WORKING PAPER SERIES

Adaptive stochastic lookahead policies for dynamic multi-period purchasing and inventory routing

Daniel Cuellar-Usaquén/Marlin W. Ulmer/
Camilo Gomez/David Álvarez-Martínez

Working Paper No. 04/2023

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

FACULTY OF ECONOMICS
AND MANAGEMENT

Impressum (§ 5 TMG)

Herausgeber:
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Der Dekan

Verantwortlich für diese Ausgabe:
Daniel Cuellar-Usaquén, Marlin W. Ulmer
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Postfach 4120
39016 Magdeburg
Germany

http://www.fww.ovgu.de/femm

Bezug über den Herausgeber
ISSN 1615-4274

Adaptive stochastic lookahead policies for dynamic multi-period purchasing and inventory routing

Daniel Cuellar-Usaquén
Universidad de Los Andes, dh.cuellar@uniandes.edu.co,

Marlin W. Ulmer
Otto-von-Guericke Universität Magdeburg, marlin.ulmer@ovgu.de,

Camilo Gomez, David Álvarez-Martínez
Universidad de Los Andes, gomez.ch@uniandes.edu.co, d.alvarezm@uniandes.edu.co,

We present a problem motivated by discussions with Colombian e-commerce platforms for agri-food products. In regular time intervals (periods), the platforms collect groceries from local farmers and stores them at a warehouse to distribute them to local customers. The supply quantities and prices per farmer and the cumulated customer demand can change from period to period. Thus, there is value in purchasing more than needed in one period to exploit cheap prices and consolidation opportunities, to hedge against future uncertainty, and to save routing cost in future periods. A careful balance between too much and not enough inventory needs to be found, especially, since inventory perishes over time. The resulting optimization problem is a stochastic dynamic multi-period routing problem with inventory and purchasing decisions. The decision space of the problem is vast as it combines purchasing, inventory, and routing decisions. Further, the value of a decisions is unknown since it depends on future developments and decisions. We propose solving the problem with a stochastic lookahead method. In every state, the method samples a set of future realizations and solves the resulting two-stage stochastic program. To cope with the complex decision space in first and second stage, we propose a "soft" decomposition where the inventory and purchasing decision are fully considered, but the routing decisions are simplified and their cost is approximated via a cost function approximation. As the routing cost also depends on future decisions, the approximated cost are learned iteratively via repeated simulation and adaption of the lookahead. We show that our method outperforms a large number of benchmark policies for a variety of instances. We further analyze the functionality of our method and investigate variation in the problem dimensions in a comprehensive analysis.

Key words: Agri-Food Supply Chains, Dynamic Multi-period Vehicle Routing, Stochastic Dynamic Decision Making, Approximate Dynamic Programming, Two-Stage Stochastic Programming, Cost Function Approximation

History:

1. Introduction

In recent years, there has been a growing demand for fresh, locally sourced, and high-quality food products worldwide (Fukase and Martin 2020), as consumers are becoming more conscious of the origin and quality of their food. In addition, locally sourced food products require less transportation, refrigeration, and packing, which can significantly reduce the food system's carbon footprint. This trend has created new business models in which local suppliers are connected to the local consumer through e-commerce platforms. These platforms allow small producers to sell their products and reach new markets, enabling easy access to fresh and locally sourced products (Gu et al. 2022).

In countries worldwide, from Denmark to Colombia, initiatives have been developed to distribute local agricultural products through e-commerce platforms (Halkier and James 2022). These initiatives have created new supply chain structures composed of multiple small producers that require coordination of the replenishment process with a network of participants rather than relying on a single supplier. This process involves purchasing products, collecting them from suppliers, and managing stocks in distribution centers (*first mile*, which is the focus of this work), followed by distribution to end customers (*last mile*, which is out of the scope of this work). These structures with high horizontal integration reduce the number of intermediaries, increase the participation of small producers, and improve the efficiency and transparency of the supply chain, in addition to reducing the gap between small and large producers in terms of competitiveness (Prajapati et al. 2022).

These new structures, despite being advantageous, pose new challenges for the supply chain management. Local customers have high expectations, whereas smaller, regionally dispersed suppliers can make ensuring consistent supply and quality challenging. In addition, the volatility of the demand, supply, and pricing creates additional complexities in the management of the agri-food supply chains. To address these challenges, it is necessary to develop methodologies that account for uncertainty and manage the complexity of joint decisions in replenishment, routing, and inventory operations in a dynamic environment (Majluf-Manzur et al. 2021). Failing to consider these factors can result in poor planning and inefficiencies.

In this paper, we consider a problem based on discussions with online agri-food platforms in Colombia; over a time horizon, a company aims at satisfying periodically occurring, uncertain customer demand for different products with fixed sale revenue. We focus on the first mile, the collection of products from a known set of suppliers to the warehouse per period assuming aggregated customer demand directly at the warehouse. Customer demand, supply volumes, and purchase prices of the products for each supplier over the periods are uncertain. This information is revealed at the beginning of each period. In every period, the company purchases and collects products

from the suppliers and store the collected products at a warehouse. For collection, the company has access to a fleet of homogeneous vehicles. Collected products can either be used to satisfy current demand or stored to cover potential demand in later periods. However, stored quantities decline over time due to perishability. The company's goal is to maximize the expected overall profit (i.e., sales revenue minus purchasing and routing cost) over the time horizon.

This problem is complex because it involves interconnected decisions of inventory, purchasing, and routing. Approaching these decisions individually can lead to cost overruns. For example, purchasing products from the cheapest suppliers may lead to expensive routing, whereas routing over nearby suppliers may lead to high purchasing cost. Additionally, the value of a decision in one period depends on stochastic demand, supply, and purchase prices, as well as decisions in future periods. Building an inventory stock may reduce expected future purchasing and routing cost, but buying too much may lead to additional cost due to perished goods. Therefore, it is necessary to strike a careful balance between purchasing, routing, and inventory decisions every period, saving costs in the current period and staying flexible for future periods.

We propose a stochastic lookahead method that takes all aspects into consideration when making decisions. The method determines purchasing and inventory decisions in each period based on future demand, supply, and price scenarios. To reduce complexity in exploring the decision space, we approximate individual routing cost for each visited supplier rather than integrating explicit routing decisions. This approximation considers the direct shipping cost and the consolidation potential of each supplier. The approximation is adaptively learned by iterative simulations of the stochastic lookahead, ultimately resulting in our **STAR**-policy (**ST**ochastic lookahead with **A**daptive **R**outing approximation).

We analyze the performance and functionality of our policy in a comprehensive computational study, comparing it to a set of benchmark policies tailored for our case study, as well as to policies from the literature. Our policy outperforms all benchmarks for a large set of instances. Our adaptively learned, approximated routing cost leads to an approximation error of only 5% and is essential for the success of our **STAR**-policy. Our computational study reveals the following managerial insights:

- Our **STAR**-policy suggests building an inventory initially and then using this inventory to flexibly decide about the products to purchase based on the realized volumes and prices.
- Compared to a **MYOPIC**-policy that does not build up inventory, the purchasing cost do not change significantly, however, the routing cost decreases.
- It is not necessarily a problem if supply volumes and purchase prices are volatile. In contrast, this volatility provides opportunities for buying larger quantities at cheap prices.

- While geographical adjacency to other suppliers is a decent indicator for a supplier's consolidation potential, other factors such as supply volume or prices play an important role, too.
- The cost of the fleet and the profit margin are the main drivers of the business model's success in our experiments.
- The perishability of products is an important factor for the platform's revenue. Significant improvements can be seen, even if only a subset of products last longer.

Our contributions are as follows. We present a comprehensive dynamic decision process model for a new practical problem based on our collaboration with e-commerce platforms in the agri-food supply chain in Colombia. We develop an anticipatory method that solves the problem and obtains significant improvements compared to a large set of benchmark policies. We analyze the components of the problem in a comprehensive sensitivity analysis and derive a set of valuable managerial insights. Methodologically, we propose embedding a cost function approximation (CFA) in a stochastic lookahead method and adaptively tuning the CFA parameters. This methodology allows us to find the purchasing, routing, and inventory management decisions and reduce complexity by having supplier-dependent estimates. The general concept might prove valuable for dynamic decision problems with complex routing decisions; an increasingly important problem field especially in transportation (Hildebrandt, Thomas, and Ulmer 2023).

The remainder of this paper is structured as follows. We present the related literature in Section 2. The problem is defined in Section 3. In Section 4, we present our approach. The design of experiments is described in Section 5. Computational studies and the detailed analysis of the results are presented in Section 6. The paper concludes with Section 7. We further present a comprehensive Appendix with more details on literature, methodology, and results.

2. Related work

The work on the individual components of our problem is vast. In this section, we focus on the most relevant literature. For a detailed overview on other related work with respect to inventory, purchasing, or multi-period routing, we refer to Appendix A.1.

To the best of our knowledge, the presented problem has not been studied in the literature as it combines dynamic multi-period inventory management of multiple, perishable products with procurement routing under uncertain demand, purchase price, and supply. There is work on the deterministic variant of the problem by Çabuk and Erol (2019). They propose a Mixed Integer Nonlinear model (MINLP) to solve a multi-period problem that integrates purchasing, routing and inventory decisions and considers price discounting as a function of quantity purchased. The authors perform a scenario analysis to evaluate the model results under varying conditions for a small instance size. Using an adaption of their approach operating on expected values (**EV**) as a

benchmark, we show that the explicit consideration of uncertainties is key for successful decision making.

In Keskin et al. (2023), over multiple periods, waste is collected from a set of facilities. To save cost, the provider can call "convenient" facilities and suggest preemptively picking up waste to avoid visits in later periods. The authors propose rules to identify promising facilities to call based on the expected waste volumes and on how they could be integrated in the routes. The problem shares the challenge of multi-period routing under uncertainty. Further, decisions are required that extend routing and inventory collection in one period to avoid cost in the next. Our proposed methodology explicitly considers future uncertainties and routing. To analyze the value of our method, we also implement a rule-based approach (policy function approximation (**PFA**), compare Powell 2021). Related to Keskin et al. (2023), this approach builds up inventory by adding supply and suppliers to visit in a period based on prices and the routing cost. We show that while such a method performs comparably well, there is significant value in explicitly incorporating future developments.

Other related work is presented by Brinkmann, Ulmer, and Mattfeld (2019, 2020), both problem- and method-wise. In the papers, the authors dynamically transport bikes to ensure sufficient inventories at bike-sharing stations. The authors propose a lookahead approach for evaluating inventory decisions. To this end, future demand is sampled and the inventory that minimizes failed demand is selected. The length of the lookahead horizon is time-dependent and learned via value function approximation. Future routing decisions are not considered in the lookahead. Our work shows similarities that we also use samples of the future to determine inventory decisions. However, we explicitly incorporate the routing and its cost in our lookahead model. Our preliminary tests showed that ignoring future routing in the lookahead leads to very poor results.

The idea of integrating approximated cost in a lookahead is also proposed by Ulmer et al. (2019) and Liu and Luo (2023). In Ulmer et al. (2019), a decomposition of the decision space is performed to allow enumeration of potential decisions. Each decision is then evaluated based on a set of scenarios where within the scenarios, decisions are made by a pre-trained policy. Our method is different as it allows for an integrated optimization considering all potential decisions with respect to all scenarios and as the training is done in an integrated manner by using the policy. In Liu and Luo (2023), for a dynamic multi-period dispatching problem, solving a stochastic program to search the decision space of the current stage is proposed. In later stages of the program, costs are approximated by using a myopic strategy. Thus, there is no iterative learning involved. In our computational study, we show that this iterative learning adds significant improvements compared to the stochastic lookahead with approximation of a myopic policy (**ST-MY**). Baty et al. (2023) address a similar problem. They do not solve a stochastic program, but approximate future cost

based on perfect information solutions. The cost is then used in a metaheuristic to search the decision space. Thus, there is no adaptive learning based on the realized cost. Our problem is too complex to determine such ex-post solutions and test the idea proposed in Baty et al. (2023) as a benchmark. However, the insights of Heinold, Meisel, and Ulmer (2023) suggest that approximation via perfect information alone often leads to too optimistic evaluations, especially if the actual cost realizations are not considered. Finally, as in our work, Haferkamp, Ulmer, and Ehmke (2023) suggest adaptively learning of cost. These costs are not used in optimization, but in a PFA for heatmap design guiding crowdsourced drivers to lucrative spots in the city.

In summary, to the best of our knowledge, we are the first to address the proposed problem and the first suggesting a dynamic policy that embeds an adaptively learned CFA in a two-stage stochastic program.

3. Problem definition

In this section, we present the problem statement. We first describe the problem. Then, we model the problem as a sequential decision process and provide an illustrative example.

3.1. Problem description

We consider the problem of an e-commerce platform purchasing and collecting agricultural products from a set of regional farmers to satisfy customer demand over a time horizon.

We assume a harvesting season with a limited number of periods (e.g., days). At the beginning of each period, customer demand quantities of the products reveal. This demand needs to be satisfied at a warehouse at the end of the period. The selling price the customers pay for a product is known and constant over the periods. The products are offered by regional suppliers. The suppliers are distributed around the warehouse. Each supplier offers a subset of products. The quantities and purchase price of the products per supplier vary from period to period and become known in the beginning of each period. For the (very unlikely) case that the realized demand is higher than all the available supply, there is also a "backorder" supplier (e.g., a wholesaler) located directly at the warehouse, offering unlimited quantities of all products at fixed high prices.

To collect products from the suppliers, the provider can hire vehicles. The vehicles have a maximum capacity and maximum working duration per period. The cost of the vehicles depends on their working time. The vehicles start and end their tours at the warehouse. We assume that split collections are prohibited. Thus, each supplier is visited by at most one vehicle. There is a service time to load the products on the vehicles. Vehicles may not only collect products demanded in the current period, but can also collect more. These additional product quantities are stored at the warehouse with unlimited capacity. A known percentage of stored inventory perishes between the periods. Additional inventory holding cost are not considered.

Every period, the e-commerce platform planner decides the product quantities to buy from each supplier (including the backorder supplier) and how to create routes for collecting them. Buying more than needed to satisfy the demand of the period is possible. A decision is feasible if the period's customer demand can be satisfied and the collection routes do not violate capacity or time constraints. The reward of a decision is the difference between the revenue of selling the products at the warehouse and the cost of purchasing and routing. The objective of the provider is to maximize the expected reward over all periods.

3.2. Sequential decision process

The problem at hand is a stochastic and dynamic decision problem. It is stochastic because the products' demand, purchase prices, and supply are only known at the beginning of each period. It is dynamic because a sequence of decisions must be made, one decision per period. In addition, current decisions change inventory volumes for future periods, thus influencing future decisions.

A stochastic dynamic decision problem can be modeled as a sequential decision process (Powell 2021), modeling the problem as a sequence of states. In each state, a decision is made and a reward is observed. Then, stochastic information is revealed and a transition leads to the next state. In the following, we define the states, decisions, reward, stochastic information, and transition function of our problem. First, we introduce the global notation.

Global Notation. The periods are denoted as $t \in T$ with $T = \{1, 2, \ldots, |T|\}$. We assume a set of products $k \in K$ and a set of suppliers $m \in M$ (including the backorder supplier, $|M|$).

We define the collection network as a complete, undirected graph $G = (V, E)$. Let $V := M \cup \{0\}$ be the set of vertices, where 0 represents the warehouse. Let E be the set of arcs where $E = \{(i, j) : i, j \in V\}$. Each arc $(i, j) \in E$ is associated with a non-negative travel time τ_{ij}. The travel time to and from the backorder supplier is 0, $(\tau_{0|M|}, \tau_{|M|0} = 0)$. There is a service time at each supplier to load the products on the vehicles. We include them in the travel times τ_{ij}, leading to asymmetric travel times from/to the warehouse. For every time unit traveled there is a cost of c. Vehicles have a maximum load capacity Q. The maximum working time per vehicle and day is denoted by $l^{\max}$.

Each supplier $m \in M$ provides a subset of the products $K_m \subseteq K$ $(K_{|M|} = K)$. Moreover, each product k is provided by a subset of suppliers M_k. Product $k \in K$ has a unit revenue r_k and a perishability percentage of $\phi_k \in [0, 1]$.

State. A decision is made every period. The state comprises all information available to make a decision. We denote the state in period $t \in T$ as S_t. For our problem, the state S_t consists of four components, two related to the warehouse and two related to the suppliers:

- the current inventory level of product k in the warehouse at period t; denoted by $\hat{I}_{kt}$,
- the demand for product k at period t, denoted as d_{kt},

- the purchase price at supplier m for product k at period t, denoted by p_{mkt},
- and the available quantity of product k at supplier m at period t, denoted by q_{mkt} ($q_{|M|kt} = \infty$).

State S_t can be summarized as $S_t = (\hat{I}_t, d_t, p_t, q_t)$ with $\hat{I}_t$ and d_t being $|K|$-dimensional vectors and p_t and q_t being $|M| \times |K|$-matrices. In the initial state S_1, the inventory at the warehouse is empty, $\hat{I}_1 = 0$.

Decision. We denote a decision at period $t \in T$ as a_t. A decision $a_t = (z_t, I_t, e_t, x_t, f_t)$ has five components that reflect purchasing and routing parts. The purchasing part is modeled via z_t and I_t. It determines the quantity of products to buy from each supplier, represented by matrix $z_t = (z_{mkt})_{m \in M, k \in K}$. The purchasing decision induces the inventory level at the end of the period as the difference between the demand and the sum of the initial inventory and the quantities purchased. The inventory at the end of the period of product k is represented by variable I_{kt} (compare Equation (1)).

The second part of the decision is the definition of collection routes, modeled via e_t, x_t, and f_t. We assume a sufficiently large available set of vehicles F (e.g., $|F| = |M|$). The routing is then modeled as follows. First, variable e_{mt} takes the value of one if supplier $m \in M$ is visited at period t, 0 otherwise. Second, variable f_{0t} is the number of vehicles dispatched from the warehouse at period t. Variable $x_{ijt} \in \{0, 1\}$ is the routing variable, indicating whether the arc from supplier i to supplier j is activated at period t. In the following, we summarize the decision space using a mixed-integer formulation.

