Research Article

Journal of Industrial Engineering, Systems,and Manufacturing Technologies

Integrated Co-optimization of Lot Sizing and Scheduling in the Agri-Food Industry: a Data-Driven Hybrid VNS–ALNS Approach

Ridha Erromdhani[1*] & Abdelwaheb Rebaï[2]

[1]*MODILS Lab, Faculty of Economics and Management, University of Sfax, Sfax, Tunisia*
[2]*MODILS Lab, Faculty of Economics and Management, University of Sfax, Sfax, Tunisia*

[*]Corresponding author: Ridha Erromdhani, MODILS Lab, Faculty of Economics and Management,University of Sfax, Sfax, Tunisia.

Submitted: 14 January 2026 **Accepted:** 21 January 2026 **Published:** 28 January 2026

Citation: *Erromdhani, R., & Rebaï, A. (2026). Integrated Co-optimization of Lot Sizing and Scheduling in the Agri-Food Industry: a Data-Driven Hybrid VNS–ALNS Approach. J of Indu Eng Sym & Manu Tech, 2(1), 01-11.*

Abstract
This paper addresses the problem of integrated co-optimization of lot-sizing and scheduling in multi-level agro-industrial production chains, characterized by product perishability, sequence-dependent setup times (SDST), and significant demand uncertainty. We propose a hybrid metaheuristic combining Variable Neighborhood Search (VNS) and Adaptive Large Neighborhood Search (ALNS), enhanced with a data-driven guidance mechanism based on machine learning. This mechanism aims to predict the potential impact of neighborhood moves in order to filter out unpromising explorations and accelerate convergence. The performance of the approach is evaluated on a set of realistic instances inspired by the agro-industrial sector, including 20 to 100 products, multi-level structures, and up to 70 stochastic demand scenarios. Experimental results demonstrate significant reductions in total cost (up to 20%), notable improvements in solution robustness (measured through CVaR and worst-case scenario cost), as well as increased stability of production plans compared to classical and unguided approaches. These findings highlight the benefits of integrating data-driven mechanisms into hybrid metaheuristics for production planning under uncertainty in the agro-industrial context, effectively balancing solution quality, decision robustness, and computational efficiency.

Keywords: Integrated lot-sizing and Scheduling, Agri-Food Production Systems, Hybrid VNS/ALNS Metaheuristics, Data-Driven Optimization, Machine Learning–Guided Search, Perishable Products, Sequence-Dependent Setup Times, Uncertainty and Robustness.

Introduction

The agro-industrial sector faces particularly complex production planning challenges, arising from the combination of strict technological, economic, and regulatory constraints. Products are often highly perishable and subject to stringent quality, food safety, and traceability requirements, while processing operations involve limited capacities and significant setup times, frequently dependent on the production sequence. In addition to these structural constraints, there is substantial demand uncertainty, influenced by seasonal, promotional, and behavioral factors, which further complicates operational decision-making.

In this context, agri-food companies must simultaneously determine the optimal production quantities (lot-sizing) and organize the detailed sequencing of operations (scheduling), while respecting capacity constraints, product shelf-life windows, and sequence-dependent setup times. However, in industrial practice, as well as in a significant part of the literature, these decisions are often addressed sequentially or independently. Such a separation can lead to major inconsistencies, including excessive inventory levels, an increased number of changeovers, greater instability of production schedules, and significant losses due to product expiration [1, 2].

Numerous studies have shown that the integrated co-optimization of lot-sizing and scheduling better captures the structural interdependencies between these decisions. The quantities produced directly influence the structure of production sequences, while the order of operations affects setup costs, the effective utilization of capacities, and the risks of non-compliance or quality issues [3, 4]. This interaction is particularly critical in the agri-food industry, where cleaning requirements, asymmetric transition times, and limited product shelf-life amplify the consequences of poorly coordinated planning decisions [5].

These challenges are further amplified in multi-level production chains, where the flows of raw materials, semi-finished products, and finished goods are highly interdependent. Insufficient coordination between lot-sizing and scheduling decisions can lead to supply shortages, costly overstocking, or significant waste due to product expiration. Moreover, simultaneously accounting for perishability, sequence-dependent setup times, and demand uncertainty is essential to ensure operational feasibility and the robustness of production plans in real-world environments [6, 7].

