change rather than from the stability measures added to the problem. To validate this intuition, the instances of the benchmark are solved when the demand is constant (as defined in the benchmark of Archetti et al. (2007)), and then when the demand is modified for the whole horizon $\mathcal{H}$ according to scenario 1 presented in Table 8. The performances for both cases are compared in Table 14 where column $|\mathcal{V}|$ represents the number of clients, $|\mathcal{H}|$ the horizon, h_i indicates whether the holding cost is low (-) or high (+). Columns $\hat{C}_{\text{LR}}$, $\hat{C}_{\text{MILP}}$, g, $cpu(s)$, $\underline{g}$ represent, respectively, the linear relaxation at the root node, the objective value of the MILP, the gap between the optimal solution and the best bound found, the execution time and finally the gap between the optimal solution and the bound at the root node, i.e. $\underline{g} = \frac{\hat{C}_{\text{MILP}} - \hat{C}_{\text{LR}}}{\hat{C}_{\text{LR}}}$, for the instance with constant demand. The next columns, marked with exponent $.^{\text{new}}$, display those same indicators for the instance with modified demands.

Table 14 shows that the model performs better with constant demand as the linear relaxations are tighter. Indeed, the instances of the benchmark provided by Archetti et al. (2007) are generated so that for each $t \in \mathcal{H}$:

$\lvert V\rvert$	$\lvert H\rvert$	h_i	$\hat{C}_{\text{LR}}$	$\hat{C}_{\text{MILP}}$	g	$cpu(s)$	$\underline{g}$	$\hat{C}^{\text{new}}_{\text{LR}}$	$\hat{C}^{\text{new}}_{\text{MILP}}$	g^{new}	$cpu^{\text{new}}(s)$	$\underline{g}^{\text{new}}$
5	3	-	1192.57	1275.86	0.00	0.02	6.98%	1233.02	1262.30	0.00	0.00	**2.37 %**
		+	2117.02	2199.90	0.00	0.02	3.91%	2199.98	2233.86	0.00	0.00	**1.54%**
	6	-	2931.20	3136.90	0.00	0.12	**7.02%**	2984.02	3252.84	0.00	0.09	9.01%
		+	5122.98	5354.20	0.00	0.12	**4.51%**	5102.35	5465.49	0.00	0.08	7.12%
10	3	-	1880.44	1910.93	0.00	0.02	**1.62%**	2017.66	2143.10	0.00	0.01	6.22%
		+	4310.50	4337.97	0.00	0.02	**0.64%**	4647.16	4830.34	0.00	0.01	3.94%
	6	-	4468.22	4612.50	0.00	0.30	**3.23%**	4709.40	5394.20	0.00	0.14	14.54%
		+	8404.77	8601.92	0.00	0.32	**2.35%**	8494.23	8950.21	0.00	0.23	5.37%
20	3	-	2466.49	2665.58	0.00	0.22	**8.07%**	2964.06	3436.15	0.00	0.39	15.93%
		+	7024.92	7225.70	0.00	0.21	**2.86%**	7435.82	7951.09	0.00	0.25	6.93%
	6	-	5944.06	6625.35	0.00	2.06	**11.46%**	6430.98	7740.53	0.00	4.31	20.36%
		+	13802.36	14602.14	0.00	1.93	**5.79%**	14326.88	15792.94	0.00	7.73	10.23%
30	3	-	3145.54	3292.93	0.00	0.43	**4.69%**	3542.42	4301.27	0.00	0.88	21.42%
		+	10756.72	10918.31	0.00	0.66	**1.50%**	11058.58	11804.55	0.00	2.32	6.75%
	6	-	7196.83	7709.87	0.00	9.60	**7.13%**	7544.74	8994.14	0.98	127.61	19.21%
		+	19687.36	20410.65	0.00	8.69	**3.67%**	20132.57	21853.02	0.03	145.77	8.55%
40	3	-	3488.61	3703.82	0.00	0.91	**6.17%**	4066.86	4840.78	0.19	124.36	19.03%
		+	12299.81	12541.06	0.00	1.02	**1.96%**	12650.32	13574.11	0.29	126.75	7.30%
50	3	-	4203.41	4397.15	0.00	2.15	**4.61%**	5097.13	6135.14	3.24	324.43	20.36%
		+	15178.98	15410.82	0.00	3.41	**1.53%**	15694.28	16707.32	1.01	431.92	6.45%