A decision $a_t = (z_t, I_t, e_t, x_t, f_t)$ at period $t \in T$ is feasible, if the following constraints hold:

$$I_{kt} = \hat{I}_{kt} + \sum_{m \in M_k} z_{mkt} - d_{kt}, \qquad \forall k \in K \tag{1}$$

$$z_{mkt} \leq q_{mkt} e_{mt}, \qquad \forall k \in K, \forall m \in M_k \tag{2}$$

$$\sum_{i \in V} x_{imt} = e_{mt}, \qquad \forall m \in M \tag{3}$$

$$\sum_{k \in K} z_{mkt} \leq Q e_{mt}, \qquad \forall m \in M \tag{4}$$

$$\sum_{(j,j') \in \delta(\{m\})} x_{jj't} = 2 e_{mt}, \qquad \forall m \in M \tag{5}$$

$$\sum_{(j,j') \in \delta(\{0\})} x_{jj't} = 2 f_{0t}, \tag{6}$$

$$Q \sum_{(i,j) \in E(M')} x_{ijt} \leq \sum_{m \in M'} \left(Q e_{it} - \sum_{k \in K_m} z_{mkt} \right), \forall M' \subseteq M, |M'| \geq 2 \tag{7}$$

$$u_{it} - u_{jt} + l^{max} x_{ijt} \leq l^{max} - \tau_{ij}, \qquad \forall i, j \in M | i \neq j \tag{8}$$

$$f_{0t} \in \mathbb{Z}, \tag{9}$$

$$e_{mt} \in \{0,1\}, \qquad \forall m \in M \tag{10}$$

$$z_{mkt} \geq 0, \qquad \forall k \in K, \forall m \in M_k \tag{11}$$

$$z_{mkt} \leq q_{mkt}, \qquad \forall k \in K, \forall m \in M_k \tag{12}$$

$$I_{kt}, \geq 0, \qquad \forall k \in K \tag{13}$$

$$x_{ijt} \in \{0,1\}, \qquad \forall (i,j) \in E, \tag{14}$$

$$u_{it} \leq l^{max}, \qquad \forall i \in M \tag{15}$$

$$u_{it} \geq \min_{j} \tau_{ij}, \qquad \forall i \in M \tag{16}$$

Equation (1) accounts for the inventory at the end of the period and guarantees the satisfaction of the demand. Equation (2) ensures that the purchase should not exceed the available capacity of the suppliers and that a vehicle can only collect quantities if it visits the supplier. Equation (3) and Equation (4) are non-split visit constraints that ensure that each supplier is visited by at most one vehicle and that no more than the vehicle's capacity is purchased. Equation (5) and Equation (6) are the degree constraints that maintain the flow in and out to the nodes. Equation (7) and Equation (8) are the sub-tour elimination constraints based on vehicle capacity and maximum travel time (Iori, Salazar-González, and Vigo (2007), Toth and Vigo (2002), Miller, Tucker, and Zemlin (1960)). Finally, Equations (9)-(16) define the domain of the variables.

The reward of decision a_t in state S_t is the sum of revenue minus the cost for purchasing and routing:

$$R(S_t, a_t) = \sum_{k \in K} \left(r_k d_{kt} - \sum_{m \in M_k} p_{mkt} z_{mkt} \right) - c \sum_{(i,j) \in E} \tau_{ij} x_{ijt} \tag{17}$$

The combination of state S_t and decision a_t leads to a post-decision state $S_t^a = (I_t)$ representing the inventory at the end of period t.

Stochastic information and transition function. After a decision a_t is taken in state S_t, stochastic information

$$\omega_{t+1} = (d_{t+1}^{\omega}, p_{t+1}^{\omega}, q_{t+1}^{\omega})$$

is revealed about the demand of the next period d_{t+1}^{ω} as well as the purchase prices p_{t+1}^{ω} and available quantities q_{t+1}^{ω} per supplier. The transition function $\mathcal{T}(S_t^a, \omega_{t+1})$ leads to a new state $S_{t+1} = (\hat{I}_{t+1}, d_{t+1}^{\omega}, p_{t+1}^{\omega}, q_{t+1}^{\omega})$. The inventory $\hat{I}_{t+1}$ depends on the inventory of the post-decision state S_t^a, the realization of ω and the perishability percentages ϕ as follows:

$$\hat{I}_{kt+1} = I_{kt}(1 - \phi_k), \qquad \forall k \in K. \tag{18}$$

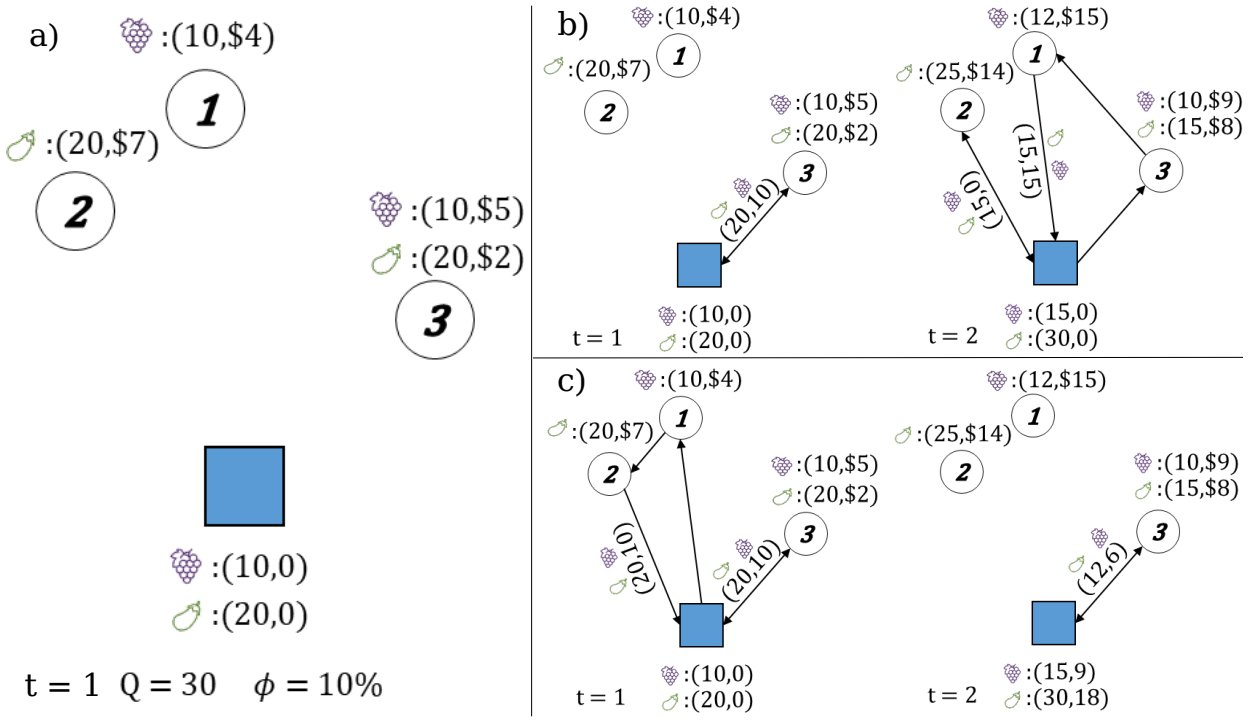

Figure 1 **Example of state, two potential decisions, and their consequence at next period**

Policy. A solution for a sequential decision process is a policy π. A policy assigns a decision $a_t = A^\pi(S_t)$ to every state S_t. The overall set of policies is defined as Π. An optimal solution $\pi^* \in \Pi$ maximizes the expected reward:

$$\pi^* = \underset{\pi \in \Pi}{\operatorname{argmax}} \mathbb{E}\left[\sum_{t \in T} (R(S_t, A^\pi(S_t)), |S_0)\right], \tag{19}$$

starting from state S_0.

3.3. Example

In Figure 1, we give an small example to illustrate the components of the sequential decision process and to motivate our solution methodology. Figure 1.a presents the state S_1 at period $t = 1$. We assume a network with three distributed suppliers (circles). For the purpose of presentation, we omit the backorder supplier and the routing network details. We further assume a system with two different products, eggplants and grapes. Each supplier provides information about the products offered, together with the available capacity and purchase price of each of them. In this example, supplier 1 offers at most 10 units of grapes for a price of $4 per unit. Supplier 2 offers at most 20 units of eggplants for a price $7 per unit. Supplier 3 offers both products with a maximum capacity of 10 units of grapes, and 20 units of eggplants. The prices per unit are $5 and $2, respectively. The square represents the warehouse, which presents the demand information for each product

and the respective initial inventory. The first entry of the vector is demand, and inventory is the second. In this example, 10 units of grapes and 20 units of eggplants are required and none are in the warehouse at the beginning of $t = 1$. In this example, the vehicle capacity is $Q = 30$, and a perishability loss rate $\phi = 10\%$ is assumed for each product.

Figures 1.b and 1.c present potential decisions for period $t = 1$ and their consequence at period $t = 2$ after exogenous information ω_2 is revealed and another decision has been taken. In Figure 1.b, the potential decision is to buy what is needed to satisfy the demand of period $t = 1$ at the lowest price without keeping units in inventory at the end of period, (e.i., $I_1 = 0$ for all products). This decision is the most cost-efficient at $t = 1$ and leads to an initial inventory $\hat{I}_2 = 0$ at $t = 2$ for each product. Note that even though supplier 1 has the lowest price for grapes, the routing cost of visiting reduces the overall reward, hence, in the decision, the grapes are purchased from supplier 3. In period $t = 2$, after new demand occurred and supply and prices were revealed, the subsequent decision is to send two vehicles due to capacity constraints, one to supplier 2 and one to suppliers 1 and 3. Thus, while cost-efficient in $t = 1$, this decision leads to inflexibility and potentially high costs in $t = 2$.

Another potential decision for period $t = 1$ is presented in Figure 1.c. In this decision, demand is satisfied by purchasing from supplier 3, and units are kept in inventory by purchasing from suppliers 1 and 2. At the end of the period $t = 1$, 20 units of grapes and 10 units of eggplants are in inventory, leading to an initial inventory of 18 and 9 at $t = 2$ after applying the perishability loss rate, respectively. In contrast to the decision in Figure 1.b, the decision presented in Figure 1.c leads to inventory at the beginning of period $t = 2$, which allow to satisfying the realized demand by only visiting supplier 3.

The example illustrates the challenge of balancing routing and purchasing cost in every state while determining effective inventories for the future. A visit to suppliers 1 and 2 at $t = 1$ as proposed in Figure 1.c generates savings in operation by anticipating changes in demand, purchase prices, and quantities available from suppliers. The purchase cost decreases because realized prices at $t = 1$ are lower compared to $t = 2$. The routing cost decreases for two reasons: the movement from 1 to 2 in $t = 1$ is made instead of going through the arc from 3 to 1 in $t = 2$; in $t = 2$, a direct shipment is made to supplier 3 instead of from supplier 2, which is farther away.

4. Method

In the following, we present the proposed method for our problem. We first give a motivation and overview of the general procedure, and then we define the individual components of our method in detail.

4.1. Motivation and overview

Solving the problem is challenging for two reasons: searching the vast decision space and evaluating the decisions with respect to future uncertainty realizations and their consequent potential decisions. We will discuss the two challenges in the following and motivate our solution approach addressing both in an integrated fashion.

- **Search:** Decisions comprise three interdependent components: the inventory level of each product, the purchasing decisions for each product across the set of suppliers to achieve the desired inventory, and an efficient routing of the vehicles to collect the purchased products. Even though routes for the problem at hand usually only have a few stops due to capacity constraints (about 1 to 3 in practice and in our experiments), the latter is an NP-hard optimization problem by itself. Furthermore, the three decisions are intertwined. Inventory decisions balance current cost and future savings but should also be made having both purchasing and routing cost in mind.
- **Evaluate:** Decisions are made under uncertainty in future realizations of demand, supply, and purchase prices. Thus, the value of a decision can only be determined at the end of the process. There is also a cost trade-off over the periods. For example, having available inventory at the beginning of the next period may require additional purchasing and routing, but likely reduces the corresponding purchasing and routing cost of that period. In contrast, high inventory levels require more purchasing and routing cost now plus a potential loss due to perishability, however, they allow more flexible decision making in the next period(s).

In summary, a methodology is required that allows an integrated consideration of the decision space while anticipating future realizations in demand, supply and purchase prices as well as potential future decisions. To this end, we present our **STAR**-approach (**ST**ochastic lookahead with **A**daptive **R**outing approximation). In every state, the **STAR**-approach samples a set of scenarios reflecting future uncertainty in demand, supply, and purchase prices over a limited set of future periods. **STAR** then solves the corresponding two-stage stochastic program, finding an integrated solution with respect to all scenarios. This solution then induces the purchasing and inventory decision for the current state. We note that this is a significant difference compared to scenario-decomposition lookahead methods like the multiple-scenario-approach (MSA, Bent and Van Hentenryck 2004) where the scenarios are solved individually and then the "average" solution is implemented.

Solving the two-stage stochastic program consists in finding an integrated inventory, purchasing and routing decision in all periods and scenarios, as well as a consideration of all individual solutions in the first stage. Even for only one scenario as required for methods such as the MSA, this is computationally challenging due to the large decision space (Zehtabian and Ulmer 2023). This challenge amplifies when routing needs to be considered in a two-stage stochastic program (Spliet

and Gabor 2015, Spliet, Dabia, and Van Woensel 2018). Thus, our **STAR**-policy considers a simplified decision space, maintaining inventory and purchasing decisions but only considering an approximation of the routing. While inventory and purchase decisions are fully considered in the optimization via variables I and z, the routing decisions in every sample path is reduced. Instead of determining the full routing via variables x, we assume direct trip costs $(c \cdot \hat{\tau}_m)$ from the warehouse to every supplier m if goods are purchased, with $\hat{\tau}_m := \tau_{0m} + \tau_{m0}, \forall\, m \in M$. Based on this approximation, we determine the inventory and purchasing decisions for a given state and solve the corresponding detailed routing part afterwards. This procedure reduces the decision space significantly and allows solving the two-stage stochastic program with commercial solvers.

Direct trips overestimate the actual routing cost as they ignore consolidation of potential close-by suppliers. Thus, we capture this consolidation potential by assigning each supplier m a weight γ_m. A weight close to zero indicates high consolidation potential while a weight close to one indicates that the supplier is usually visited via direct trip. The routing cost of a supplier m is then the cost of the direct trip multiplied by the weight $(\gamma_m \cdot c \cdot \hat{\tau}_m)$.

Determining the weights analytically is challenging as we show in our computational study. Recalling the example from the previous section, it is likely that $\gamma_1 < \gamma_3$, as the central position of supplier 1 might increase consolidation potential. However, the consolidation potential does not only depend on the location of a supplier, but also on their subset of products and quantities offered as well as the purchase prices. Even though a supplier might be very close to another supplier, there might be only limited consolidation, e.g., because the other supplier is very expensive or because the first supplier offers so much supply that the vehicle's capacity is already reached. Furthermore, the actual consolidation potential depends on the policy applied. For example, a policy aiming for high inventory values may lead to more direct trips compared to a policy that only purchases the bare minimum every day. Thus, we propose adaptively learning the γ-values via simulation. We start with initial γ-values prescribing a STAR-policy π^γ. Then, over a number of iterations, we repeatedly apply policy π^γ and observe the actual routing decisions. We use the observations to update the routing approximation γ-values and consequently the policy π^γ for the next iteration. We repeat this procedure several times until convergence in routing approximation and decision making is achieved. The procedure is summarized in Figure 2. In step (1), the approximated routing cost defines the **STAR**-policy π^γ. In step (2), the policy is applied in simulations. In step (3), the simulations outcome is used to update the γ-values and the cycle begins again.

Given the framework by Powell (2021), we propose embedding a parametric cost function approximation (CFA) in a stochastic lookahead method and adaptively tuning the CFA-parameters. Ensuring linearity via the γ-values allows us to search the complex decision space in every state

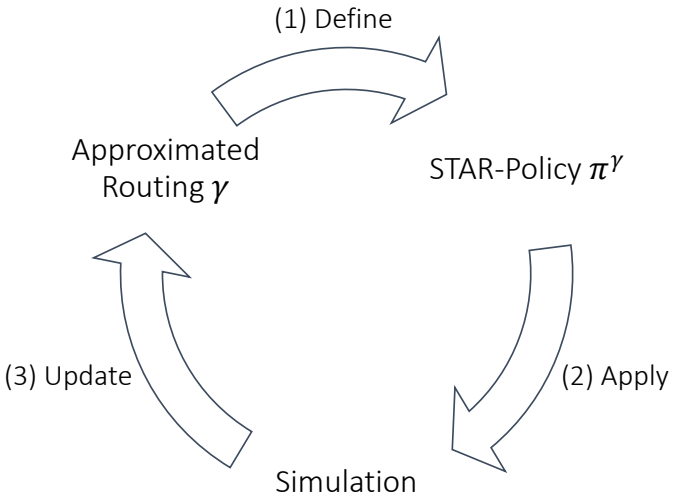

Figure 2 The iterative procedure to tune the γ-values of the STAR-approach

while also considering long-term impact of our decisions in detail. To the best of our knowledge, our method is the first to propose this general concept to the transportation literature.

In the following, we present the details of our algorithm. We first present the **STAR**-policy assuming γ-values are given and discuss how we determine the detailed routing decision given a purchasing decision z in a state. We then describe how we adaptively learn the γ-values.

4.2. Stochastic lookahead model

In every state S_t in period t, the **STAR**-policy solves a stochastic lookahead model. The model integrates a set of scenarios, i.e., sample paths $\omega \in \Omega$ and the current γ-values within a forward period horizon T'. Each sample path determines realized demand and supply volumes as well as prices for every period $t' \in T'$. The stochastic lookahead model is a two-stage stochastic program, and it is presented in Eqs. (20)-(30). The structure of the stochastic program is similar to the decision space defined in Section 3.2. However, it models several scenarios and time periods. Furthermore, the routing variables x are replaced with supplier selection variables e. The objective function, presented in Eq. (20), maximizes the expected reward over all scenarios. Eq. (21) guarantees the balance of inventory and demand satisfaction. Eqs. (22) and (23) guarantee that the purchase does not exceed the quantity of product on hand and the vehicle capacity. Eqs. (24), (25) and (26) are the non-anticipativity constraints, which ensure that the first-period decisions on the horizon T', corresponding to state S_t, are the same for all scenarios. Finally, Eqs. (27)-(30) define the variables domain.

$$\max \sum_{\omega \in \Omega} \frac{1}{|\Omega|} \left(\sum_{t' \in T'} \left(\sum_{k \in K} \left(r_k d'_{kt'\omega} - \sum_{m \in M_k} p'_{mkt'\omega} z_{mkt'\omega} \right) - c \sum_{m \in M} \gamma_m \hat{\tau}_m e_{mt'\omega} \right) \right) \tag{20}$$

s.t.

$$I_{kt'\omega} = I_{kt'-1\omega}(1-\phi_k) + \sum_{m\in M_k} z_{mkt'\omega} - d'_{kt'\omega}, \forall k \in K, \forall t' \in T', \forall \omega \in \Omega \tag{21}$$

$$z_{mkt'\omega} \leq q'_{mkt'\omega} e_{mt'\omega}, \qquad \forall k \in K, \forall m \in M_k, \forall t' \in T', \forall \omega \in \Omega \tag{22}$$

$$\sum_{k\in K} z_{mkt'\omega} \leq Q e_{mt'\omega}, \qquad \forall m \in M, \forall t' \in T', \forall \omega \in \Omega \tag{23}$$

$$z_{mkt\omega} = \sum_{\omega'\in\Omega} \frac{z_{mkt\omega'}}{|\Omega|}, \qquad \forall k \in K, \forall m \in M_k, \forall \omega \in \Omega \tag{24}$$

$$I_{kt\omega} = \sum_{\omega'\in\Omega} \frac{I_{kt\omega'}}{|\Omega|}, \qquad \forall k \in K, \forall \omega \in \Omega \tag{25}$$

$$e_{mt\omega} = \sum_{\omega'\in\Omega} \frac{e_{mt\omega'}}{|\Omega|}, \qquad \forall m \in M, \forall \omega \in \Omega \tag{26}$$

$$e_{mt'\omega} \in \{0,1\}, \qquad \forall m \in M, \forall t' \in T', \forall \omega \in \Omega \tag{27}$$

$$z_{mkt'\omega} \geq 0, \qquad \forall k \in K, \forall m \in M_k, \forall t' \in T', \forall \omega \in \Omega \tag{28}$$

$$z_{mkt'\omega} \leq q'_{mkt'\omega}, \qquad \forall k \in K, \forall m \in M_k, \forall t' \in T', \forall \omega \in \Omega \tag{29}$$

$$I_{kt'\omega} \geq 0, \qquad \forall k \in K, \forall t' \in T', \forall \omega \in \Omega \tag{30}$$

4.3. Routing algorithm

The stochastic program obtains the value of purchasing decisions z_t, inventories I_t, and suppliers selection e_t given a state S_t. Based on the variables, the **STAR**-policy determines routing decisions x_t and f_t. In the following, we describe the process conceptually. For details, we refer to Appendix A.2.1. First, a complete tour with the selected suppliers is created through a nearest neighbour algorithm. The creation of the complete tour follows Cuellar-Usaquén, Gomez, and Álvarez-Martínez (2021). An augmented graph is constructed with this complete tour and the purchased quantities (z_t). Then, the pool of routes that follow the complete tour order are extracted using the split procedure from Prins (2004) respecting the vehicle capacities and the maximum travel time. After the construction of the augmented graph that follows an acyclic-directed graph (ADG) structure, we solve the shortest path problem using the BellmanFord algorithm for ADG to find the best set of routes that minimize the travel time (Goldberg and Radzik 1993). These routes are then implemented in the decision x_t. The number of routes selected is equivalent to the number of vehicles used in decision f_t.