From a methodological perspective, the integrated resolution of Lot-Sizing and Scheduling (LSS) problems is recognized as NP-hard, especially when considering multi-level structures, realistic planning horizons, and sources of uncertainty. Exact models based on mixed-integer linear programming (MILP) allow for a detailed representation of these problems but quickly become intractable for large-scale industrial instances. In contrast, sequential heuristic approaches, although more tractable, often overlook key interactions between decisions and lead to suboptimal solutions that are not robust against operational uncertainties.

Given these limitations, recent literature highlights the growing interest in integrated and robust production planning approaches under uncertainty [8-10]. Nevertheless, solving integrated, multi-level LSS models with sequence-dependent setup times and perishable products generally exceeds the capabilities of exact methods, justifying the use of advanced metaheuristics. Approaches such as Variable Neighborhood Search (VNS) and Adaptive Large Neighborhood Search (ALNS) have proven effective for tackling complex combinatorial problems, providing high-quality solutions within reasonable computation times [11, 12]. However, their performance remains highly dependent on the choice of operators and their ability to exploit the specific structure of the instances being solved.

Simultaneously, the increasing digitalization of agri-food production systems—through ERP, MES, and traceability systems—generates a large volume of historical data on demand, machine performance, and setup times. Leveraging these data opens new opportunities for designing data-driven optimization methods capable of guiding heuristic search, reducing the solution space, and enhancing decision robustness [13, 14].

In this work, we propose an integrated, data-driven approach for the co-optimization of lot-sizing and scheduling in multi-level agri-food production systems under uncertainty. We develop a hybrid metaheuristic combining VNS and ALNS, enhanced with machine learning mechanisms designed to predict the quality of neighborhood moves, filter unpromising explorations, and accelerate convergence. The method's performance is evaluated on a set of realistic instances inspired by the agri-food sector, demonstrating significant gains in total cost, solution robustness, and schedule stability compared to classical approaches.

The main contributions of this paper are as follows:

- An integrated multi-level lot-sizing and scheduling model, specifically designed for the agri-food industry, incorporating perishability, limited capacities, sequence-dependent setup times, and demand uncertainty.
- A hybrid VNS/ALNS metaheuristic capable of efficiently solving large-scale industrial instances by combining structural diversification with adaptive intensification.
- A data-driven guidance mechanism based on machine learning, allowing dynamic selection of neighborhood operators and reducing the search space.
- A comprehensive computational validation, highlighting significant improvements in cost, robustness, and schedule stability.

The remainder of the paper is organized as follows. Section 2 presents a literature review. Section 3 details the proposed modeling framework. Section 4 describes the hybrid metaheuristic guided by data. Section 5 analyzes the experimental results, and Section 6 concludes the paper and outlines future research directions.

Literature Review

This section reviews the main studies related to integrated lot-sizing and scheduling problems, sequence-dependent setup times, production planning under uncertainty in the agri-food context, as well as hybrid metaheuristics and data-driven approaches. The objective is to highlight the existing advances while identifying the methodological gaps that this work aims to address.

Integrated Lot-Sizing–Scheduling (LS–S) Approaches

The integration of lot sizing and scheduling is a central topic in production optimization. Early studies have shown that the sequential separation of these decisions leads to significant inefficiencies, particularly in the presence of limited capacities and substantial setup costs [15]. These inefficiencies manifest themselves in excessive inventory levels, suboptimal resource utilization, and increased schedule instability.

Integrated Lot-Sizing and Scheduling (LSS) models were initially developed for single-machine environments and were subsequently extended to multi-product and multi-machine settings [16, 17]. More recent studies have introduced multi-level structures, which are better suited to complex industrial production systems [18]. These models are predominantly based on exact Mixed-Integer Linear Programming (MILP) formulations, allowing for a detailed representation of the interactions between production quantities and scheduling sequences.

However, several major literature surveys point out that these integrated formulations rapidly become intractable as the problem size increases or when realistic constraints are incorporated, such as long planning horizons or multi-level structures. This limitation is particularly pronounced in the agri-food context, where production decisions are highly interdependent and sensitive to operational constraints.