Table 14: Constant demand vs. Scenario 1 demand

1. the demand for a client $i \in \mathcal{V}\backslash\{0\}$ is constant: $\forall t \in \mathcal{H}\backslash\{0\} D_i^t = D_i$

2. the vehicle capacity is three halves of the total demand for a period t: $C = \frac{3}{2} \times \sum\limits_{i \in \mathcal{V}\backslash\{0\}} D_i^t$

3. the quantity produced by the supplier for a period t is equal to the total demand for that period: $R^t = \sum\limits_{i \in \mathcal{V}\backslash\{0\}} D_i^t$

4. the inventory capacity is equal to twice or three times the demand for a period t: $S_i^{\max} = \{2, 3\} \times D_i^t$

5. the initial inventory is equal to the inventory capacity minus the demand for one period: $S_i^0 = S_i^{\max} - D_i$

Assumption 2 combined with assumption 3 makes it possible to replenish all clients at any period of the horizon $\mathcal{H}$ at least to satisfy their demands for the period in question. Assumptions 4 and 5 on the other hand ensure that when $|\mathcal{H}| = 3$ for example, only one delivery for each client can satisfy its demand for the whole horizon. All of the above make Archetti et al. (2007)'s instances easier to solve. However, in the case where the demand is not constant anymore, as it is the case in the re-optimisation problem, these assumptions do not hold anymore. Therefore, we believe that it is necessary to propose a new benchmark of the IRP that is more realistic, the assumptions cited above being rarely met in a real-life instances.

6 Conclusion and perspectives

This article tackles the inventory routing problem under uncertainty by proposing stability metrics to be used when re-optimising the IRP. Re-optimisation is considered as an approach to deal with unexpected events for two reasons. First, the literature lacks "*a posteriori*" methods in the context of IRP. Second, the parameters of the IRP themselves (time dimension, flexibility brought by the inventory) make re-optimisation a relevant approach to investigate. When re-optimising, new objectives appear, besides the cost, to define the performance of the solution. Indeed, it might not be desirable for the new solution to be radically different from the original one – in other words, the stability of the re-optimised solutions compared to the original ones needs to be ensured. To determine how to ensure this stability and due to the lack of agreed-upon definition of the concept, we carried out a literature review of the stability metrics used for components problems of the IRP, respectively, routing and inventory management, and sequencing problems such as scheduling. The stability metrics resulting from this literature review are then adapted to the IRP, their advantages and drawbacks presented, and their mathematical formulations proposed. A framework of comparison validates the mathematical formulations of these metrics as well as investigates their behaviour in relation to each other and to the initial objective function, the cost. We show that the metrics have a tight relationship to each other and are equivalent in most cases. Furthermore, we show that their impact on the cost of the solution seem to be rather small. Finally, by investigating the hardness of solving the re-optimisation models, we show that the benchmark of the literature is easy to solve due to the way it is generated, hence the need of proposing new benchmarks that are compatible with real-life situations.

A future work on this topic is to extend the numerical evaluation to the rest of the proposed stability metrics. As some of these metrics are only compatible with richer variants of the IRP, in addition to comparing a larger panel of stability metrics, it would also help assessing the effectiveness of the three formulated stability metrics in a richer context (multi-depot

(Bertazzi et al., 2019), multi-echelon (Guimarães et al., 2019) or the multi-attribute (Coelho et al., 2020)...) rather than the basic IRP. Other works could focus on investigating the impact of other sources of uncertainty, such as travelling times, on the stable solutions. Furthermore, it would be interesting to confront the re-optimisation approach with stability metrics to real-life instances. This can reinforce the efficiency of the approach if the same results are obtained. Another perspective would be to compare re-optimisation to *"a priori"* methods such as robust and stochastic optimisation. The proposed comparison framework could be used by scholars of other fields to compare different metrics related to various OR problems.