4.4. Adaptive routing cost approximation

Finally, we describe how we adaptively approximate the routing cost parameters γ over the iterations. We refer to Appendix A.2.2 for the algorithmic details. In the first iteration, initial values γ_m^0 are set, e.g., $\gamma_m^0 = 1.0$ for all suppliers m. These values induce an initial $\mathbf{STAR}^0$-policy. This policy is evaluated for a set of n simulations of the process ($n = 20$ in our computational study). In each simulation j, for each supplier m, the corresponding routing decisions are tracked and

for each observed routing in a period t, the corresponding $\hat{\gamma}_{j,t,m}$ is calculated based on the routes travel duration, the number of suppliers visited and the direct trip duration for the corresponding supplier. More specifically, $\hat{\gamma}_{j,t,m}$ is determined by the overall route duration containing supplier m in period t divided by the number of suppliers in that route, and then, normalized by dividing by the direct trip duration for supplier m. We note that such $\hat{\gamma}_{j,t,m}$-values are only calculated in periods t where supplier m was visited.

After all $\hat{\gamma}_{j,t,m}$-values of simulation j are collected, an average value $\bar{\gamma}_{j,m}$ is calculated. Based on preliminary tests, we also integrate γ_m^{i-1} in this calculation to avoid outlier values in case of only very few observations in simulation j. After all n simulations, the overall average $\bar{\gamma}_m$ is calculated based on the $\bar{\gamma}_{j,m}$-values of the individual simulation. Next, we set the new value γ_m^i as the weighted combination of old value γ_m^{i-1} and average value $\bar{\gamma}_m$:

$$\gamma_m^i = (1-\alpha_m)\cdot\gamma_m^{i-1} + \alpha_m \cdot \bar{\gamma}_m$$

Value α_m determines the step size of the updates. In our computations, we set $\alpha_m = 1/\sqrt{o_m}$ with o_m being the times supplier m was updated before. This allows us to focus more on later observations.

We illustrate the procedure via an example in Figure 3. In the example, we approximate the value for supplier 5 after iteration 4, i.e., γ_5^4. For the purpose of presentation, we only consider one simulation per iteration and a problem horizon of six periods. We recall that we integrate the service times in travel time values τ. We assume the travel time between supplier 5 and warehouse 0 is 95 minutes, and the service time for loading the products on the truck is 10 minutes. Thus, the direct trip travel time for supplier 5 is 200 minutes. Within the periods, supplier 5 was visited three times, in periods 1, 3, and 6. Each time, supplier 5 was part of a different route. In period 1, the supplier was visited with two other suppliers (1 and 4) and the overall tour duration was 300 minutes. In period 3, supplier 5 was visited together with supplier 9 and a tour duration of 280 minutes. In period 6, supplier 5 was visited in a direct-trip tour with duration of 200 minutes. For each of the three observations, we now calculate the relative γ-values as the ratio between partial routing time and direct trip duration. The partial routing time is the overall routing duration divided by the number of suppliers in the tour. In the first period, the relative $\hat{\gamma}_{1,1,5}$-value for supplier 5 is $\frac{100}{200} = 0.5$, the partial route duration $300/3 = 100$ divided by 200 minutes direct trip time. In period 3, the $\hat{\gamma}_{1,3,5}$-value is $\frac{140}{200} = 0.7$ and in period 6, $\hat{\gamma}_{1,6,5}$ is $\frac{200}{200} = 1.0$. If we assume a previous value of $\gamma_5^3 = 1.0$, we calculate

$$\bar{\gamma}_{1,5} = \frac{1.0 + 0.5 + 0.7 + 1.0}{4} = 0.8.$$

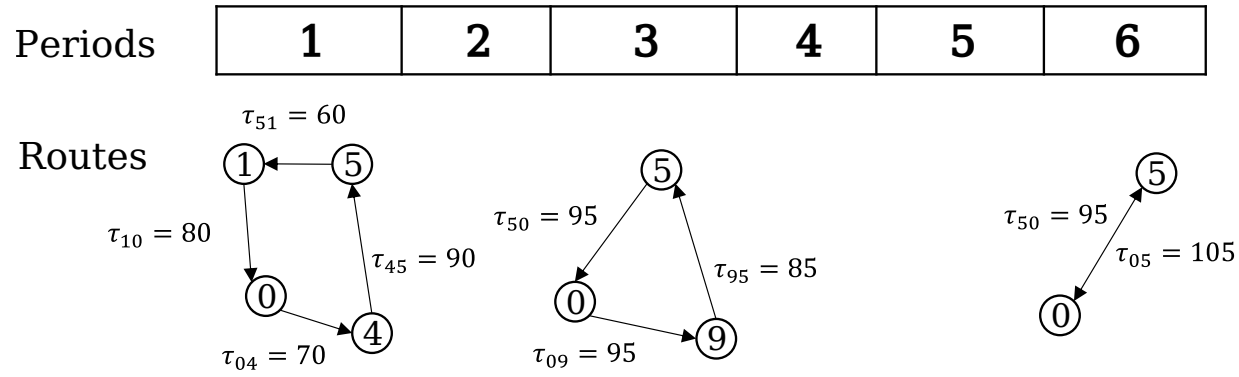

Figure 3 **Example for estimation γ for supplier** 5

Because the number of simulations is one for this example, $\bar{\gamma}_5 = \bar{\gamma}_{1,5}$. The new value γ_5^4 is then calculated as

$$\gamma_5^4 = (1 - 1/\sqrt{4}) \cdot \gamma_5^3 + (1/\sqrt{4}) \cdot \bar{\gamma}_5 = 0.5 \cdot 1.0 + 0.5 \cdot 0.8 = 0.9$$

We note that this procedure may produce values $\gamma_m > 1.0$ for some suppliers, mainly with small direct trip duration. For such suppliers, being consolidated with other suppliers the relative routing cost is higher than the direct trip cost. However, since the γ-values are considered jointly in the optimization, restricting values by 1.0 yields inferior results.

4.5. Implementation details

The parameters related to the **STAR**-approach were tuned based on preliminary experiments. The following parameter values were set. We set $h = 3$, the lookahead horizon to three periods, i.e., the stochastic program considers four periods total; the gap for the optimization solver to 5%; and the number of scenarios to $|\Omega| = 10$. This setting strikes the right balance between runtime and system performance, as shown in Appendix A.2.3. In the adaptive learning process, we use a number of $n = 20$ simulations per iteration and 60 iterations overall. Unless stated differently, the initial γ-values are fixed to 1 for all suppliers. For all the experiments, a computer with an Intel(R) Core(TM) i7-8650U CPU @ 1.90GHz 2.11 GHz was used with Windows 10 and 16 GB RAM. All implementation are coded on Python 3.9, and Gurobi 9.1.1 is used as optimizer.

5. Design of experiments

In the following, we present the instance setting details and the benchmark strategies for our computational experiments.

5.1. Instance setting

We consider a basis instance setting for our main experiments. This setting is also the foundation for our analysis described later. We base our instance design on publicly accessible data (de Planeación

and Sectorial 2005, Perfetti et al. 2013, Rodolfo Enrique and Mendoza Valencia 2017, UPRA 2018, DANE 2022, Valev 2023) and on our discussions the Colombian companies from the agri-food sector.

Layout. Our basis instance consists of 21 nodes (20 suppliers and one warehouse), five products, and a horizon of 20 periods ($|V| = 21, |M| = 20, |K| = 5, |T| = 20$). The warehouse is located at the center of a 250km times 250km region. The coordinates of the suppliers were sampled uniformly in the region.

Vehicles. Vehicles are provided by local freelancers. We assume vehicles have a capacity of six tons, $Q = 60$ (in units of hundreds of kilograms), as common for the smaller trucks operating in the region. We further assume a maximum working time of $l^{max} = 480$ minutes for travel and loading. We assume that the time it takes to load the products on the vehicle is 10 minutes per supplier. Travel times are calculated "as the crow flies", i.e., are Euclidean. We assume a travel speed of 60 km per hour. We assume compensation of one dollar $c = \$1$ per minute worked since it captures both vehicle and driver costs.

Products. Not all 20 suppliers offer all five products. Instead, we assume that a product is offered by at least 30% of suppliers. The suppliers for each product k were randomly selected with probability 30%. Because this is done individually for each product, some suppliers may offer a wider range of products than others. In the unusual case that a supplier m ends up with no product at all, we assign a single, random product. We set the perishability of all products to $\phi_k = 10\%, \forall k \in K$.

Supply and demand. The expected supply $\mu_{q_{mk}}$ of each supplier m and product k (if available) is drawn from a uniform distribution between 400 and 900 kilograms. The realization of each supply in a period follows a normal distribution with a coefficient of variation of 0.1. That means the standard deviation is 10% of the expected product supply. The expected demand μ_{d_k} of a product k is modeled similarly and is uniformly drawn between $1,000$ and $1,500$ kilograms. Again, we set the coefficient of variation to 0.1. Even though negative values are highly unlikely, we truncate all distributions at 0 kilograms (and $2 \cdot \mu_{q_{mk}}$ and $2 \cdot \mu_{d_k}$ kilograms to ensure symmetry). Finally, we set the demand for all products at the first period $t = 1$ to zero to allow an initial system "setup", $d_{k1} = 0, \forall k \in K$. Preliminary tests showed that this initial system setup does not affect the performance of the policies significantly compared to a setting without an initial setup period.

Modeling supply and demand based on the requirements leads to instances where the backorder option is generally not needed since the demand in a period can be satisfied by the available supply in the period.

Prices. For each supplier m and product k, we sample the expected prices $\mu_{p_{mk}}$ uniformly in the range from \$50 to \$120 per 100 kilograms. Again, we model the realizations with a normal distribution and a coefficient of variation of 0.1 and truncate the distributions at \$1 (and $2 \cdot \mu_{p_{mk}} - \$1$ to ensure symmetry). The revenue per product is 10% higher than the average expected values generated for product prices. The prices for the backorder option are set high, 250 times higher than the expected revenue. In our experiments, the backorder option was never used.

In our main experiments, we determine the expected demand, supply, and purchase prices values once and create 20 instance replications. In our analysis, we developed additional settings, e.g., by varying coefficient of variation for each source of uncertainty, by assuming correlation in suppliers' prices and supply volumes, by testing different cost and revenue percentages, by varying the perishability of products, and by assuming different vehicle capacity values. These additional settings are presented in Section 6.4.

5.2. Benchmarks

We test our policy to nine benchmark policies. We present problem-oriented and method-oriented benchmarks. Testing a policy without consideration of routing approximation (i.e., $\gamma = 0$) leads to substantially worse performance than all other tested policies. Thus, all benchmark policies rely on some form of approximation of γ.

Problem. We implement three problem-oriented strategies:

- **MYOPIC**: The first is a **MYOPIC**-policy maximizing the revenue per day by avoiding any additional travel and inventory. To approximate routing, the **MYOPIC**-policy assumes the same γ-value for all suppliers which is determined via enumeration (see Appendix A.3.1 for details).
- **EV**: As second benchmark, we implement an expected value policy (**EV**) related to Çabuk and Erol (2019). This policy plans on the expected values in purchase prices, supply, and demand. Algorithmically, it only considers one deterministic scenario of the expected values in prices, supply, and demand. The γ-values are tuned as for **MYOPIC**.
- **PFA**: The third benchmark is a policy function approximation (PFA). The **PFA**-policy mimics the practical idea to buy more than needed in case the prices are below average. To this end, for all products, the policy considers a percentage of the revealed demand at the beginning of each period and buys up to this percentage more if the product is cheaper than expected. The best percentage is determined via enumeration. The same γ-value is used as for **MYOPIC**. Details of the **PFA**-policy are presented in Appendix A.3.2.

Method. We test six method-oriented policies to investigate the value of scenario-generation and adaptive routing approximation. To this end, we test different alternatives to approximate the γ-values for the suppliers (for algorithmic details, we refer to Appendix A.3.3):

- **ST-ONE**: We assume a single γ-value for all suppliers. This value is determined via enumeration.
- **ST-MY**: This policy follows Liu and Luo (2023). The γ-value of each supplier is determined by running policy **MYOPIC** and approximating the γ-values as described in Section 4.
- **ST-DIST**: This policy links the γ-values of a supplier to the distance to other suppliers. To this end, we set γ_m relative to the the number of suppliers in supplier m's neighborhood defined via a travel time radius. We search for the best travel time radius via enumeration.
- **ST-CAPA**: This policy links the γ-values to the relative product subset and quantities offered of a supplier assuming that a supplier with a wider subset of products and higher supply quantities is part of more routes. We calculate a score for each supplier based on products and expected quantities. The γ-values are then determined based on the suppliers' scores relative to other suppliers. Suppliers with more products and offer receive a smaller γ-value.
- **ST-PRICE**: This policy is similar to **ST-CAPA**, but makes γ-values dependent on the purchase prices. It assumes that a supplier is more often part of a route if the prices are comparably cheap. The suppliers' scores depend on the relative expected prices of the offered products. A supplier with cheaper prices is assigned a smaller γ-value.
- **ST-DCP**: This policy combines the previous three. The γ-values are set as the average of the three individual values.

6. Computational study

In this section, we present our computational study. We first compare the solution quality of the policies. We then take a closer look at our policy's functionality analyzing the approximated γ-values, the accuracy of the routing cost approximation, and the changes in decision making. Finally, we present a sensitivity analysis on selected problem parameters.

6.1. Solution quality

To compare the solution quality of the policies, we draw on the **MYOPIC**-policy as basis benchmark and calculate the improvement in reward as follows. Let $V^{\mathbf{MYOPIC}}$ be the objective value of the myopic policy and V^{π} the value of another policy, then the improvement of the other policy over **MYOPIC** is calculated as

$$\frac{V^{\pi} - V^{\mathbf{MYOPIC}}}{V^{\mathbf{MYOPIC}}}.$$

First, we compare **STAR**-policy to the problem-oriented benchmarks without individual routing cost approximation (**PFA**, **EV**, **ST-ONE**) and to the benchmark of Liu and Luo (2023), **ST-MY**. The results are shown in Figure 4. The x-axis depicts the policies, the y-axis shows the

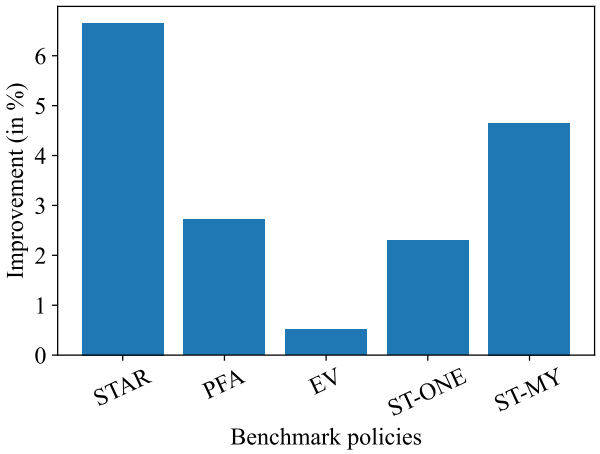

Figure 4 **Comparison of main policies to MYOPIC**

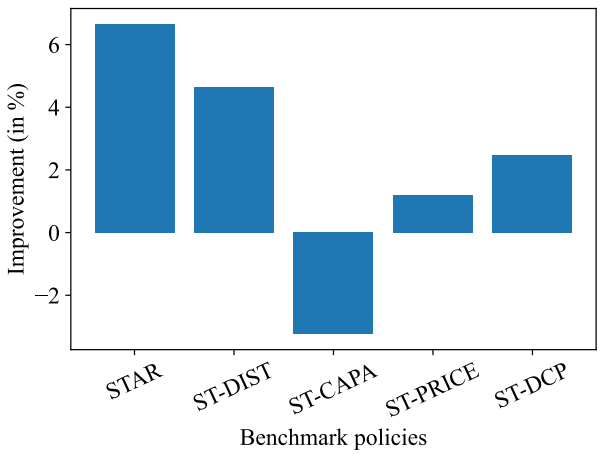

Figure 5 **Alternative γ-value estimations**

improvement over **MYOPIC**. We observe that all policies improve compared to myopic decision making. Thus, building up inventory for future periods proves valuable. We also observe that the improvement of **EV** with less than 1% is very limited compared to **ST-ONE** with more than 2%. This indicates that consideration of the uncertainty is essential for successful decision making. Finally, we see a gradually improvement over **ST-ONE**, **ST-MY** (ca. 5%), and **STAR** (nearly 7%). Thus, approximating individual routing cost is important. However, as **STAR** clearly outperforms **ST-MY**, the adaptive learning of the γ-values is very valuable. Interestingly, as we show in Appendix A.4.1, while the **STAR**-policy achieves the highest objective values, it does so with one of the smallest variances among all policies, thus, also leading to more reliable solutions.