Sequence-Dependent Setup Times (SDST)

The consideration of sequence-dependent setup times (SDST) significantly increases the complexity of Lot-Sizing and Scheduling (LSS) problems. These setup times, associated with cleaning operations, tool changes, or production line reconfigurations, are often asymmetric and variable, particularly in the agri-food industry. Seminal works by Fleischmann and Meyr and Sahling demonstrated that the explicit integration of SDST fundamentally alters the structure of optimal solutions and makes exact formulations difficult to exploit beyond small-sized instances [19, 20]. Formulations based on sequencing or positional variables typically lead to a combinatorial explosion in the number of variables and constraints.

Recent surveys confirm that, despite methodological advances, the exact solution of integrated LSS problems with SDST remains computationally intractable for realistic industrial environments. As a result, most contributions rely on heuristic or metaheuristic approaches, often at the cost of significant simplifications of the problem structure.

Optimization Under Uncertainty and Agri-Food Specificities

Agri-food production systems are inherently exposed to high levels of uncertainty, stemming from demand variability, supply disruptions, and operational disturbances. To address these challenges, a variety of modeling paradigms have been proposed, including stochastic programming, robust optimization, and distribution-free approaches.

Despite these advances, the explicit integration of uncertainty within multi-level LSS models incorporating sequence-dependent setup times and product perishability remains limited. A significant portion of the literature accounts for uncertainty solely at the lot-sizing level, neglecting its implications for detailed scheduling decisions, or relies on simplified representations of perishability. Such methodological decoupling restricts the ability of existing models to produce production plans that are both robust and operationally feasible. In particular, the complex interactions between demand uncertainty, sequencing decisions, and perishability-related losses are still largely underexplored, especially in multi-level agri-food production environments.

Metaheuristics and VNS–ALNS Hybridization

Given the computational intractability of exact formulations, metaheuristic approaches have become the predominant tools for solving large-scale Lot-Sizing and Scheduling (LSS) problems. Variable Neighborhood Search (VNS) is well known for its ability to systematically explore neighborhoods of increasing complexity, thereby ensuring effective diversification of the search process [21]. In contrast, Adaptive Large Neighborhood Search (ALNS) relies on adaptive destroy-and-repair operators to intensify the search in promising regions of the solution space, making it particularly well suited for complex combinatorial problems [22].

Although both approaches have been successfully applied to a wide range of production planning and scheduling problems, their joint use within an explicitly integrated LSS framework remains relatively limited. Most existing studies either focus on sequential heuristics or employ a single metaheuristic paradigm, without fully exploiting the complementarity between diversification and intensification. This limitation is especially pronounced in multi-level production systems operating under uncertainty, where the search landscape is highly rugged and strongly constrained.

Data-Driven Approaches and Research Gaps

The increasing digitalization of industrial production systems has recently fostered the development of data-driven optimization approaches that exploit historical data related to demand, production processes, and resource performance. Recent studies have shown that the integration of predictive models and adaptive learning mechanisms can effectively guide metaheuristic search, reduce the exploration space, and enhance solution robustness and convergence speed.

Despite these advances, the application of data-driven techniques to integrated lot-sizing and scheduling problems in multi-level agri-food production systems remains very limited. In particular, existing studies rarely address simultaneously: (i) product perishability, (ii) sequence-dependent setup times, (iii) demand uncertainty, and (iv) data-driven guidance embedded within a hybrid metaheuristic framework. Most contributions focus on isolated aspects of these challenges or apply learning-based techniques to simplified problem settings, thereby limiting their practical relevance for real-world agri-food production environments. This lack of integrated, data-driven solution approaches highlights a clear methodological gap, which the present work aims to address by combining hybrid VNS–ALNS metaheuristics with learning-based guidance in a comprehensive and realistic production planning framework.

Positioning and Contribution of the Present Work

The literature review highlights several important research gaps. First, integrated LSS models that simultaneously account for product perishability, sequence-dependent setup times, multi-level production structures, and demand uncertainty remain scarce. Second, existing approaches are either computationally intractable when realistic industrial constraints are considered or rely on unguided heuristic methods, which often exhibit limited robustness and poor schedule stability. Finally, the explicit hybridization of complementary metaheuristics enriched with data-driven guidance mechanisms has been largely underexplored in realistic agri-food production contexts.

Against this background, the present work proposes an integrated, hybrid, and data-driven approach for the co-optimization of lot sizing and scheduling in multi-level agri-food production systems under uncertainty. The proposed framework combines the strengths of VNS and ALNS metaheuristics and embeds learning-based mechanisms to guide the search process, enhance robustness, and ensure operationally stable and high-quality production plans.