Moreover, we believe that scholars should ponder on the applicability of re-optimisation for other problems, and especially integrated ones, such as the Integrated Process Planning and Scheduling (IPPS), the Location Routing Problem (LRP), the Vehicle Routing Scheduling problem (VRSP). Re-optimisation can be especially relevant for such integrated problems because the existence of different sub-problems brings extra flexibility: the optimisation of one problem can benefit to the other one in the face of disruptions. Future works could therefore focus on adapting the stability metrics presented in this paper, or proposing new relevant ones, for these problems.

References

Aghezzaf, E. H. (2008). Robust distribution planning for supplier-managed inventory agreements when demand rates and travel times are stationary. *Journal of the Operational Research Society*, 59(8):1055–1065.

Agra, A., Christiansen, M., Delgado, A., and Hvattum, L. M. (2015). A maritime inventory routing problem with stochastic sailing and port times. *Computers and Operations Research*, 61:18–30.

Agra, A., Christiansen, M., Hvattum, L. M., and Rodrigues, F. (2018). Robust optimization for a maritime inventory routing problem. *Transportation Science*, 52(3):509–525.

Al-Ameri, T. A., Shah, N., and Papageorgiou, L. G. (2008). Optimization of vendor-managed inventory systems in a rolling horizon framework. *Computers and Industrial Engineering*, 54(4):1019–1047.

Andersson, H., Hoff, A., Christiansen, M., Hasle, G., and Løkketangen, A. (2010). Industrial aspects and literature survey: Combined inventory management and routing. *Computers & Operations Research*, 37(9):1515–1536.

Archetti, C., Bertazzi, L., Laporte, G., and Speranza, M. G. (2007). A branch-and-cut algorithm for a vendor-managed inventory-routing problem. *Transportation Science*, 41(3):382–391.

Archetti, C., Bertazzi, L., and Speranza, M. G. (2003). Reoptimizing the traveling salesman problem. *Networks: An International Journal*, 42(3):154–159.

Bertazzi, L., Bosco, A., and Laganà, D. (2015). Managing stochastic demand in an Inventory Routing Problem with transportation procurement. *Omega (United Kingdom)*, 56:112–121.

Bertazzi, L., Coelho, L. C., De Maio, A., and Laganà, D. (2019). A matheuristic algorithm for the multi-depot inventory routing problem. *Transportation Research Part E: Logistics and Transportation Review*, 122:524–544.

Bertazzi, L. and Speranza, M. G. (2013). Inventory routing problems with multiple customers. *EURO Journal on Transportation and Logistics*, 2(3):255–275.

Bertsimas, D. J., Gupta, S., and Tay, J. (2016). Scalable Robust and Adaptive Inventory Routing. *Optimization Online*, pages 1–29.

Chen, Y. M. and Lin, C. T. (2009). A coordinated approach to hedge the risks in stochastic inventory-routing problem. *Computers and Industrial Engineering*, 56(3):1095–1112.

Cho, J., Lim, G. J., Kim, S. J., and Biobaku, T. (2018). Liquefied natural gas inventory routing problem under uncertain weather conditions. *International Journal of Production Economics*, 204.

Chow, J. Y. (2016). Dynamic UAV-based traffic monitoring under uncertainty as a stochastic arc-inventory routing policy. *International Journal of Transportation Science and Technology*, 5(3):167–185.

Christiansen, M. and Nygreen, B. (2005). Robust inventory ship routing by column generation. In *Column Generation*. Springer-Verlag.

Coelho, L. C., Cordeau, J.-F., and Laporte, G. (2014). Thirty years of inventory routing. *Transportation Science*, 48(1):1–19.

Coelho, L. C., De Maio, A., and Laganà, D. (2020). A variable mip neighborhood descent for the multi-attribute inventory routing problem. *Transportation Research Part E: Logistics and Transportation Review*, 144:102137.

Coelho, L. C. and Laporte, G. (2014). Improved solutions for inventory-routing problems through valid inequalities and input ordering. *International Journal of Production Economics*, 155:391–397.

Cowling, P. and Johansson, M. (2002). Using real time information for effective dynamic scheduling. *European Journal of Operational Research*, 139(2):230–244.