To analyze the value of adaptively learning individual γ-values in more detail, we compare the performance of **STAR** to the remaining approximation policies **ST-DIST**, **ST-CAPA**, **ST-PRICE**, and **ST-DCP**. The results are shown in Figure 5. We observe that **STAR** outperforms all other policies. We further observe that the approximation based on distance via **ST-DIST** results in an acceptable performance while consideration of capacity (**ST-CAPA**) and price (**ST-PRICE**) as well as their combinations (**ST-DCP**) perform rather poorly. This indicates that the location of a supplier indeed plays a major role in the routing and corresponding routing cost. Policy **ST-CAPA** leads to even worse results than **MYOPIC**. Thus, the expected capacity and product range cannot be transferred easily to approximated routing cost, likely, because while large expected supply volumes may lead to more visits in routes (small γ), it may also consume most of the vehicle's capacity, i.e., no other suppliers can be visited in the same route (large γ).

6.2. Iterative routing cost approximation

In the following, we investigate the approximation process and the final γ-values of our policy **STAR** in more detail. To analyze how well the γ-values approximate the real routing cost, we

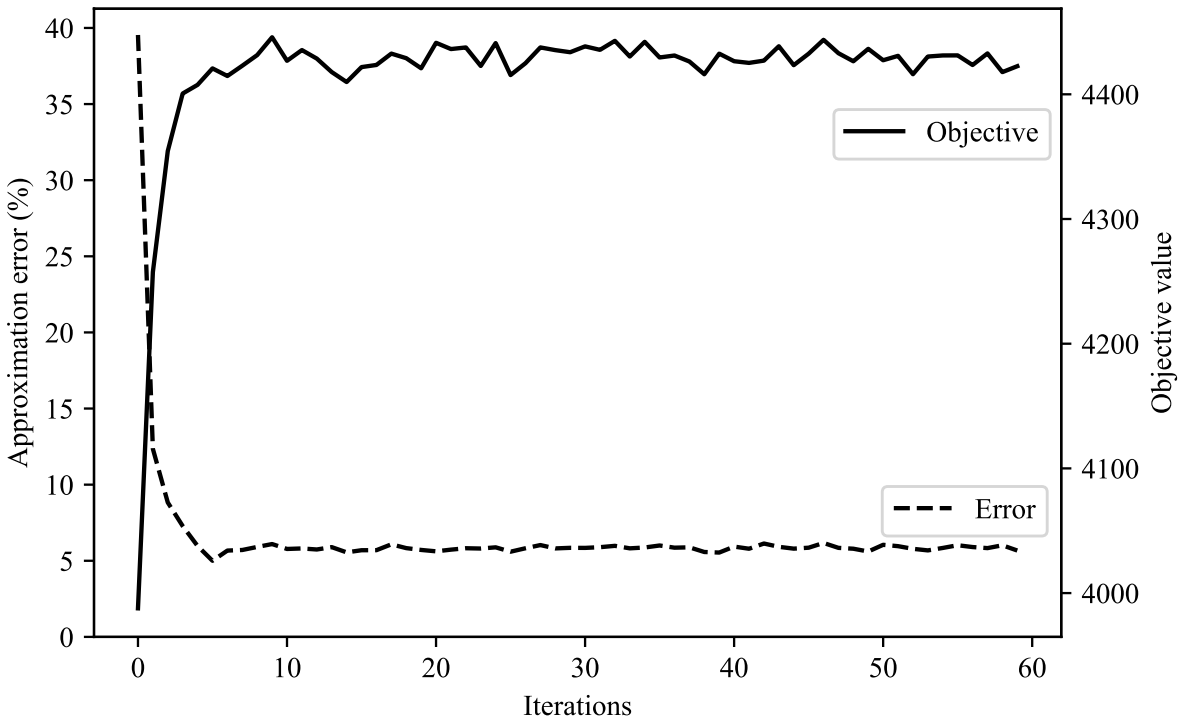

Figure 6 Absolute error of routing cost and approximate cost

calculate the average difference between real routing cost and approximated routing cost in every iteration. The results are shown in Figure 6 by the dashed line and left y-axis. We observe that originally, at iteration 0, the difference is substantial with about 40% of error. In that initial iteration, the values are $\gamma_m = 1$ for all suppliers $m \in M$. Thus, direct trips are assumed while, in reality, consolidation occurs. After this initial iteration, we observe a fast decrease in approximation error and convergence after five iterations to around 5% of error. Thus, on average, the approximation of our γ-values is only about 5% different than the real cost and, therefore, quite accurate.

Notably, while the routing approximation error converges after five iteration, the **STAR**-policy still improves in later iterations as shown in solid line on the right y-axis of Figure 6. In line with the fast improvement in approximation error in the first iterations, we observe a fast increase in the objective value. Still, convergence is reached after about 10 iterations. Thus, even once the routing approximation is accurate, decision making changes for a few more iterations.

Next, we analyze the approximated γ-values in detail. First, we plot them in space, shown in Figure 7. The figure represents the 250km times 250km service area with the 20 suppliers. The depot 0 is located in the center. The γ of each supplier is adjacent to the supplier's location, e.g., supplier 1 on the left has a value of $\gamma_1 = 1.0$. We observe that there is no clear picture in how the γ-values form. However, some insights can be identified:

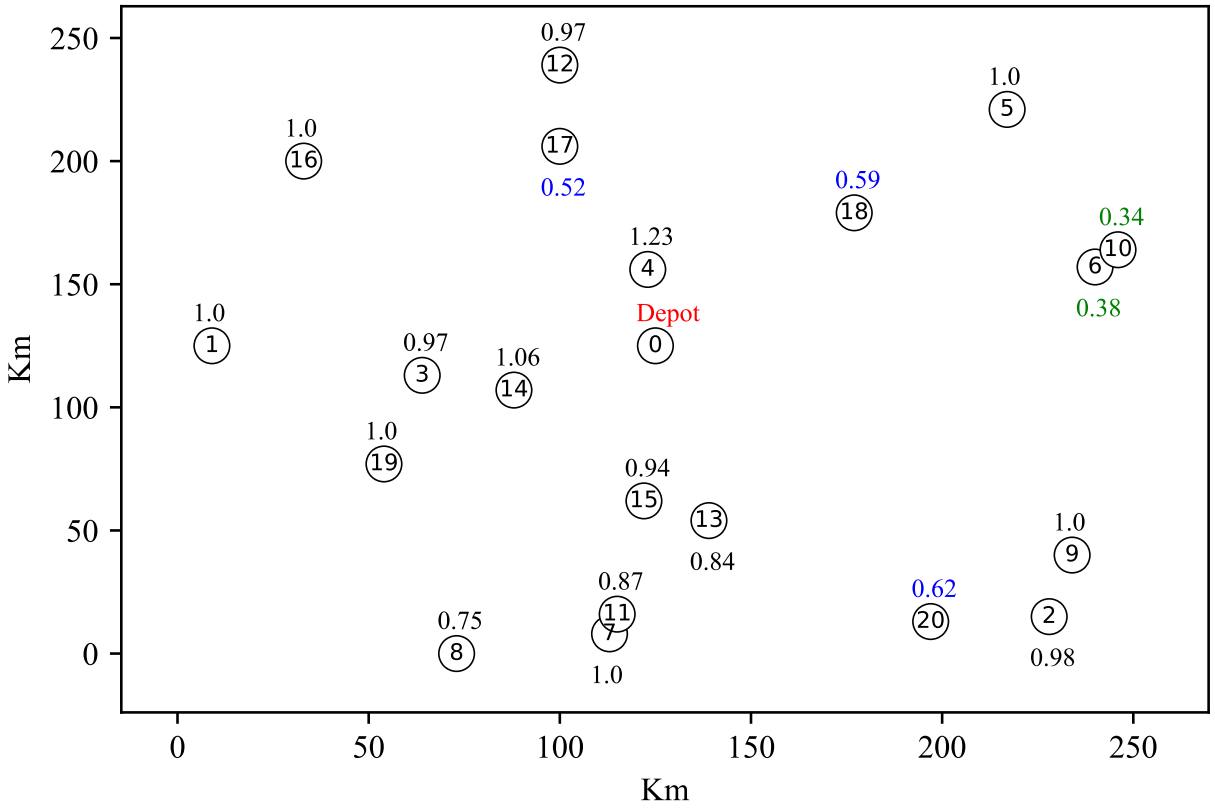

Figure 7 **Geographical distribution γ-values**

- As expected, suppliers that are rather isolated have a higher γ-value, e.g., suppliers 1, 5, or 16 have values of $\gamma = 1.0$. This indicates that consolidation with the suppliers is usually not possible and explains the relatively good performance of **ST-DIST**.
- Suppliers with smaller γ-values either occur in clusters (e.g., suppliers 6 and 10) or are "on the road" to other suppliers (e.g., supplier 17, 18, or 20). This can be expected as well. However, being in a cluster or "on the road" to another supplier does not automatically lead to smaller γ-values (e.g., for suppliers, 7, 11, or 15). Here, other factors such as expected prices and capacity come into play.
- In two cases, the values are even slightly higher than one: $\gamma_{14} = 1.06$ and $\gamma_4 = 1.23$. These two supplies have rather short direct trip distances and the shared γ-calculation may lead to the values higher than one. Interestingly, we tested limiting γ-values in the range $[0, 1]$ and the results were inferior. As the γ-values are used in a joint calculation of the routing cost for several suppliers, limiting the values led to an underestimation of the concerted cost.

For an additional analysis of the γ-values with respect to prices and capacity, we refer to Appendix A.4.2.

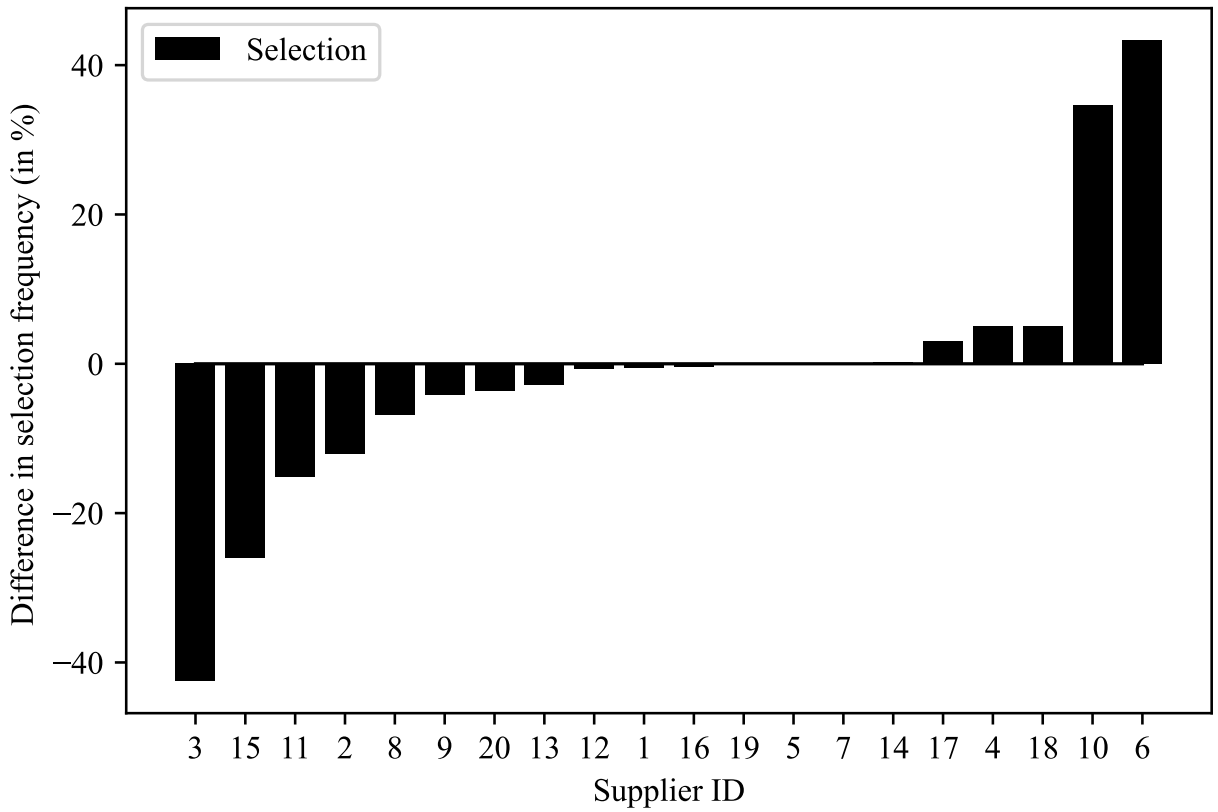

Figure 8 **Difference in frequency of STAR and MYOPIC supplier selection**

6.3. Decision making

In the following, we analyze how anticipation by the **STAR**-policy changes decision making compared to the **MYOPIC**-policy. To this end, we analyze how supplier selection is made. Then, we show how the policies affect inventory quantities and purchases, and finally, how this affects purchasing and routing costs.

Figure 8 presents the frequency of supplier selection. The x-axis presents the IDs of the suppliers, and the y-axis the difference in selection frequency between **STAR** and **MYOPIC**. We observe that for many suppliers the difference is small. These suppliers can be seen as "obvious" choices in both policies as they are either cheap or well-located (or the opposite). However, for a few suppliers, we see significant differences. The global routing approximation of the **MYOPIC**-policy does not discover every consolidation potential with fitting suppliers in its selection, so it selects suppliers 3, 11, and 15 more often. Even though the suppliers have low prices, the neighboring suppliers are not as attractive as indicated by their high γ-values shown in Figure 7. Due to the individual routing approximation, the **STAR**-policy selects suppliers 6, 10, and 18 more often, which despite being far from the depot, allow consolidations among them at low prices, indicated by their smaller γ-values shown in Figure 7.

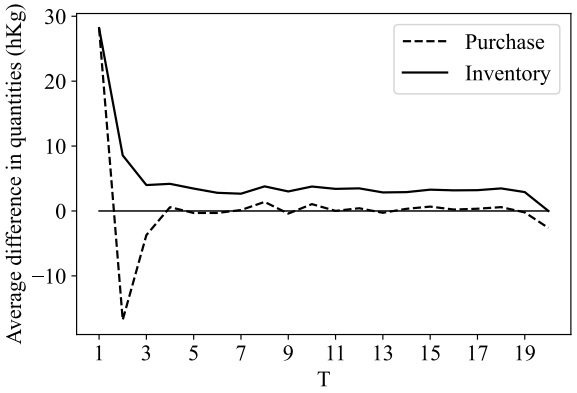

Figure 9 **Average difference in quantities purchased and inventory levels by STAR and MYOPIC**

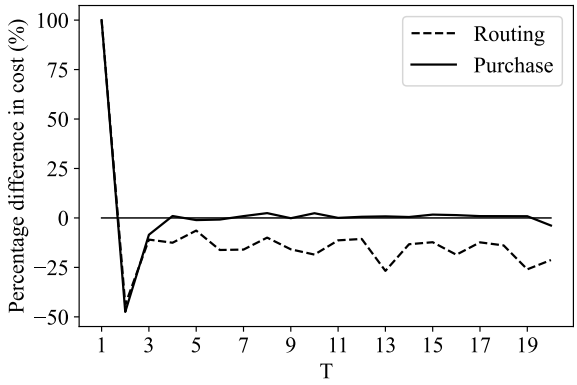

Figure 10 **Average percentage difference in purchase and routing cost by STAR and MYOPIC**

The anticipation of our **STAR**-policy also changes how much is purchased in every period and how much inventory is stored. Figure 9 presents the changes in purchase and inventory quantities. The x-axis shows the periods. The y-axis shows the average difference between the quantities purchased and quantities kept in inventory by **STAR** and **MYOPIC** policies. It can be seen that the most significant difference is found in the first three periods. In the first period that is without any demand, the **STAR**-policy decides to replenish almost 30 hundred kilograms, as opposed to the **MYOPIC**-policy, which buys nothing. In the second and third period, **MYOPIC** purchases more than **STAR** and the surplus in inventory decreases gradually until it reaches a constant level of about 500kg per period until period 19. In the final period, the **STAR**-policy consumes all remaining inventory.

While in periods 4 to 19, the amount of purchased products is about the same for **STAR** and **MYOPIC**, the cost for purchasing differs significantly. This can be seen in Figure 10.

The x-axis shows the periods. The y-axis shows the percentage difference in purchase and route costs between the **STAR** and **MYOPIC** policies. It can be observed that after the purchase stabilizes, from period 4 onwards, the purchase costs per period are very similar between the policies. However, the **STAR**-policy generates savings of 10% to nearly 30% in routing cost in every period. Thus, purchasing and holding the "right" inventory in every period allows for more flexible and cheaper routing in the next.

6.4. Problem's parameters analysis

In the following, we perform an analysis of the problem's parameters. To this end, we generate instances, each differing in exactly one dimension. We train our method and all benchmarks for the individual instance settings. As before, we split training and evaluation instances.

Table 1 Performance for different coefficients of variation (left) and for correlation (right). Relative difference over the general "standard" configuration on the top and improvement of STAR over MYOPIC on the bottom.

	COV	Demand	Prices	Supply	Prices (ρ)	Supply (ρ)
vs. Standard	0	-0.1%	-9.4%	0.6%	-59.9%	-18.1%
	0.1	0.0%	0.0%	0.0%	-39.6%	-19.2%
	0.2	0.2%	18.3%	-1.7%	0.5%	-22.3%
vs. **MYOPIC**	0	6.3%	4.5%	5.6%	10.0%	8.9%
	0.1	6.7%	6.7%	6.7%	4.1%	7.3%
	0.2	6.9%	5.6%	6.5%	4.1%	7.2%

Uncertainty. In our main experiments, we assumed uncertainty in demand, supply, and prices with a coefficient of variation of 0.1. We now analyze how more of less uncertainty impacts the performance. To this end, we vary the coefficient of variation (COV) for one of the three sources of uncertainty from 0.0 and 0.2. Setting the value to 0.0 results in the corresponding source of uncertainty to be deterministic. The results are depicted on the left side of Table 1. On the top part of the table, we compare the objective value of **STAR** for the specific setting to its objective for the standard setting. In the bottom part, we compare the improvement of **STAR** over **MYOPIC** for the individual settings. For completeness, the middle row shows the standard setting with a COV of 0.1. Each column represents one source of uncertainty and the impact of varying their COV ceteris paribus. For example, the first value in the Demand-column in the top of the table, -0.1%, indicates the changes of the objective value if demand-values are certain and prices and supply still have a COV of 0.1 compared to the standard setting with a COV of 0.1 also for demand. The first entry in the bottom part, 6.3%, indicates the improvement of **STAR** compared to **MYOPIC** for this specific setting. Looking at the bottom part of the table, we observe that **STAR** outperforms **MYOPIC** regardless of the instance setting with improvements between 4.1% and 10.0%. When looking at the top of the table, we observe that the overall objective is affected by the changes in COV. We observe that uncertainty in supply and demand has relatively small impact on the overall objective value with all changes being below 2%. However, having different volatility in the prices impacts the performance significantly. Interestingly, having fixed prices (COV of 0) is not necessarily beneficial, but leads to a reduction in objective value of 9.4% while an increasing in price volatility can even increase the objective. This rather counter-intuitive observation can be explained by the increasing opportunities for cheap purchases in case of varying prices.