Advanced Problem Modeling

We consider a multi-level agri-food production system composed of raw materials, intermediate products, and finished goods, where each level is interdependent and subject to capacity, storage, and perishability constraints. The objective is to simultaneously determine production quantities (lot sizing) and the sequencing of operations (scheduling), while minimizing the overall total cost and ensuring operational feasibility. The main challenges specific to the agri-food context include:

- Product perishability: Each product has a limited shelf life, which imposes constraints on inventory holding and generates waste-related costs.
- Sequence-dependent setup times (SDST): Transitions between products require setup times that vary according to

the production sequence, reflecting cleaning operations, tool changes, or product changeovers.

- Demand variability and uncertainty: Demand is modeled using a set of stochastic scenarios to capture the fluctuations commonly observed in real industrial environments.
- Multi-level production structures: Material flows from raw materials to intermediate and finished products must remain consistent across levels to ensure overall system feasibility.

Sets, Indices, and Parameters

Sets and Indices

- P: Set of products (including multi-level products).
- M : Set of resources/machines.
- T={1,…,T} : Discrete planning horizon (periods).
- S : Set of demand scenarios (stochastic) with associated probability πs.
- $A_p=\{0,\ldots, A_p-1\}$: Shelf life (in periods) of product p modeled using age-based compartments.

Paramètres

- $d^s_{p,t}$: Demand for product p in period t under scenario s.
- c^{prod}_p : Unit production cost of product p.
- c^{inv}_p : Unit holding cost per period for product p.
- $c^{setup}_{i,j,m}$: Transition (setup) cost for switching from product i to product j on machine m.
- $\tau_{i,j,m}$ Setup time for switching from product i to product j on machine m.
- $r_{p,m}$: Unit production time (or processing rate) of product p on machine m.
- $C_{m,t}$: Available capacity (time) of resource/machine m in period t.
- h_p : Shelf life (horizon of perishability, in periods) of product p.
- c^{waste}_p : Waste cost per unit of expired product p.
- c^{back}_p : Backorder cost or late-penalty per unit of product p.
- α : Service level parameter (optional, for service-level constraints).

Decision Variables

Decision variables may be indexed by scenario s (second-stage variables) to account for demand uncertainty.

Production Quantities and Inventories

- $q^s_{p,t} \geq 0$: Quantity of product p produced in period t under scenario s.
- $I^s_{p,t} \geq 0$: Inventory level of product p at the end of period t under scenario s.
- $I^s_{p,t,a}$: Inventory of product p in period t with age a (age-bucket representation), under scenario s.

Production Activation / Setup Decisions

- $y^s_{p,t} \in \{0,1\}$: Binary variable equal to 1 if product p is produced in period t (on a given machine or on at least one machine), and 0 otherwise.

Sequencing Variables (SDST)

To model sequence-dependent setup times, successor variables are introduced for each machine m and period t:

- $u^s_{i,j,m,t} \in \{0,1\}$: Binary variable equal to 1 if, in period t on machine m, product j is scheduled immediately after product i (i.e., i precedes j in the production sequence), and 0 otherwise.
- $v^s_{p,m,t} \in \{0,1\}$: Binary variable equal to 1 if product p is the first product processed (or is produced) in the sequence of period t on machine m, and 0 otherwise.

Alternative, more compact formulations based on position or permutation variables may also be considered; the choice depends on the required level of detail and the size of the model.

Penalty Variables

- $b^s_{p,t} \geq 0$: Backorder quantity of product p in period t under scenario s.
- $w^s_{p,t} \geq 0$: Quantity of product p wasted (expired) in period t under scenario s.

Objective Function

The objective is to minimize the expected total cost over all demand scenarios:

$$Min\ \mathrm{E}[Z] = \sum_{s \in S} \pi_s \left[\sum_{t \in T} \sum_{p \in P} \left(c^{prod}_p q^s_{p,t} + c^{inv}_p I^s_{p,t} + c^{back}_p b^s_{p,t} + c^{wast}_p w^s_{p,t} \right) + \sum_{t \in T} \sum_{m \in M} \sum_{i \in P} \sum_{j \in P} c^{setup}_{i,j,m} u^s_{i,j,m,t} \right]$$

where the expectation is taken over the stochastic demand scenarios s∈S, weighted by their probabilities πs.