Curry, J. and Peters, B. (2005). Rescheduling parallel machines with stepwise increasing tardiness and machine assignment stability objectives. *International Journal of Production Research*, 43(15):3231–3246.

Desaulniers, G., Rakke, J. G., and Coelho, L. C. (2016). A branch-price-and-cut algorithm for the inventory-routing problem. *Transportation Science*, 50(3).

Dettenbach, A. M. C. and Ubber, S. (2015). Managing disruptions in last mile distribution. In *48th Hawaii International Conference on System Sciences (HICSS)*, pages 1078–1087.

Diz, G. S. d. S., Hamacher, S., and Oliveira, F. (2019). A robust optimization model for the maritime inventory routing problem. *Flexible Services and Manufacturing Journal*, 31(3):675–701.

Dong, Y., Maravelias, C. T., and Jerome, N. F. (2018). Reoptimization framework and policy analysis for maritime inventory routing under uncertainty. *Optimization and Engineering*, 19(4):937–976.

Ercan, S. and Cinar, D. (2017). A multiproduct multi-vehicle inventory routing problem with uncertainty. *Springer Optimization and Its Applications*, 129:181–198.

Fardi, K., Jafarzadeh Ghoushchi, S., and Hafezalkotob, A. (2019). An extended robust approach for a cooperative inventory routing problem. *Expert Systems with Applications*, 116:310–327.

Ghasemi, E. and Bashiri, M. (2018). A selective covering-inventory- routing problem to the location of bloodmobile to supply stochastic demand of blood. *International Journal of Industrial Engineering & Production Research*, 29(2):147–158.

Gruler, A., Panadero, J., de Armas, J., Moreno Pérez, J. A., and Juan, A. A. (2018). Combining variable neighborhood search with simulation for the inventory routing problem with stochastic demands and stock-outs. *Computers and Industrial Engineering*, 123(June):278–288.

Guimarães, T. A., Coelho, L. C., Schenekemberg, C. M., and Scarpin, C. T. (2019). The two-echelon multi-depot inventory-routing problem. *Computers & Operations Research*, 101:220–233.

Harris, F. W. (1919). How many parts to make at once. *Factory, The Magazine of Management*, 10(2):135–136.

Herrera, C., Belmokhtar-Berraf, S., Thomas, A., and Parada, V. (2016). A reactive decision-making approach to reduce instability in a master production schedule. *International Journal of Production Research*, 54(8):2394–2404.

Herroelen, W. and Leus, R. (2005). Project scheduling under uncertainty: Survey and research potentials. *European Journal of Operational Research*, 165(2):289–306.

Hu, H., Li, J., and Li, X. (2018). A credibilistic goal programming model for inventory routing problem with hazardous materials. *Soft Computing*, 22(17):5803–5816.

Huang, S. H. and Lin, P. C. (2010). A modified ant colony optimization algorithm for multi-item inventory routing problems with demand uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 46(5):598–611.

Hvattum, L. M. and Løkketangen, A. (2009). Using scenario trees and progressive hedging for stochastic inventory routing problems. *Journal of Heuristics*, 15(6):527–557.

Inderfurth, K. (1994). Nervousness in inventory control : analytical results. *Operations-Research-Spektrum*, 16(2):113–123.

Jafarian, A., Asgari, N., Sina, S., and Fatemi-sadr, E. (2019). The inventory-routing problem subject to vehicle failure. *Transportation Research Part E*, 126:254–294.

Juan, A. A., Grasman, S. E., Caceres-Cruz, J., and Bektaş, T. (2014). A simheuristic algorithm for the Single-Period Stochastic Inventory-Routing Problem with stock-outs. *Simulation Modelling Practice and Theory*, 46:40–52.

Kadipasaoglu, S. N. and Sridharan, S. V. (1997). Measurement of instability in multi-level MRP systems. *International Journal of Production Research*, 35(3):713–737.

Karoonsoontawong, A. and Unnikrishnan, A. (2013). Inventory routing problem with route duration limits and stochastic inventory capacity constraints : Tabu search heuristics. *Transportation Research Record: Journal of the Transportation Research Board*, 2378(1):43–53.