Correlation. In our main experiments, we assume expected prices $\mu_{p_{mk}}$ and supply volumes $\mu_{q_{mk}}$ of different products k are independent per supplier m. Now, we generate a correlated setting where supplier have "low" and "high" prices (or supply volumes). We do this by sampling the value for the first product freely. If it is below average (i.e., "low"), we ensure that all other values are below average as well by "mirroring" the sampled values on the expected value if necessary. We

calculate the improvements of **STAR** to the standard setting and compared to **MYOPIC** for the three aforementioned COVs. We keep the tuning parameters of our method the same, however, based on preliminary tests, we start with initial γ-values of zero. The results are shown on the right side of Table 1, indicated by "ρ". Again, our policy clearly outperforms the benchmark also for the correlated cases. Interestingly, correlation especially in the prices leads to substantially worse objective values. The reasons are twofold. First, correlation reduces flexibility in the routing and purchasing as some suppliers likely are too expensive regardless of the product. Thus, the set of suppliers to purchase from is reduced significantly leading to less consolidation or longer routes. Same is true for the supply, likely, because some suppliers might not be worth visiting since their supply is too small. Second, in our model, repeated visits of suppliers are prohibited. With correlated prices or supply, it might become important to visit a supplier with more vehicles, either, because of the cheap prices for all products or because of the vast amounts of supply.

Cost and revenue. In our general settings, we assume vehicle cost of \$1 per minute and a revenue of 10% above the average purchase price. In our analysis, we generate instances setting the cost to \$0.5 ,\$1.0, \$1.5, and \$2.0. We further generate instances varying the revenue, setting it to 20% and 50% more than the average purchase price. The results are shown in Figures 11 and 12. The x-axis depicts the changes in cost (revenue). The y-axis shows the performance of **STAR** and **MYOPIC** relative to the standard setting. This allows us to depict both the value of anticipation and the impact of varying cost and revenue. E.g., a value of -100% can be seen as a break-even point ceteris paribus.

As expected, the overall objective value decreases with increasing routing cost. We further observe the improvement of **STAR** increases as well. If routing is cheap, **MYOPIC** performs well since it buys cheap and does not built any inventory that perishes over the periods. If routing is expensive, the gap between **STAR** and **MYOPIC** becomes substantial. Here, the savings in daily routing seen in the previous section multiple. Thus, anticipatory decision making and active building of inventory is of particular importance if routing costs are high. For the revenue margin, the development is the other way around. The explanation is similar as for the routing cost. With high revenue margins, the routing cost become relatively smaller and the **MYOPIC** strategy performs comparably well.

Vehicle capacity. Next, we evaluate the value of having a truck of different size. To this end, we increase or decrease in vehicle capacity by testing $Q = \{40, 80, 100\}$ instead of 60. The results are shown in Figure 13. We observe that for our setting, the small trucks already perform relatively well and there is only a mild improvement of about 2.5% for larger trucks. However, we observe a slight increase in the gap between **STAR** and **MYOPIC** with increasing truck capacity. With larger capacity, more inventory can be built without additional routing cost. Thus, there is benefit for **STAR** but not for **MYOPIC**.

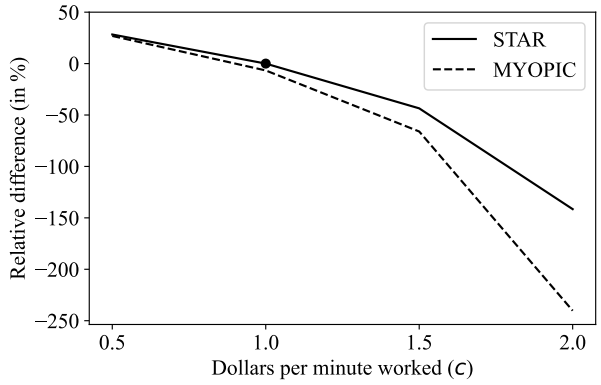

Figure 11 **Changing routing cost**

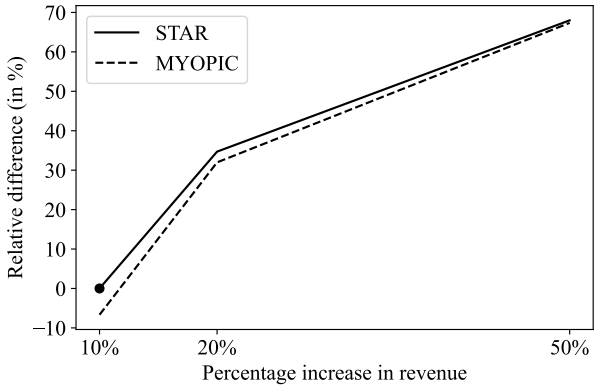

Figure 12 **Changing revenue margin**

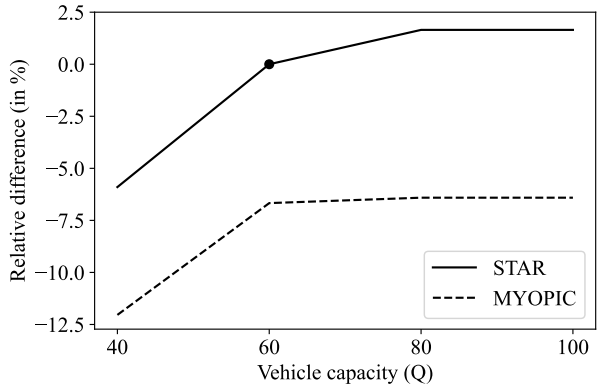

Figure 13 **Changing vehicle capacity**

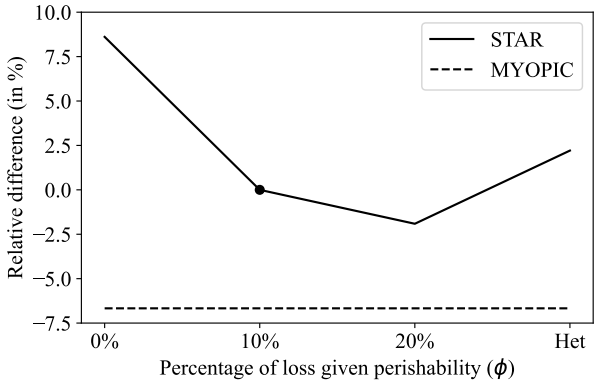

Figure 14 **Changing perishability**

Perishability. Finally, we analyze how perishability impacts the policies' performances. To this end, we generate settings with $\phi = 0\%$ and 20% perishability instead of 10%. We further generate a set of heterogeneous perishability for different products where we set the perishability to 0% for the first two products, 10% for the third product, and 20% for the last two products. Thus, the average perishability remains at 10%. The results are shown in Figure 14. As the **MYOPIC** policy does not build up any inventory, the objective value does not change with changing perishability. However, we observe that perishability has an impact on the performance of **STAR**. Without any perishability, significantly more revenue can be achieved while with a perishability of 20%, the objective value decreases slightly, as expected. Notably, in case of heterogeneous perishability, the objective value increases, likely, because the policy learns to purchase the longer-lasting products with $\phi = 0$. Thus, it might be valuable for companies to consider investing in improved storage facilities, even if only for a subset of products.

7. Conclusion

In this paper we have shown how anticipatory decision making via stochastic lookahead can improve performance significantly in dynamic multi-period purchasing inventory routing for agri-food supply chains. We have further shown how to adaptively tune and embed a cost function approximation in the lookahead model and the benefits it brings. There are several avenues for future research.

Our computational study has shown that the supplier characteristics with respect to supply and prices are very important for successful operations. Future research may analyze how the setup impacts the system's performance in more detail. Based on such an analysis, some suppliers might be encouraged to offer a specific product or a guaranteed supply quantity via fixed contracts. Here, the selection of suppliers and their guaranteed quantities might be of particular interest. Further, repeated visits of suppliers within a period might be included in the model. At the same time, future research may analyze how to improve the experience of the suppliers and ensure long-term participation in the system, e.g., by consistent purchases or regular visits at regular times (Zehtabian and Ulmer 2023). While our work showed how to improve operations on the first mile between suppliers and warehouse, future work may consider a more integrated view of first and last mile. For example, customers with larger demand quantities (e.g., canteens or restaurants) may be served directly by the collection fleet and may even have their own inventory.

Further, we have modeled and analyzed several sources of uncertainty with demand, supply, and prices. However, additional uncertainties might be considered. For example, the perishability may be uncertain or may even depend on the suppliers the products were purchased from. For some suppliers, the volumes may be uncertain until the vehicle arrives at their location. We have also seen that the routing cost are a significant factor in the profitability of the business. In practice, the cost and availability of vehicles is often uncertain too. Future research may incorporate this additional uncertainty, e.g., by purchasing more when many cheap vehicles are available. Finally, there might be other companies involved in the business and the prices and available supply may not only vary over the days, but even within the day. Anticipatory, real-time updates of routes might be required to adapt to changing supply and prices.

While we have designed our **STAR**-method for the specific problem at hand, its functionality is a general contribution to the literature on approximate dynamic programming for dynamic problems with combinatorial decision spaces. This area is still relatively unexplored (Liu and Luo 2023, Hildebrandt, Thomas, and Ulmer 2023). Future research may transfer our method's general idea of integrating an adaptively trained cost function approximation in a stochastic lookahead to other problems in the space. We have also shown that our "soft" decomposition of the complex decision space (full inventory decisions, approximated routing decisions) allows for high-quality decision

making if the routing cost approximation is learned iteratively. While we selected our decomposition a priori based on domain knowledge, future research could develop automated methods that stepwise change the decomposition based on its approximation error.

Acknowledgments

The authors thank Merqueo S.A.S. for their valuable input in preparation of this work. Marlin Ulmer's work is funded by the Emmy Noether Program, number 444657906 of the Deutsche Forschungsgemeinschaft (DFG). This support is gratefully acknowledged.

References

Adelman D, 2004 *A price-directed approach to stochastic inventory/routing. Operations Research* 52(4):499–514.

Albareda-Sambola M, Fernández E, Laporte G, 2014 *The dynamic multiperiod vehicle routing problem with probabilistic information. Computers & Operations Research* 48:31–39.

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

Angelelli E, Bianchessi N, Mansini R, Speranza MG, 2009 *Short term strategies for a dynamic multi-period routing problem. Transportation Research Part C: Emerging Technologies* 17(2):106–119.

Angelelli E, Mansini R, Vindigni M, 2016 *The stochastic and dynamic traveling purchaser problem. Transportation Science* 50(2):642–658.

Aouadni S, Aouadni I, Rebaï A, 2019 *A systematic review on supplier selection and order allocation problems. Journal of Industrial Engineering International* 15(1):267–289.

Avraham E, Raviv T, 2021 *The steady-state mobile personnel booking problem. Transportation Research Part B: Methodological* 154:266–288.

Baty L, Jungel K, Klein PS, Parmentier A, Schiffer M, 2023 *Combinatorial optimization enriched machine learning to solve the dynamic vehicle routing problem with time windows. arXiv preprint arXiv:2304.00789* .

Bent RW, Van Hentenryck P, 2004 *Scenario-based planning for partially dynamic vehicle routing with stochastic customers. Operations Research* 52(6):977–987.

Beraldi P, Bruni ME, Manerba D, Mansini R, 2017 *A stochastic programming approach for the traveling purchaser problem. IMA Journal of Management Mathematics* 28(1):41–63.

Bertazzi L, Bosco A, Guerriero F, Lagana D, 2013 *A stochastic inventory routing problem with stock-out. Transportation Research Part C: Emerging Technologies* 27:89–107.

Bertazzi L, Laganà D, Ohlmann JW, Paradiso R, 2020 *An exact approach for cyclic inbound inventory routing in a level production system. European Journal of Operational Research* 283(3):915–928.

Bianchessi N, Irnich S, Tilk C, 2021 *A branch-price-and-cut algorithm for the capacitated multiple vehicle traveling purchaser problem with unitary demand. Discrete Applied Mathematics* 288:152–170.

Bianchessi N, Mansini R, Speranza MG, 2014 *The distance constrained multiple vehicle traveling purchaser problem. European Journal of Operational Research* 235(1):73–87.

Brinkmann J, Ulmer MW, Mattfeld DC, 2019 *Dynamic lookahead policies for stochastic-dynamic inventory routing in bike sharing systems. Computers & Operations Research* 106:260–279.

Brinkmann J, Ulmer MW, Mattfeld DC, 2020 *The multi-vehicle stochastic-dynamic inventory routing problem for bike sharing systems. Business Research* 13:69–92.

Çabuk S, Erol R, 2019 *Modeling and analysis of multiple-supplier selection problem with price discounts and routing decisions. Applied Sciences* 9(17):3480.

Cheng C, Qi M, Wang X, Zhang Y, 2016 *Multi-period inventory routing problem under carbon emission regulations. International Journal of Production Economics* 182:263–275.

Chitsaz M, Cordeau JF, Jans R, 2019 *A unified decomposition matheuristic for assembly, production, and inventory routing. INFORMS Journal on Computing* 31(1):134–152.

Chitsaz M, Cordeau JF, Jans R, 2020 *A branch-and-cut algorithm for an assembly routing problem. European Journal of Operational Research* 282(3):896–910.

Cobb BR, 2016 *Inventory control for returnable transport items in a closed-loop supply chain. Transportation Research Part E: Logistics and Transportation Review* 86:53–68.

Coelho LC, Cordeau JF, Laporte G, 2014a *Heuristics for dynamic and stochastic inventory-routing. Computers & Operations Research* 52:55–67.

Coelho LC, Cordeau JF, Laporte G, 2014b *Thirty years of inventory routing. Transportation Science* 48(1):1–19.

Crama Y, Rezaei M, Savelsbergh M, Van Woensel T, 2018 *Stochastic inventory routing for perishable products. Transportation Science* 52(3):526–546.

Cuellar-Usaquén D, Gomez C, Álvarez-Martínez D, 2021 *A grasp/path-relinking algorithm for the traveling purchaser problem. International Transactions in Operational Research* 30(2):831–857.

DANE, 2022 *Tercer censo nacional agropecuario*. URL `http://microdatos.dane.gov.co/index.php/catalog/513/get_microdata`.

de Planeación OA, Sectorial GP, 2005 *Caracterización del transporte en Colombia: Diagnostico y proyectos de transporte e infraestructura*. URL `https://www.mintransporte.gov.co/descargar.php?id=455`.

Esmaeili-Najafabadi E, Nezhad MSF, Pourmohammadi H, Honarvar M, Vahdatzad MA, 2019 *A joint supplier selection and order allocation model with disruption risks in centralized supply chain. Computers & Industrial Engineering* 127:734–748.

Fukase E, Martin W, 2020 *Economic growth, convergence, and world food demand and supply. World Development* 132:104954.

Goldberg AV, Radzik T, 1993 *A heuristic improvement of the Bellman-Ford algorithm. Applied Mathematics Letters* 6(3):3–6.

Gu W, Archetti C, Cattaruzza D, Ogier M, Semet F, Speranza MG, 2022 *A sequential approach for a multi-commodity two-echelon distribution problem. Computers & Industrial Engineering* 163:107793.

Haferkamp J, Ulmer MW, Ehmke JF, 2023 *Heatmap-based decision support for repositioning in ride-sharing systems. Transportation Science* .

Halkier H, James L, 2022 *Learning, adaptation and resilience: The rise and fall of local food networks in Denmark. Journal of Rural Studies* 95:294–301.

Hammami R, Frein Y, Hadj-Alouane AB, 2012 *An international supplier selection model with inventory and transportation management decisions. Flexible Services and Manufacturing Journal* 24(1):4–27.

Hammami R, Temponi C, Frein Y, 2014 *A scenario-based stochastic model for supplier selection in global context with multiple buyers, currency fluctuation uncertainties, and price discounts. European Journal of Operational Research* 233(1):159–170.

Heinold A, Meisel F, Ulmer MW, 2023 *Primal-dual value function approximation for stochastic dynamic intermodal transportation with eco-labels. Transportation Science* .

Hildebrandt FD, Thomas BW, Ulmer MW, 2023 *Opportunities for reinforcement learning in stochastic dynamic vehicle routing. Computers & Operations Research* 106071.

Hosseini ZS, Flapper SD, Pirayesh M, 2022 *Sustainable supplier selection and order allocation under demand, supplier availability and supplier grading uncertainties. Computers & Industrial Engineering* 165:107811.

Iori M, Salazar-González JJ, Vigo D, 2007 *An exact approach for the vehicle routing problem with two-dimensional loading constraints. Transportation Science* 41(2):253–264.

Kang S, Ouyang Y, 2011 *The traveling purchaser problem with stochastic prices: Exact and approximate algorithms. European Journal of Operational Research* 209(3):265–272.

Keskin M, Branke J, Deineko V, Strauss AK, 2023 *Dynamic multi-period vehicle routing with touting. European Journal of Operational Research* .

Klapp MA, Erera AL, Toriello A, 2018a *The dynamic dispatch waves problem for same-day delivery. European Journal of Operational Research* 271(2):519–534.

Klapp MA, Erera AL, Toriello A, 2018b *The one-dimensional dynamic dispatch waves problem. Transportation Science* 52(2):402–415.

Laganà D, Laporte G, Vocaturo F, 2021 *A dynamic multi-period general routing problem arising in postal service and parcel delivery systems. Computers & Operations Research* 129:105195.

Liu S, Luo Z, 2023 *On-demand delivery from stores: Dynamic dispatching and routing with random demand. Manufacturing & Service Operations Management* 25(2):595–612.

Mafakheri F, Breton M, Ghoniem A, 2011 *Supplier selection-order allocation: A two-stage multiple criteria dynamic programming approach. International Journal of Production Economics* 132(1):52–57.

Majluf-Manzur ÁM, González-Ramirez RG, Velasco-Paredes RA, Villalobos JR, 2021 *An operational planning model to support first mile logistics for small fresh-produce growers. Production Research* 205–219.

Malladi KT, Sowlati T, 2018 *Sustainability aspects in inventory routing problem: A review of new trends in the literature. Journal of Cleaner Production* 197:804–814.

Manerba D, Mansini R, 2015 *A branch-and-cut algorithm for the multi-vehicle traveling purchaser problem with pairwise incompatibility constraints. Networks* 65(2):139–154.

Manerba D, Mansini R, 2016 *The nurse routing problem with workload constraints and incompatible services. IFAC-PapersOnLine* 49(12):1192–1197.

Manerba D, Mansini R, Riera-Ledesma J, 2017 *The traveling purchaser problem and its variants. European Journal of Operational Research* 259(1):1–18.

Mendoza A, Ventura JA, 2008 *An effective method to supplier selection and order quantity allocation. International Journal of Business and Systems Research* 2(1):1–15.

Miller CE, Tucker AW, Zemlin RA, 1960 *Integer programming formulation of traveling salesman problems. Journal of the ACM* 7(4):326–329.

Mjirda A, Jarboui B, Macedo R, Hanafi S, Mladenović N, 2014 *A two phase variable neighborhood search for the multi-product inventory routing problem. Computers & Operations Research* 52:291–299.

Moin NH, Salhi S, Aziz N, 2011 *An efficient hybrid genetic algorithm for the multi-product multi-period inventory routing problem. International Journal of Production Economics* 133(1):334–343.