Possible extensions include the incorporation of probabilistic service-level constraints or risk-averse criteria such as Conditional Value-at-Risk (CVaR) to explicitly capture decision-makers' risk aversion under uncertainty.

- The setup cost term can be replaced or complemented by the consumption of setup times within the capacity $C_{m,t}$ constraints (as detailed below).
- Fixed start-up costs may also be added if required (denoted, for example, by).

Main Constraints

Multi-Level Flow Balance

For each product p∈P, period t∈T, and scenario s∈S:

$$I^s_{p,t-1} + q^s_{p,t} + \sum_{\ell \in \boldsymbol{p}\ (p)} \beta_{\ell \to p} q^s_{\ell,t} - lag_{\ell \to p} = d^s_{p,t} + I^s_{p,t} + b^s_{p,t} + w^s_{p,t}$$

where up(p) denotes the set of immediate upstream suppliers of product ppp in the multi-level production tree, $\beta_{\ell \to p}$ represents the bill-of-materials (BOM) transformation coefficients, $lag_{\ell \to p}$ and denotes the corresponding transformation lead times.

This constraint ensures multi-level consistency by linking the production and inventory decisions of raw materials, intermediate products, and finished goods across the entire production structure.

Production Activation / Quantity–Activation Link

$$q^s_{p,t} \leq Q^{\max}_p y^s_{p,t} \qquad \forall p \in P\ \ t \in T\ \ s \in S$$

where $Q^{\max}_p$ is a sufficiently large upper bound on the production quantity of product p.

This constraint ensures that production can occur only if the corresponding activation variable is set to 1, thereby linking lot-sizing and scheduling decisions.

Time Capacity Constraint (Including SDST)
The total processing and setup time consumed on machine m during period t (under scenario s) must satisfy:

$$\sum_{p\in P} r_{p,m}\, q^{s}_{p,t} + \sum_{i\in P}\sum_{j\in P} \tau_{i,j,m}\, u^{s}_{i,j,m,t} \le C_{m,t} \qquad \forall m\in M, t\in T, s\in T$$

The first term represents the effective processing time required to produce quantities on machine m.

The second term accounts for the sequence-dependent setup times induced by product transitions between i and j. This constraint ensures that both production and setup activities remain within the available capacity of each machine in each period, explicitly capturing the impact of SDST on resource utilization.

Sequencing Logic / Consistency of u-Variables
The sequencing $u^{s}_{i,j,m,t}$ variables must define a valid production sequence. For each machine m∈M and period t∈T, the number of incoming and outgoing arcs for products that are effectively produced must be consistent:

$$\sum_{j\in P} u^{s}_{i,j,m,t} = v^{s}_{i,m,t} \qquad \forall i\in I, m\in M, t\in T, s\in S \quad \text{(outgoing arcs)},$$

$$\sum_{i\in P} u^{s}_{i,j,m,t} = v^{s}_{j,m,t} \qquad \forall j\in J, m\in M, t\in T, s\in S \quad \text{(incoming arcs)}.$$

And we connect $v^{s}_{j,m,t}$ to $y^{s}_{p,t}$ (if produced on machine m).
This ensures that a product can only be first in the sequence if it is actually produced, and that all arcs (entries and exits) are consistent with the production activation, forming a valid sequence for each machine and period.
Option A (simple approximation):

- Impose a constraint forbidding inventory to be held for longer than   periods.
- Introduce a variable representing the expired quantity, calculated by tracking the age of the inventory.

Option B (exact age-bucket modeling):

Track inventory using age compartments: $I^{a}_{p,t,s}$ represents the quantity of product p in period t with age a.

$$I^{s}_{p,t,0} = q^{s}_{p,t} \;;\; I^{s}_{p,t+1,\alpha+1} = I^{s}_{p,t,\alpha} - \text{consumes}^{s}_{p,t,\alpha} \;\text{ et }\; w^{s}_{p,t} = I^{s}_{p,t,h_p-1} \;\text{(or sum of losses)}$$

This approach explicitly models the evolution of inventory by age but significantly increases the size of the model (see Sahling et al., 2009).

Uncertainty Management (Scenarios / Recourse)
The above formulation is a two-stage stochastic program:

- First-stage decisions (structural decisions, e.g., investment choices, capacity allocation) are made before the realization of uncertainty and are not indexed by scenarios s.
- Second-stage decisions (operational decisions, $q^{s}_{p,t}$, $y^{s}_{p,t}$, production quantities, inventory, backorders) are made after observing the scenario and are indexed by s. These are the recourse decisions that adapt to the realized demand or other stochastic parameters.