Kazemi, S. M., Rabbani, M., Tavakkoli-Moghaddam, R., and Shahreza, F. A. (2017). Blood inventory-routing problem under uncertainty. *Journal of Intelligent and Fuzzy Systems*, 32(1):467–481.

Kimms, A. (1998). Stability measures for rolling schedules with applications to capacity expansion planning, master production scheduling, and lot sizing. *Omega*, 26(3):355–366.

Kleywegt, A. J., Nori, V. S., and Savelsbergh, M. W. P. (2004). Dynamic programming approximations for a stochastic inventory routing problem. *Transportation Science*, 38(1):42–70.

Lefever, W. (2018a). An effective sample average approximation algorithm for the Inventory-Routing Problem with Transshipment under stochastic demand. In *ILS 2018 - Information Systems, Logistics and Supply Chain, Proceedings*, pages 115 – 124.

Lefever, W. (2018b). *Stochastic and robust optimization algorithms for the inventory-routing problem*. PhD thesis, Universiteit Gent.

Lefever, W., Touzout, F. A., Hadj-Hamou, K., and Aghezzaf, E. H. (2019). Benders' decomposition for robust travel time-constrained inventory routing problem. *International Journal of Production Research*, 0(0):1–25.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Cybernetics and Control Theory*, 10(8):845–848.

Li, M., Wang, Z., and Chan, F. T. S. (2016a). A robust inventory routing policy under inventory inaccuracy and replenishment lead-time. *Transportation Research Part E: Logistics and Transportation Review*, 91:290–305.

Li, M., Wang, Z., and Chan, F. T. S. (2016b). An inventory routing policy under replenishment lead time. *IEEE Transactions on Systems, Man and Cybernetics*, 47(12):3150–3164.

Malekly, H. (2015). The inventory pollution-routing problem under uncertainty. In *Green Logistics and Transportation*, pages 83–117. Springer, Cham.

Micheli, G. J. and Mantella, F. (2018). Modelling an environmentally-extended inventory routing problem with demand uncertainty and a heterogeneous fleet under carbon control policies. *International Journal of Production Economics*, 204(November 2017):316–327.

Mirzapour Al-e-hashem, S. M., Rekik, Y., and Mohammadi Hoseinhajlou, E. (2019). A hybrid L-shaped method to solve a bi-objective stochastic transshipment-enabled inventory routing problem. *International Journal of Production Economics*, 209(July 2018):381–398.

Narayanan, A. and Robinson, P. (2010). Evaluation of joint replenishment lot-sizing procedures in rolling horizon planning systems. *International Journal of Production Economics*, 127(1):85–94.

Niakan, F. and Rahimi, M. (2015). A multi-objective healthcare inventory routing problem; a fuzzy possibilistic approach. *Transportation Research Part E: Logistics and Transportation Review*, 80:74–94.

Nikzad, E., Bashiri, M., and Oliveira, F. (2019). Two-stage stochastic programming approach for the medical drug inventory routing problem under uncertainty. *Computers and Industrial Engineering*, 128(July 2017):358–370.

Nolz, P., Absi, N., and Feillet, D. (2014). A stochastic inventory routing problem for infectious medical waste collection. *Networks*, 63(1):82–95.

Nori, V. S. and Savelsbergh, M. W. P. (2002). The stochastic inventory routing problem with direct deliveries. *Transportation Science*, 36(1):94–118.

Pfeiffer, A., Kádár, B., and Monostori, L. (2007). Stability-oriented evaluation of rescheduling strategies, by using simulation. *Computers in Industry*, 58(7):630–643.

Rahbari, A., Nasiri, M. M., and Ghorbani, E. (2017). An inventory routing problem for perishable products with demand and transportation time uncertainty: A robust optimization approach. In *The 10th International Conference of Iranian Operations Research Society*.

Rahim, M. K. I. A., Zhong, Y., Aghezzaf, E. H., and Aouam, T. (2014). Modelling and solving the multiperiod inventory-routing problem with stochastic stationary demand rates. *International Journal of Production Research*, 52(14):4351–4363.