Mousavi R, Bashiri M, Nikzad E, 2022 *Stochastic production routing problem for perishable products: Modeling and a solution algorithm. Computers & Operations Research* 142:105725.

Naqvi MA, Amin SH, 2021 *Supplier selection and order allocation: a literature review. Journal of Data, Information and Management* 3(2):125–139.

Onggo BS, Panadero J, Corlu CG, Juan AA, 2019 *Agri-food supply chains with stochastic demands: A multi-period inventory routing problem with perishable products. Simulation Modelling Practice and Theory* 97:101970.

Papageorgiou DJ, Cheon MS, Nemhauser G, Sokol J, 2015 *Approximate dynamic programming for a class of long-horizon maritime inventory routing problems. Transportation Science* 49(4):870–885.

Pazhani S, Ventura JA, Mendoza A, 2016 *A serial inventory system with supplier selection and order quantity allocation considering transportation costs. Applied Mathematical Modelling* 40(1):612–634.

Perfetti JJ, Hernández A, Leibovich J, Balcázar Á, et al., 2013 *Políticas para el desarrollo de la agricultura en Colombia*. URL `https://www.repository.fedesarrollo.org.co/bitstream/handle/11445/61/?sequence=1`.

Powell WB, 2021 *From reinforcement learning to optimal control: A unified framework for sequential decisions. Handbook of Reinforcement Learning and Control*, 29–74 (Springer).

Prajapati D, Chan FT, Daultani Y, Pratap S, 2022 *Sustainable vehicle routing of agro-food grains in the e-commerce industry. International Journal of Production Research* 60(24):7319–7344.

Prins C, 2004 *A simple and effective evolutionary algorithm for the vehicle routing problem. Computers & Operations Research* 31(12):1985–2002.

Riera-Ledesma J, Salazar-González JJ, 2012 *Solving school bus routing using the multiple vehicle traveling purchaser problem: A branch-and-cut approach. Computers & Operations Research* 39(2):391–404.

Rivera AEP, Mes MR, 2017 *Anticipatory freight selection in intermodal long-haul round-trips. Transportation Research Part E: Logistics and Transportation Review* 105:176–194.

Rodolfo Enrique SE, Mendoza Valencia DP, 2017 *Costos logísticos del transporte terrestre de carga en Colombia: estrategias para la generación de valor en la logística del transporte terrestre con plus agregado*. URL `https://repositorio.sena.edu.co/bitstream/11404/4125/7/costos_logist_tmp_ri.pdf`.

Roy A, Gao R, Jia L, Maity S, Kar S, 2020 *A noble genetic algorithm to solve a solid green traveling purchaser problem with uncertain cost parameters. American Journal of Mathematical and Management Sciences* 40(1):17–31.

Spliet R, Dabia S, Van Woensel T, 2018 *The time window assignment vehicle routing problem with time-dependent travel times. Transportation Science* 52(2):261–276.

Spliet R, Gabor AF, 2015 *The time window assignment vehicle routing problem. Transportation Science* 49(4):721–731.

Toriello A, Nemhauser G, Savelsbergh M, 2010 *Decomposing inventory routing problems with approximate value functions. Naval Research Logistics* 57(8):718–727.

Toth P, Vigo D, 2002 *An overview of vehicle routing problems. The Vehicle Routing Problem* 1–26.

Ulmer MW, Goodson JC, Mattfeld DC, Hennig M, 2019 *Offline–online approximate dynamic programming for dynamic vehicle routing with stochastic requests. Transportation Science* 53(1):185–202.

Ulmer MW, Soeffker N, Mattfeld DC, 2018 *Value function approximation for dynamic multi-period vehicle routing. European Journal of Operational Research* 269(3):883–899.

UPRA, 2018 *Identificación general de la frontera agrícola en Colombia*. URL `https://acortar.link/vQaW7M`.

Valev N, 2023 *Retail prices around the world*. URL `https://www.globalproductprices.com`.

van Heeswijk WJA, Mes MRK, Schutten JMJ, 2019 *The delivery dispatching problem with time windows for urban consolidation centers. Transportation Science* 53(1):203–221.

Violi A, Laganá D, Paradiso R, 2020 *The inventory routing problem under uncertainty with perishable products: an application in the agri-food supply chain. Soft Computing* 24(18):13725–13740.

Wen M, Cordeau JF, Laporte G, Larsen J, 2010 *The dynamic multi-period vehicle routing problem. Computers & Operations Research* 37(9):1615–1623.

Yadav S, Singh SP, 2022 *Modelling procurement problems in the environment of blockchain technology. Computers & Industrial Engineering* 172:108546.

Zehtabian S, Ulmer MW, 2023 *Consistent time window assignments for stochastic multi-depot multi-commodity pickup and delivery. Working Paper Series* .

Appendix

In this Appendix, we present a more detailed literature review, details on the algorithm and benchmark policies, and additional results of our experiments.

A.1. Detailed literature review

This section presents work related to each decision component of our problem. First, we present the works related to inventory routing, procurement with inventory management, procurement routing, and dynamic multi-period routing. Then, we present a summary table.

A.1.1. Inventory routing problem. The Inventory Routing Problem (IRP) aims to find optimal inventory policies and vehicle routing programs to reduce supply chain costs. A general review of the IRP is presented in Andersson et al. (2010), Coelho, Cordeau, and Laporte (2014b), and Malladi and Sowlati (2018). As per the flow of products in the supply chain, inventory routing can be classified as inbound routing (replenishment) or outbound routing (delivery), as mentioned in Cobb (2016). Below we discuss the studies given related to these two categories.

Inventory problems with inbound routes have not yet been extensively studied in the literature. The papers found considered deterministic and static versions of the problem. They involve managing the collection of products from suppliers to a distribution center or production plant, the latter being in charge of the collection logistics. Moin, Salhi, and Aziz (2011) and Mjirda et al. (2014) consider an assembly problem where each supplier provides a single type of part. In both cases, the solution approach is approximate optimization. Cheng et al. (2016) proposes an MINLP and a genetic algorithm for a supplier pickup and assembly problem at the silver plant considering carbon emission regulations. In Chitsaz, Cordeau, and Jans (2019), a decomposition matheuristic is developed to solve an assembly, production, and inventory routing problem with inbound transportation. The problem consists of selecting the suppliers to visit, their order, and the inventory level at the supplier and the plant, considering only one product type per supplier. Subsequently, in Chitsaz, Cordeau, and Jans (2020), the same problem is solved by a branch-and-cut (B&C) algorithm. However, in this version, the suppliers offer different products. Finally, in Bertazzi et al. (2020), a material assembly problem is solved where the decisions of visit, inventory management, and quantity to be picked are managed cyclically. A Branch-and-cut algorithm is proposed to solve the problem.

Inventory problems with outbound routes have been widely studied. The Vendor Managed Inventory (VMI) problem is the most common form known in the literature. In the VMI, customers transfer the inventory management responsibility to a vendor, who knows the stock levels of their customers and must plan a distribution scheme to maintain adequate levels for all customers' products. A deterministic inventory routing problem is solved in Toriello, Nemhauser, and Savelsbergh (2010) and Papageorgiou et al. (2015). A Mixed Integer Linear Program (MILP) and a Value Function Approximation (VFA) are proposed to solve the problems, respectively. In Adelman (2004), a stochastic dynamic inventory routing problem (SDIRP) is solved. A VFA is developed based on dual relaxations to anticipate future routing costs. Each customer has a stochastic demand per day. Bertazzi et al. (2013) propose a hybrid Rollout Algorithm (RA) to solve a similar problem with backlogs. The solution of a MILP is used in the RA to anticipate stochastic demand and

potential future decisions. In Coelho, Cordeau, and Laporte (2014a), an IRP with a single vehicle is solved by optimizing a static instance whenever new information becomes available. They use forecasts to generate an approximation of the unknown future demand. Lateral transshipment between customers is allowed. An SDIRP in the context of Bike Sharing Systems is presented in Brinkmann, Ulmer, and Mattfeld (2019). They propose a dynamic lookahead to relocate bikes dynamically during the day. Simulation of future periods is done to anticipate future demands in current inventory decisions. Subsequently, in Brinkmann, Ulmer, and Mattfeld (2020), the same problem is solved but they extend it to a multi-vehicle problem. A lookahead method is proposed that anticipates future developments and coordination of the fleet.

Perishable products are also considered in stochastic inventory routing problems with outbound logistics. A single perishable product SIRP is solved in Crama et al. (2018) and Onggo et al. (2019). Crama et al. (2018) tested five policies to solve the problem: three based on expected values and two based on features of the problem (inventory levels, routing). On the other hand, Onggo et al. (2019) proposes a simheuristic algorithm, which integrates Monte Carlo simulation within an iterated local search to solve the problem. In Violi, Laganá, and Paradiso (2020), a rolling horizon approach based on a multistage stochastic linear program is proposed. They solve an IRP by considering risk measures. Finally, a matheuristic is developed in Mousavi, Bashiri, and Nikzad (2022) for a SIRP considering production decisions.

Our paper extends the work related to inventory routing problems. In contrast to previous work, we do not assume total control over suppliers' offer (customers in the case of outbound logistics), which leads to a more volatile and dynamic environment. Moreover, not having control over inventory levels and the need to contemplate the purchase prices of the products increases the complexity of the problem.

A.1.2. Procurement with inventory management. The purchasing decisions and inventory management are related to the problems of supplier selection and order allocation (SS&OA). We refer to Aouadni, Aouadni, and Rebaï (2019) and Naqvi and Amin (2021) for recent overviews. Generally, the routing decisions are not considered in SS&OA problems, or an approximations function is used to estimate the routing cost. The deterministic version of the problem uses qualitative and quantitative criteria to rank the suppliers, then, using exact and approximate methods, operational decisions (quantities and inventories) are made. Mendoza and Ventura (2008) presented a mathematical model for determining the optimal inventory policy. This model uses a power-of-two (POT) approach for the effectiveness of the inventory system under control. A two-stage multiple criteria dynamic programming approach is proposed in Mafakheri, Breton, and Ghoniem (2011). They consider decisions under time-varying prices/costs, capacity, and demand volumes and profiles. A MINP is proposed in Pazhani, Ventura, and Mendoza (2016) to determine the optimal inventory policy and allocation of orders among the suppliers at a multi-stage supply chain. Different vehicles' size are considered, and the transportation cost is modeled using piecewise function. A procurement problem in a blockchain context is solved in Yadav and Singh (2022). They propose a MILP that incorporates the block development cost while purchasing, ordering, transporting and holding processes.

Different uncertainty sources have been studied in SS&OA problems. In Hammami, Frein, and Hadj-Alouane (2012) they consider a supplier selection problem with uncertain lead time in an international context. They use a MILP model to solve the problem. Uncertainty in the purchase prices related to exchange

rates is considered in Hammami, Temponi, and Frein (2014). A mixed integer scenario-based stochastic programming method is developed to minimize the total system expected cost. Esmaeili-Najafabadi et al. (2019) solve a problem of SS&OA with supplier availability under disruption risk. A mixed-integer nonlinear programming model is proposed. Finally, Hosseini, Flapper, and Pirayesh (2022) solve a problem with uncertain demand and suppliers' availability. They integrated a solution approach based on stochastic programming and dynamic programming.

Different from the problems mentioned above, we do not consider the selection of suppliers for long periods. In the context of agri-food chains, the characteristics of suppliers change rapidly, requiring a flexible policy. Additionally, we contemplate the effect of three sources of uncertainty in decision making which has yet to be done in the SS&OA literature.

A.1.3. Procurement with routing decisions. The Traveling Purchaser Problem (TPP) considers jointly procurement and routing decisions from suppliers. We refer to Manerba, Mansini, and Riera-Ledesma (2017) for a general overview. The TPP can use one or a fleet of vehicles. When more than one vehicle is considered this problem is known as the Multi-Vehicle Traveling Purchaser Problem (MVTPP), only deterministic versions have been solved. Riera-Ledesma and Salazar-González (2012) solved an MVTPP to model a school bus routing problem. They presented a branch-and-cut (B&C) algorithm based on a two-index single-commodity flow formulation. Bianchessi, Mansini, and Speranza (2014) proposed a branch-and-price (B&P) algorithm to solve an MVTPP with length route bounded. The pricing problem is modeled and solved as an elementary path problem with resource constraints (SPPRC). Manerba and Mansini (2015) introduced a variant named MVTPP with pairwise incompatibility constraints (PIC), involving possible incompatibilities among products that forbid loading two incompatible products into the same vehicle. A B&C algorithm based on a three-index formulation is proposed. In Manerba and Mansini (2016), a MVTPP is used to model a nurse routing problem with inter-route incompatibilities constraints and bounds on the duration of the routes. A B&P algorithm is proposed to solve the problem. Finally, a Branch-and-price-and-cut (B&P&C) algorithm is developed in Bianchessi, Irnich, and Tilk (2021) to solve the MVTPP with additional constraints.

Articles that consider uncertainty and dynamism solve the TPP just with one vehicle. Product prices and quantities available from suppliers are the most studied uncertain parameters in the literature (see Kang and Ouyang (2011), Beraldi et al. (2017)). In Roy et al. (2020), uncertain travel times are considered in addition to price and quantities. Regarding dynamic variants of the TPP in Angelelli, Mansini, and Vindigni (2016), they solve the problem by presenting changes in the units available at suppliers, which decrease over time.

Problems that consider purchasing decisions and routing to suppliers usually solve single-period problems. In contrast to our work, we propose a multi-period problem with the interest of looking at the effect of purchasing decisions in future periods and the impact on routing and inventory management decisions.

A.1.4. Multi-period routing under uncertainty. There is also increasing work on multi-period routing under uncertainty (MP-R). We refer to Avraham and Raviv (2021) for a recent overview. In the work on multi-period routing, usually the demand of future days is uncertain. Decisions are made about the period demand should be scheduled, either with the goal of minimizing cost or with the goal of minimizing waiting

Table A1 Literature classification

Category	Author	Decisions			Sources of uncertainty				Anticipation	Multi period
		Purchase	Routing	Inventory	Prices	Supply	Demand	Others		
IRP	Toriello, Nemhauser, and Savelsbergh (2010)		✓	✓					n/a	✓
	Moin, Salhi, and Aziz (2011)		✓	✓					n/a	✓
	Mjirda et al. (2014)		✓	✓					n/a	✓
	Papageorgiou et al. (2015)		✓	✓					n/a	✓
	Cheng et al. (2016)		✓	✓					n/a	✓
	Chitsaz, Cordeau, and Jans (2019)		✓	✓					n/a	✓
	Bertazzi et al. (2020)		✓	✓					n/a	✓
	Chitsaz, Cordeau, and Jans (2020)		✓	✓					n/a	✓
	Adelman (2004)		✓	✓			✓		✓	
	Bertazzi et al. (2013)		✓	✓			✓		✓	✓
	Coelho, Cordeau, and Laporte (2014a)		✓	✓			✓		✓	✓
	Crama et al. (2018)		✓	✓			✓		✓	
	Brinkmann, Ulmer, and Mattfeld (2019)		✓	✓			✓		✓	
	Onggo et al. (2019)		✓	✓			✓		✓	✓
	Brinkmann, Ulmer, and Mattfeld (2020)		✓	✓			✓		✓	
	Violi, Laganá, and Paradiso (2020)		✓	✓			✓		✓	✓
	Mousavi, Bashiri, and Nikzad (2022)		✓	✓			✓		✓	✓
SS&OA	Mendoza and Ventura (2008)	✓		✓					n/a	
	Mafakheri, Breton, and Ghoniem (2011)	✓		✓					n/a	✓
	Pazhani, Ventura, and Mendoza (2016)	✓		✓					n/a	
	Yadav and Singh (2022)	✓		✓					n/a	✓
	Hammami, Frein, and Hadj-Alouane (2012)	✓		✓				✓		✓
	Hammami, Temponi, and Frein (2014)	✓		✓	✓				✓	✓
	Esmaeili-Najafabadi et al. (2019)	✓		✓		✓				
	Hosseini, Flapper, and Pirayesh (2022)	✓		✓		✓	✓		✓	✓
TPP	Riera-Ledesma and Salazar-González (2012)	✓	✓						n/a	
	Bianchessi, Mansini, and Speranza (2014)	✓	✓						n/a	
	Manerba and Mansini (2015)	✓	✓						n/a	
	Manerba and Mansini (2016)	✓	✓						n/a	
	Bianchessi, Irnich, and Tilk (2021)	✓	✓						n/a	
	Kang and Ouyang (2011)	✓	✓		✓					
	Angelelli, Mansini, and Vindigni (2016)	✓	✓			✓				
	Beraldi et al. (2017)	✓	✓		✓	✓			✓	
	Roy et al. (2020)	✓	✓		✓			✓		
MP-R	Angelelli et al. (2009)		✓				✓			✓
	Wen et al. (2010)		✓				✓			✓
	Albareda-Sambola, Fernández, and Laporte (2014)		✓				✓			✓
	Klapp, Erera, and Toriello (2018b)		✓				✓		✓	✓
	van Heeswijk, Mes, and Schutten (2019)						✓		✓	✓
	Rivera and Mes (2017)						✓		✓	✓
	Klapp, Erera, and Toriello (2018a)		✓				✓		✓	✓
	Ulmer, Soeffker, and Mattfeld (2018)		✓				✓		✓	✓
	Avraham and Raviv (2021)		✓				✓	✓	✓	✓
	Laganà, Laporte, and Vocaturo (2021)		✓				✓			✓
	Keskin et al. (2023)		✓				✓	✓		✓
	Çabuk and Erol (2019)	✓	✓	✓					n/a	✓
	Our work	✓	✓	✓	✓	✓	✓		✓	✓

time for the customers. Early work has been presented by Angelelli et al. (2009), Wen et al. (2010), Albareda-Sambola, Fernández, and Laporte (2014). All three papers suggests rule-based approaches to decide whom to serve in this period and whom to serve in the next. Later, such rule-based method were either adapted for more complex problems (Laganà, Laporte, and Vocaturo 2021) or translated in learning algorithms (Ulmer, Soeffker, and Mattfeld 2018, Avraham and Raviv 2021). Some work proposes anticipation via stochastic programs (Klapp, Erera, and Toriello 2018b,a) while others apply reinforcement learning (Rivera and Mes 2017, van Heeswijk, Mes, and Schutten 2019). We note that none of the previous works combines both. Further, most of the work avoid complex combinatorial decision making, either by limiting the problem size or by not modeling routing decisions explicitly.

A.1.5. Summary. Table A1 presents a summary of the relevant literature to this paper. We differentiate work into the following categories: *Decisions*, which means if the problem considers purchase, routing,

or inventory management decisions; *Stochasticity*, whether the work considers uncertainty in at least one problem component; *Anticipation*, whether the proposed method anticipates the impact of a decision on future costs. All deterministic problems have a "n/a"-entry in the Anticipation column, as Anticipation is only applicable in a stochastic problem; and *Multi-period*, whether the decisions have an impact over a time horizon longer than one period.