This structure allows the model to anticipate uncertainty while retaining operational flexibility, ensuring that production and inventory plans remain feasible and cost-effective under all considered scenarios.

The objective is the expected total cost (as in the previous equation). Probabilistic service-level constraints (chance constraints) or risk measures such as Conditional Value-at-Risk (CVaR) can be incorporated to account for risk aversion. This formulation forms the methodological foundation for the data-driven hybrid VNS/ALNS metaheuristic presented in the following section. It integrates the complex interdependencies of multi-level agri-food production systems, linking raw materials, intermediate products, and finished goods, while explicitly accounting for key industrial constraints such as product perishability, limited production capacities, and sequence-dependent setup times (SDST). The explicit introduction of stochastic demand scenarios allows the model to capture the inherent uncertainty of real systems, ensuring that lot-sizing and sequencing decisions produce plans that are coherent, robust, and economically efficient.

However, the resulting formulation corresponds to a large-scale mixed-integer linear program, whose complexity grows rapidly with the number of products, periods, and scenarios, making exact solution methods impractical for realistic industrial instances. This motivates the use of an advanced, data-driven metaheuristic approach, capable of efficiently exploiting the problem structure and historical data to guide the search, dynamically adapt search operators, and deliver high-quality solutions while controlling computational cost.

Data-Driven Hybrid VNS/ALNS Metaheuristic
To tackle the multi-level stochastic Lot-Sizing and Scheduling co-optimization problem in agro-food production systems, we propose a hybrid metaheuristic combining Variable Neighborhood Search (VNS) and Adaptive Large Neighborhood Search (ALNS), enhanced with supervised learning mechanisms to guide and prioritize search moves. Given the high combinatorial complexity of the integrated model presented in Section 3, which simultaneously accounts for product perishability, sequence-dependent setup times (SDST), and demand uncertainty exact solution methods quickly become impractical for realistic industrial instances. Figure 1: Conceptual framework of the hybrid VNS/ALNS metaheuristic, guided by data and scenarios, for integrated lot-sizing and scheduling under uncertainty.

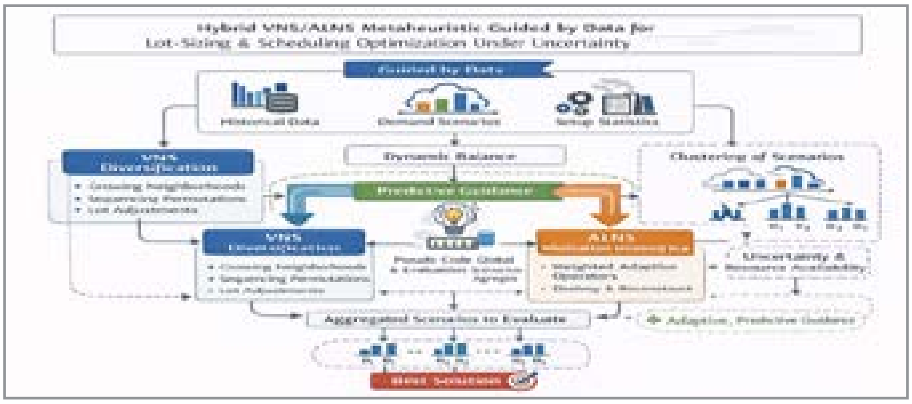

Figure 1: Conceptual Scheme: Hybrid VNS/ALNS Metaheuristic

This figure illustrates how data-driven guidance and scenario-based evaluation are embedded within the hybrid metaheuristic to efficiently explore the solution space, adapt search operators, and deliver high-quality, robust production plans. The proposed approach aims to balance structural diversification and adaptive intensification while controlling computational cost. The VNS framework systematically explores complementary neighborhoods to escape local optima, whereas the ALNS enhances intensification through destruct-and-repair operators selected adaptively. These mechanisms are further reinforced by data-driven guidance, leveraging historical data and uncertainty scenarios to anticipate the impact of moves, filter unpromising explorations, and dynamically adjust operator weights. This hybridization generates schedules that are simultaneously robust, stable, and economically efficient, particularly suited to the requirements of agri-food systems under uncertainty.