Rahimi, M., Baboli, A., and Rekik, Y. (2017). Multi-objective inventory routing problem: A stochastic model to consider profit, service level and green criteria. *Transportation Research Part E: Logistics and Transportation Review*, 101:59–83.

Rakke, J. G., Stålhane, M., Moe, C. R., Christiansen, M., Andersson, H., Fagerholt, K., and Norstad, I. (2011). A rolling horizon heuristic for creating a liquefied natural gas annual delivery program. *Transportation Research Part C: Emerging Technologies*, 19(5):896–911.

Rangsaritratsamee, R., Ferrell, W. G., and Kurz, M. B. (2004). Dynamic rescheduling that simultaneously considers efficiency and stability. *Computers and Industrial Engineering*, 46(1):1–15.

Rodrigues, F., Agra, A., Christiansen, M., Hvattum, L. M., and Requejo, C. (2019). Comparing techniques for modelling uncertainty in a maritime inventory routing problem. *European Journal of Operational Research*, 277(3):831–845.

Sahin, F., Robinson, E. P., and Gao, L. L. (2008). Master production scheduling policy and rolling schedules in a two-stage make-to-order supply chain. *International Journal of Production Economics*, 115(2):528–541.

Salavati-Khoshghalb, M., Gendreau, M., Jabali, O., and Rei, W. (2019). A hybrid recourse policy for the vehicle routing problem with stochastic demands. *EURO Journal on Transportation and Logistics*, 8(3):269–298.

Schönberger, J. and Kopfer, H. (2008). Schedule nervousness reduction in transport re-planning. *Communications of SIWN*, 3:7–14.

Solyalı, O., Cordeau, J.-F., and Laporte, G. (2012). Robust inventory routing under demand uncertainty. *Transportation Science*, 46(3):327–340.

Sörensen, K. (2006). Route stability in vehicle routing decisions: a bi-objective approach using metaheuristics. *Central European Journal of Operations Research*, 14(2):193–207.

Soysal, M. (2016). Closed-loop Inventory Routing Problem for returnable transport items. *Transportation Research Part D: Transport and Environment*, 48(December 2015):31–45.

Soysal, M., Bloemhof-Ruwaard, J. M., Haijema, R., and Van Der Vorst, J. G. (2015). Modeling an inventory routing problem for perishable products with environmental considerations and demand uncertainty. *International Journal of Production Economics*, 164:118–133.

Soysal, M., Bloemhof-Ruwaard, J. M., Haijema, R., and van der Vorst, J. G. (2018). Modeling a green inventory routing problem for perishable products with horizontal collaboration. *Computers and Operations Research*, 89:168–182.

Sridharan, S. V., Berry, W. L., and Udayabhanu, V. (1988). Measuring Master Production Schedule Stability Under Rolling Planning Horizons. *Decision Sciences*, 19(1):147–166.

Trudeau, P. and Dror, M. (1992). Stochastic inventory routing: Route design with stockouts and route failures. *Transportation Science*, 26(3):171–184.

Wang, X., Ruan, J., Shang, H., and Ma, C. (2011). *A combinational disruption recovery model for vehicle routing problem with time windows.* Springer, Berlin, Heidelberg.

Wu, S. D., Storer, R. H., and Chang, P. C. (1993). One-machine rescheduling heuristics with efficiency and stability as criteria. *Computers and Operations Research*, 20(1):1–14.

Yang, Z., Emmerich, M., Bäck, T., and Kok, J. (2016). Multi-objective inventory routing with uncertain demand using population-based metaheuristics. *Integrated Computer-Aided Engineering*, 23(3):205–220.

Yu, Y., Chu, C., Chen, H., and Chu, F. (2012). Large scale stochastic inventory routing problems with split delivery and service level constraints. *Annals of Operations Research*, 197(1):135–158.

Zhang, C., Nemhauser, G., Sokol, J., and Papageorgiou, M.-s. C. D. (2015). Robust inventory routing with flexible time window allocation. Technical report, Georgia Institute of Technology, Atlanta.

Zhang, X. and Tang, L. (2007). Disruption management for the vehicle routing problem. In *International Conference on Intelligent Computing*, pages 225–234.