Algorithm 1 Generate routes

Input: List τ_{ij}, z_t, e_t, Q, l^{max}

Output: x_t : routes details, f_t : number of vehicles dispatched

1: $Tour \leftarrow NearestNeighbourAlgorithm(e_t, \tau_{ij})$
2: **for each node** i **in** $Tour$ **do**
3: $SplitNodes \leftarrow AddNode(i)$
4: **end for**
5: **for each contiguous subset** N' **of nodes in** $Tour$ **do**
6: **if** $AccumulatedPurchase(N', z_t) \leq Q$ **then**
7: **if** $AccumulatedTravelTime(N', \tau_{ij}) \leq l^{max}$ **then**
8: $TravelTime \leftarrow AccumulatedTravelTime(N', \tau_{ij}))$
9: $SplitArcs \leftarrow AddWeightedArc(N'.First, N'.Last, TravelTime)$
10: **end if**
11: **end if**
12: **end for**
13: $G \leftarrow BuildGraph(SplitNodes, SplitArcs)$
14: $x_t, f_t \leftarrow BellmanFordShortestPathAlgorithm(Graph)$
15: **return** x_t, f_t

A.2. Algorithmic details

In this section, additional details are given about our routing heuristic, the adaptive routing cost approximation, and the tuning of our method.

A.2.1. Routing heuristic. Algorithm 1 presents the process for constructing the routes from the stochastic program solution. The nearest neighbor algorithm (NNA) is run to obtain a complete tour based on the selected suppliers, e_t (line 1). Then, an augmented graph is constructed with this complete tour and the quantities to be purchased from each supplier, z_t (lines 2 - 13). Following the split procedure, where through the nodes of the complete tour, the possible vehicle routes are built respecting the vehicle capacities and the maximum travel time. After the construction of the augmented graph, we solve a shortest path problem (line 14) using the *BellmanFord* algorithm, to find the routes that minimize the travel time, and the number of vehicles required. Finally, Algorithm 1 returns the routes and the number of vehicles (line 15).

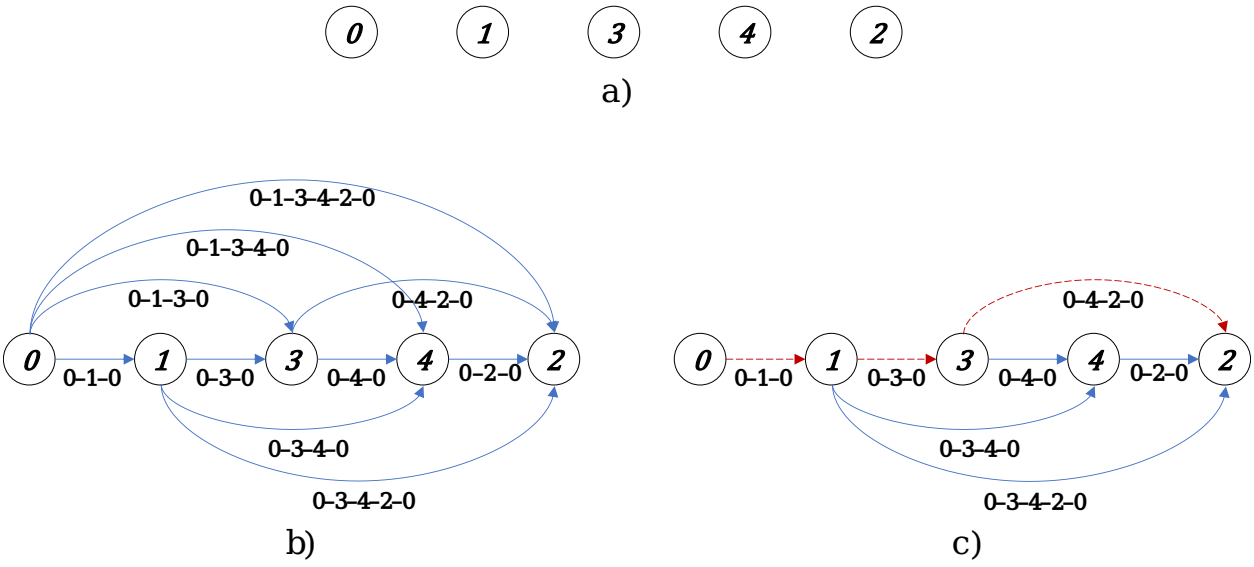

Figure A1 **Example split procedure**

Creating the augmented graph is necessary, as it involves validating all the routing decision constraints (Algorithm 1, lines 2 - 13). The first step in creating the graph is adding the nodes corresponding to the depot and each selected supplier, according to the order in the complete tour (lines 2 - 4). Next, the arcs representing consolidated trips between contiguous nodes of the tour are added; the possible consolidations are identified (lines 5 - 12). These arcs are created if the capacity and maximum travel time constraints are met for the consolidation (lines 6 - 11). Finally, the augmented graph is constructed and returned with the generated nodes and arcs (line 13).

Figure A1 presents an example of the construction of the routes from the stochastic program solution. Figure A1.a shows the complete tour obtained after using the NNA over the selected suppliers. From the order of the complete tour, the possible routes are constructed by considering the subsets of contiguous nodes. Each arc (i, j) in the graph describes the route that starts and ends at the depot and travels between i and j, excluding i. Figure A1.b presents the possible routes to be generated. For computational efficiency, during the creation of arcs, if a route is infeasible, evaluating the subsets of consecutive nodes containing this route will be aborted. In our example, the route 0 - 1 - 3 - 0 is infeasible, then the routes 0 - 1 - 3 - 4 - 0 and 0 - 1 - 3 - 4 - 2 - 0 were not evaluated. Route infeasibility can be caused by exceeding the vehicle capacity or the maximum time limit per route. After having the augmented network constructed with the feasible routes, the shortest path problem is solved from the depot to the last supplier in the sequence to get a detailed understanding of the routes and the quality of the solution. Figure A1.c shows the resulting augmented network and the routes that minimize the total distance to visit the suppliers. One vehicle is assigned for each route.

A.2.2. Adaptive routing cost approximation. The data-driven method for estimating the approximate routing cost of visiting a supplier is presented in the Algorithm 2. First, we initialize the γ-values, and *count*-array in the iteration 0 (lines 1 - 4). We have *count* as a $|M|$-dimensional vector that keeps a

Algorithm 2 Adaptive learning process

Input: List T, M, K, M_k, τ_{ij}, r_k, ϕ_k, Q, l^{max}

Parameters: h: size of the forward horizon, *MIPGap*: minimum quality of the solution returned by the optimization solver, $|\Omega|$: Number of sample of future information, *MaxIter*: maximum number of iterations, n: maximum number of simulations, *Default*: initial γ-values

Output: γ: routing cost approximation tuned

1: **for** $m \in M$ **do**
2: $\gamma^0[m] \leftarrow Default$
3: $count^0[m] \leftarrow 1$
4: **end for**
5: **for** $i \leftarrow 1$ **To** *MaxIter* **do**
6: **for** $j \leftarrow 1$ **To** n **do**
7: $\hat{I}_1 \leftarrow 0$
8: $d_1, p_1, q_1 \leftarrow UncertaintyRevealed(\omega_1)$
9: $S_1 \leftarrow (\hat{I}_1, d_1, p_1, q_1)$
10: **for** $t \in T$ **do**
11: $a_t \leftarrow STARPolicy(t, S_t, T, M, K, M_k, \tau_{ij}, r_k, \phi_k, Q, l^{max}; h, MIPGap, |\Omega|, \gamma^{i-1})$
12: $S_t^a \leftarrow GeneratesPostDecisionState(S_t, a_t)$
13: $S_{t+1} \leftarrow \mathcal{T}(S_t^a, \omega_{t+1})$
14: $routes[j,t] \leftarrow ExtractRoutesInformation(a_t)$
15: **end for**
16: **end for**
17: $\gamma^i, count^i \leftarrow UpdateGammaValues(M, T, routes, count^{i-1}, \tau_{ij}, \gamma^{i-1}; n)$
18: **end for**
19: **return** γ

count of the cumulative number of times each supplier has been visited throughout the iterative process. The algorithm performs a *MaxIter* number of iterations (line 5); in each iteration i, a n number of simulations is performed (line 6). In each simulation j, the S_1 state information is initialized (lines 7 - 9), and then the decision sequence is run over the entire planning horizon T (line 10). In each period $t \in T$, the **STAR**-approach is executed with the γ^{i-1}-values fixed, and the action a_t is obtained (line 11). Having the action a_t, the post-decision state S_t^a is generated (line 12), and using the transition function, the state S_{t+1} is calculated (line 13). The routing information generated in each period $t \in T$ for each simulation j is stored in the *routes*-array (line 14). At the end of the simulations, with the information on the generated routes and the current γ^{i-1}-values, the new values of the approximate routing cost to be used in the next iteration are estimated γ^i-values; it also updates the $count^i$-array with overall count of supplier visits. (lines 17). This

procedure is performed until the number of iterations has been completed. Algorithm 2 returns the final γ-values (line 17).

Algorithm 3 Update gamma values

Input: List M, T, *routes*, *count*, τ_{ij}, γ

Parameters: n: maximum number of simulation in the set of training instances

Output: *new*γ: new estimates for γ-values, *count*: global counter of visits to suppliers.

1: **for** $m \in M$ **do**
2: $supplierInRouteFlag \leftarrow False$
3: **List** $\bar{\gamma} \leftarrow \emptyset$
4: **for** $j \leftarrow 1$ **To** n **do**
5: **List** $\hat{\gamma} \leftarrow \gamma[m]$
6: **for** $t \in T$ **do**
7: **for each** *route* **in** $routes[j,t]$ **do**
8: **if** m **in** *route* **then**
9: $supplierInRouteFlag \leftarrow True$
10: $avgTime \leftarrow \frac{route.time}{route.number\_suppliers}$
11: $\hat{\gamma} \leftarrow AddGamma(\frac{avgTime}{\tau_{0m}+\tau_{m0}})$
12: **end if**
13: **end for**
14: **end for**
15: $\bar{\gamma} \leftarrow AddGamma(Average(\hat{\gamma}))$
16: **end for**
17: $new\gamma[m] \leftarrow (1 - \frac{1}{\sqrt{count[m]}}) \cdot \gamma[m] + (\frac{1}{\sqrt{count[m]}} \cdot Average(\bar{\gamma}))$
18: **if** $supplierInRouteFlag = True$ **then**
19: $count[m] \leftarrow count[m] + 1$
20: **end if**
21: **end for**
22: **return** *new*γ, *count*

Algorithm 3 presents the process for updating the γ-values. The process is performed for each supplier $m \in M$ (line 1). *supplierInRouteFlag* is a flag that indicates if the supplier m has been visited at least one period of any simulation, it is initialized to *False* (line 2). We use the $\bar{\gamma}$-list to storage the estimations of each simulation j, it is initialized as empty (line 3). The route information generated throughout the n simulations

in the different periods is collected (lines 4 - 16). The $\hat{\gamma}$-list stores the estimated γ-values along the decision sequence of a horizon T; the first value in the list is the current $\gamma[m]$-value (line 5). To update the γ-values, in each period $t \in T$ (line 6), it is checked whether the supplier m is on any route in the *routes*-array (lines 7 and 8). If true, the *supplierInRouteFlag* flag changes its value (line 9), and the routing cost extraction is done (lines 10 - 11). First, the average time to visit a supplier *avgTime* is calculated. This value is obtained by dividing the time of the route where supplier m is part by the number of suppliers composing that route (line 10). Then, the proportion between the average time and the direct shipping cost is calculated; this information is added to the $\hat{\gamma}$-list (line 11). At the end of the periods of horizon T, the values stored in the $\hat{\gamma}$-list are averaged; this value corresponds to the average γ-values estimated for the simulation j; the new value is stored in the $\bar{\gamma}$-list (line 15). The process is repeated for each simulation j. At the end of the simulations, with the current γ-value and the the average of the values in the $\bar{\gamma}$-list representing the final information extracted from the simulations for supplier m, the value for the next iteration is updated *new*$\gamma[m]$ (line 17). If the supplier m has been visited during any simulation period, its global visit counter *count* is updated (lines 18 - 20). When the new γ-values for all suppliers have been estimated, the new values are returned along with global count (line 22). This algorithm has been elaborated in detail to represent a learning function, line 5 attempts to stabilise past knowledge as an inertia concept, which together with line 17, becomes reinforcement learning.

A.2.3. Tuning global approach parameters. In this section, we present the impact in the stochastic lookahead algorithm performance of tuning the gap for the optimization solver (*MIPGap*), the number of sample paths per uncertainty source ($|\Omega|$) and the lookahead horizon (h). Experiments were run using 20 independent instances realizations.

First, we analyzed the impact of *MIPGap* and $|\Omega|$ values on the performance of solutions. We used $MIPGap = \{1\%, 5\%, 10\%\}$ and $|\Omega| = \{3, 5, 10, 15, 20\}$. We fixed $h = 3$ in the experiments. As the accuracy of the optimizer's stopping criterion (*MIPGap* with small values) and the number of scenarios per source of uncertainty ($|\Omega|$) is increased, the value of the objective function increases (see Figure A2). This same pattern is repeated for the computation time. The higher the accuracy and the greater the number of scenarios, the longer the computation time required. (see Figure A3).

Since we are looking for a methodology that finds quick solutions and is flexible to possible changes in the sources of uncertainty not contemplated in the scenarios, we set $MIPGap = 5\%$ and $|\Omega| = 10$. These values were selected since with $MIPGap = 5\%$ the difference in solution quality compared to $MIPGap = 1\%$ and computation time compared to $MIPGap = 10\%$ is not significantly increased. In addition, increasing the number of scenarios to more than 10 does not improve the solutions with a $MIPGap = 5\%$, significantly.

With *MIPGap* and $|\Omega|$ parameters tuned, we analyzed the impact of the changing size of the forward horizon h. We used $h = \{1, 2, \ldots 6, 7\}$. Figure A4 shows performance for values used. We set $h = 3$ as by increasing or decreasing its value, the difference between the solution quality is no more than 1%.

A.3. Benchmark policy details

In this section, we present details on the tuning of the benchmark policies with global γ-parameters as well as the PFA and the alternative approaches for approximation of the γ-values.

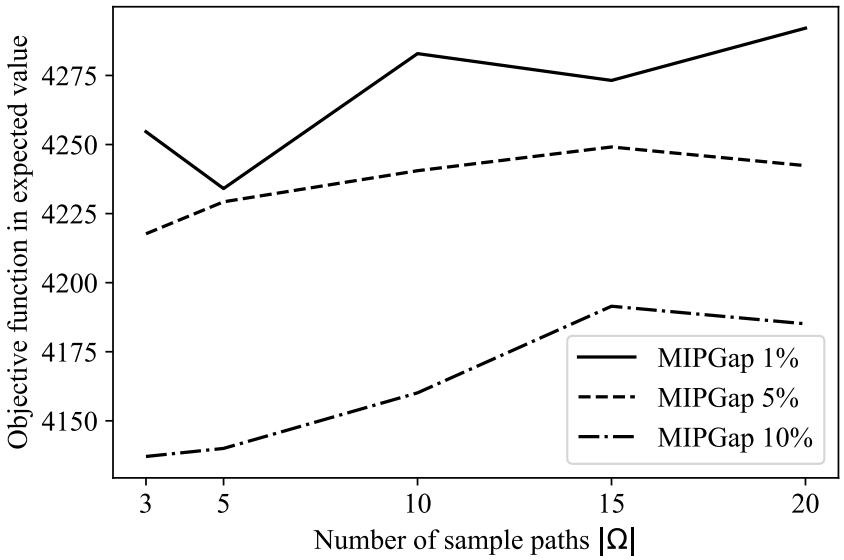

Figure A2 **Objective function for each pair of MIPGap values and number of sample paths ($|\Omega|$)**

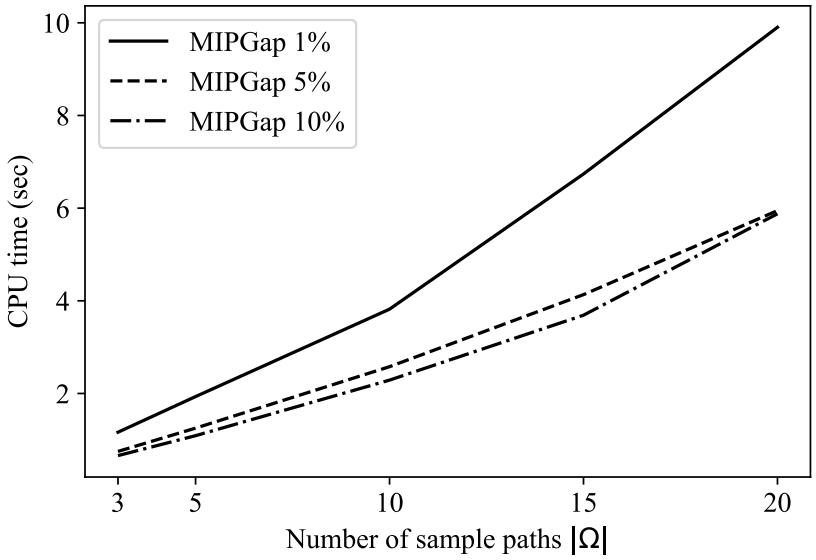

Figure A3 **CPU time for each pair of MIPGap values and number of sample paths ($|\Omega|$)**

A.3.1. Tuning routing approximation via enumeration. In this section, we present the results of the global γ-values tuning for **MYOPIC**, **EV** and **ST-ONE** policies via enumeration. For these policies, we plot the performance for $\gamma_m = \{0.0, 0.1, \ldots, 1.0\}, \forall m \in M$ in Figure A5. We generated 20 independent instance realizations and evaluated each γ-value for the entire set of instances.

The behavior for all three policies is similar. Initially, with increasing γ, the objective values increase as well. They all reach a peak at $\gamma = 0.6$ and then start declining. This confirms that both no routing cost

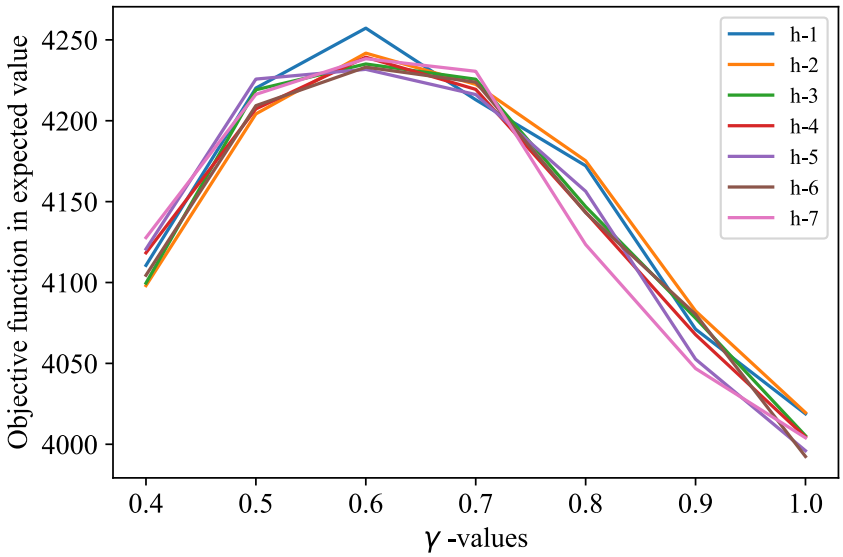

Figure A4 **Objective function as a function of different forward horizon values (h)**

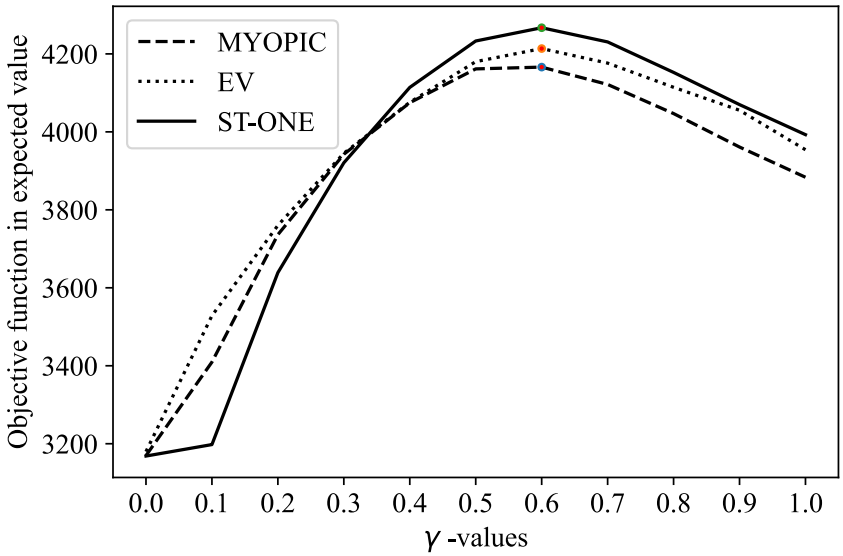

Figure A5 **Policy performance as a function of γ-values**

consideration as well as assumption of direct trips only lead to inferior performance. Particularly notable is the the very poor performance for $\gamma = 0.0$. That means, a clear decomposition, purchasing first, route second without routing cost approximation, is insufficient for the problem at hand.

A.3.2. Policy function approximation (PFA). In this section, we present the policy function approximation algorithm (PFA) details. First, we explain how the policy builds a solution. Later, we show how the θ-parameter is tuned.

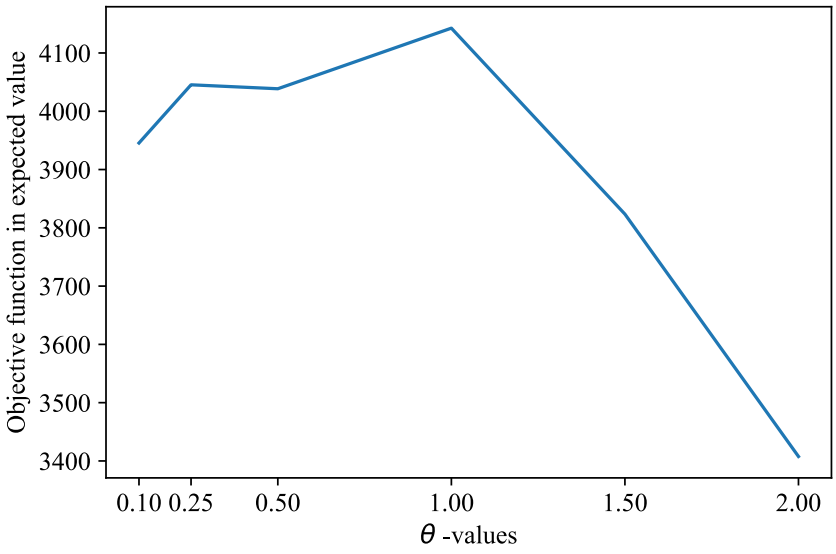

Figure A6 Tuning of the policy function approximation (PFA)

The **PFA** seeks to keep units in inventory by purchasing an additional θ-percentage over known demand if purchase prices are low than the expected value. **PFA** is equivalent to running stochastic program presented on Section 4.2 having a forward period horizon $h=1$, and $|\Omega|=1$. However, information for future periods is not generated by scenarios. With $h=1$ we have a set $T'=\{t,t+1\}$. The information for decision making in period t is related to the state $S_t=(\hat{I}_t,d_t,p_t,q_t)$, and the information for period $t+1$ is constructed as follows: $d_{t+1}=d_t\cdot\theta$, $p_{t+1}=\mu_p$, and $q_{t+1}=q_t$. Given a two-period problem, if the purchase prices in period t are lower than the expected value of the purchase prices in $t+1$, the optimization model will try to satisfy the demand in period $t+1$ by purchasing in period t as long as the suppliers' capacity allows it. The **PFA** used the same γ-values as **MYOPIC**-policy.

Different values of $\theta=\{0.1,0.25,0.5,1,1.5,1.5,2\}$ have been tested. We generated 20 independent instance realizations. Figure A6 shows the results using the same θ-value for the entire set of instances. When the θ-value increases, the profit on the solutions also increases; however, when the θ-value is greater than 1, the profit decreases. These results could be related to the loss due perishability of the products. For the benchmark comparison, $\theta=1$ is set.

A.3.3. Method-oriented policies approximation. In this section, we present the details to approximate the γ-values based on travel time radios, available quantities, and purchase prices at the suppliers. To calculate these approximations, an indicator function $\mathbb{I}(\cdot)$ is used, which takes the value of one if the condition is satisfied. This indicator function generates a score for each supplier to then calculate the γ-values. First, we present the procedure for the **ST-DIST**-policy, then the policies **ST-CAPA**, **ST-PRICE**, and **ST-DCP** are described.

Eq. A1 presents the way to estimate the γ-values for **ST-DIST**-policy. Estimates are made with the number of suppliers that can be visited within a maximum δ^τ travel time radio. We tested $\delta^\tau=\{15,30,45,60,75,90,105,120,135,160\}$ as maximum visiting travel time for each supplier.

$$\gamma_m^{\tau} = \frac{1}{1 + \sum_{i \in M} \mathbb{I}(\tau_{mi} \leq \delta^{\tau})}, \forall m \in M | m \neq i \tag{A1}$$

Experiments were run using 20 independent instances realizations. Figure A6 shows the results using the same δ^{τ}-value for all suppliers. As the time limit to visit a supplier increases, the quality performance of the solution decreases. The best results are obtained with small time limits, which makes sense since the aim is to generate consolidations among close suppliers. For the benchmark comparison, $\delta^{\tau} = 30$ is set.

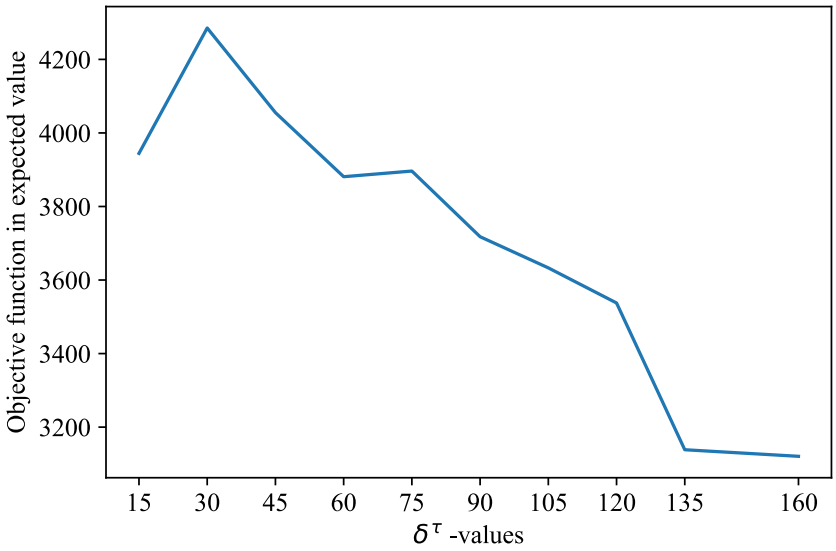

Figure A7 **Tuning δ^{τ} for ST-DIST-policy**

Eq. A2, and Eq. A3 present the calculations of the γ-values for policies **ST-CAPA** and **ST-PRICE**, respectively. The calculations relate the information of the available quantities and purchase price of each product at each supplier with the average information of all suppliers. The average value for the offer and the purchase price are calculated as follows: $\bar{\mu}_{q_k} = \frac{\sum_{m \in M_k} \mu_{q_{mk}}}{|M_k|}, \forall k \in K$, $\bar{\mu}_{p_k} = \frac{\sum_{m \in M_k} \mu_{p_{mk}}}{|M_k|}, \forall k \in K$, respectively. These values are taken as a reference to calculate the score for each supplier. The number of products whose expected value is above (offer) or below (purchase price) the general average value of the suppliers is counted. This count is then used to calculate the γ-values.

The last method is called **ST-DCP**, presented in Eq. A4. The **ST-DCP** method relates the information of the previous estimates, and the γ-values are obtained by calculating the average of the estimations.

$$\gamma_m^{\mu_q} = \frac{1}{1 + \sum_{k \in K_m} \mathbb{I}(\mu_{q_{mk}} \geq \bar{\mu}_{q_k})}, \forall m \in M \tag{A2}$$

$$\gamma_m^{\mu_p} = \frac{1}{1 + \sum_{k \in K_m} \mathbb{I}(\mu_{p_{mk}} \leq \bar{\mu}_{p_k})}, \forall m \in M \tag{A3}$$

$$\gamma_m^{avg} = \frac{\gamma_m^{\tau} + \gamma_m^{\mu_q} + \gamma_m^{\mu_p}}{3}, \forall m \in M \tag{A4}$$

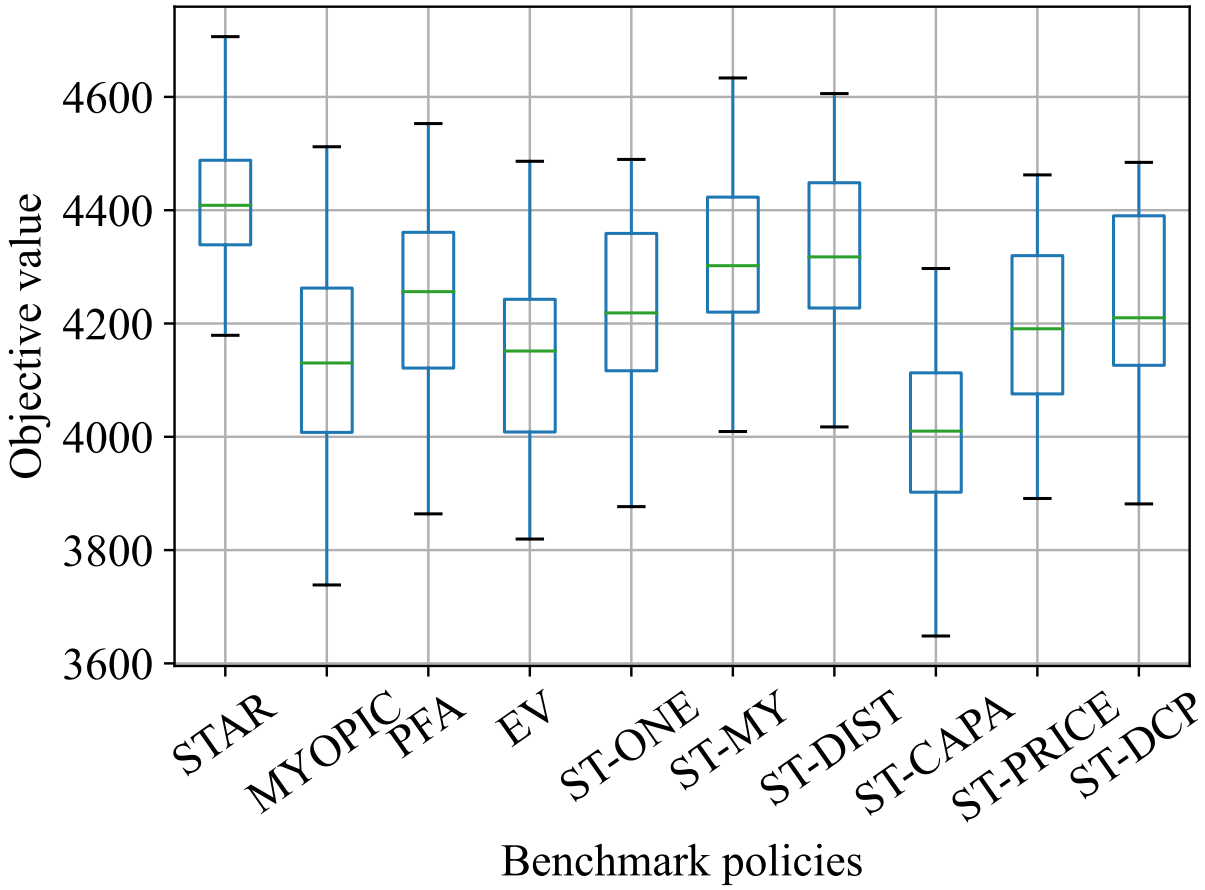

Figure A8 **Analysis of variability in performance of benchmark policies**

A.4. Additional results of the experiments

In this section, we present additional results of our computational study related to variability in the solutions and the relationship of the γ-values with other characteristics of the problem.

A.4.1. Variance in the policies' performance. In this section, we analyzed the variability behavior in the quality of the benchmark policies' solutions. The results are presented in Figure A8. We observe a positive bias (higher values) in the behavior of the **STAR**-policy. In addition, there is less dispersion in the solution's quality close to the mean compared to the other policies. Furthermore, the mean value of the **STAR**-policy is, in most cases, above the values that compose the third quartile of the benchmark policies, showing a better performance in the quality of the solutions. To sum up, the **STAR**-policy achieves the highest objective values; it does so with a comparably smaller variance.

A.4.2. Analysis of γ-values in relation to purchase prices, location and supply. In the following, we analyze the relationship between purchase prices, location, and offer features to the γ-values obtained for each supplier. Figure A9 illustrates the behavior of each feature in a heatmap. The x-axis displays the suppliers arranged according to their IDs, ordered from the lowest to the highest γ-values obtained after applying the adaptive learning method. The y-axis denotes the name of each feature. The heatmap represents

the normalized value of each feature-supplier pair. To obtain the normalized values, we first calculated the average value of each feature for each supplier. Then, we divided each value by the maximum value for all suppliers. A dark box indicates a high value compared to the other suppliers.

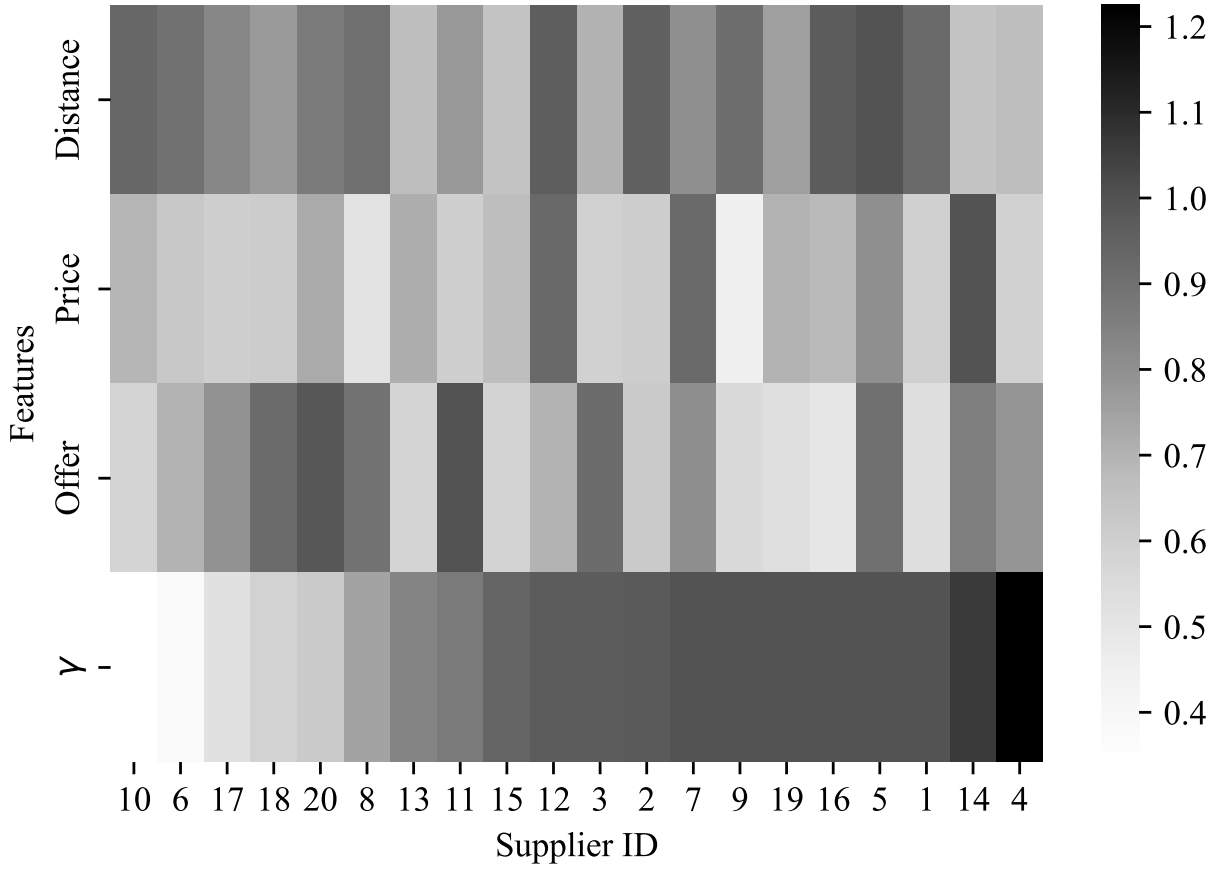

Figure A9 Heatmap distribution of γ-values characteristics

Several patterns can be observed from the γ-values. Suppliers 10, 6, 17, 18, 20, and 8 exhibits the lowest γ-values, which can be attributed to their high average distance. However, they compensate for this by having lower average prices and capacity greater than or equal to the average supply of all suppliers. In contrast, suppliers 4 and 14 have the highest γ-values despite having the smallest average distance. It is because their proximity to the warehouse makes it challenging to consolidate with other suppliers. For the remaining suppliers, their γ-values are either very close to each other or equal to 1. It suggests that direct cost provides a reasonable estimate, given the characteristics of the other variables. While the final γ-values display a certain structure, the complex interplay between purchasing and routing decisions makes it challenging to explain the values clearly or set them a priori without learning.

Otto von Guericke University Magdeburg
Faculty of Economics and Management
P.O. Box 4120 | 39016 Magdeburg | Germany

Tel.: +49 (0) 3 91 / 67-1 85 84
Fax: +49 (0) 3 91 / 67-1 21 20

www.fww.ovgu.de/femm

ISSN 1615-4274