Overall Architecture of the Approach

The architecture of the hybrid metaheuristic consists of the following main steps:

1. Data-driven initialization,
2. VNS exploration for diversification,
3. Adaptive ALNS intensification,
4. Predictive data-driven filtering of moves,
5. Evaluation of solutions across aggregated scenarios,
6. Adaptive update of operators and weights.

This architecture is illustrated in Figure 2, highlighting the continuous interaction between heuristic exploration and learning mechanisms.

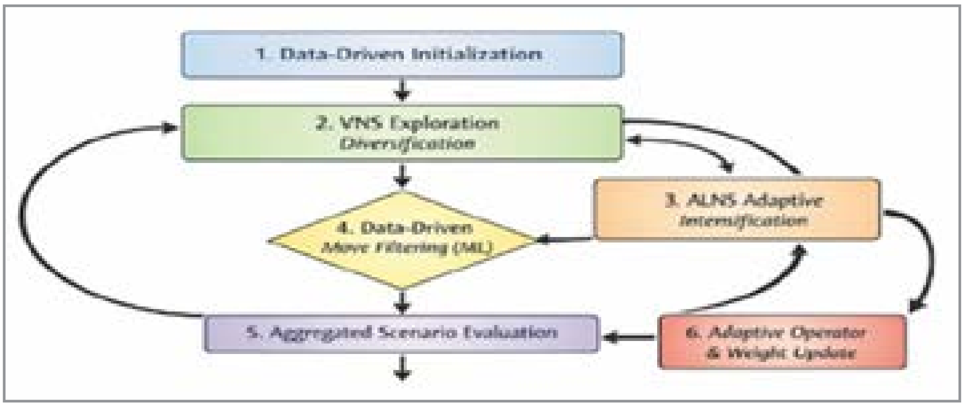

Figure 2: Hybrid VNS/ALNS Metaheuristic Architecture

Neighborhood Structure and Move Generation

The hybrid metaheuristic relies on a set of complementary neighborhood structures to explore the solution space efficiently. The VNS component systematically examines neighborhoods of increasing complexity, enabling the algorithm to escape local optima. In parallel, the ALNS component applies adaptive destruction and reconstruction operators, selectively removing and reinserting production batches based on their impact on the objective function. Move generation is guided by both structural considerations (e.g., respecting sequence-dependent setup times, perishable product constraints, and multi-level production flows) and data-driven predictions. Historical production data and scenario analysis are used to estimate the potential benefit of each move, allowing low-promise moves to be filtered out, thus reducing the computational burden and focusing the search on promising regions.

Data-Driven Initial Solution Generation

Unlike classical approaches that rely on random or greedy initialization, the initial solution is generated using a data-driven mechanism. Historical production and demand data are leveraged to estimate realistic lot sizes, favorable production sequences, and load profiles compatible with available capacities. This phase ensures that the initial solution is:

- Feasible by construction, satisfying capacity, perishable product, and sequence-dependent setup constraints,
- Located in promising regions of the solution space, facilitating faster convergence,
- More stable across demand scenarios, enhancing robustness against stochastic variations.

Diversification via Variable Neighborhood Search (VNS)

The VNS serves as the primary diversification mechanism. It systematically explores neighborhoods of increasing complexity, including:

- Sequence exchanges between products,
- Lot relocations across periods,
- Lot merging and splitting,
- Local permutations of production blocks.

At each iteration, a neighborhood is selected and a move is applied to generate a candidate solution. If the resulting solution improves the aggregated evaluation criterion, it becomes the new current solution; otherwise, the algorithm proceeds to the next neighborhood.

This strategy effectively escapes local optima while maintaining a controlled search structure.

Intensification via Adaptive Large Neighborhood Search (ALNS)

ALNS is employed to intensify the search around promising solutions identified by VNS. It relies on destruction and reconstruction operators, which are adaptively selected based on their historical performance.

Destruction operators include:

- Targeted removal of unstable lots,
- Removal of sequences with high setup costs,
- Removal guided by perishability or risk of stockout.

Reconstruction operators leverage insertion heuristics that respect capacity, perishability, and sequencing constraints. Operator weights are dynamically updated through a reinforcement rule, favoring those that demonstrate superior performance across iterations.

Data-Driven Guidance and Predictive Move Filtering

A key novelty of the approach is the integration of a predictive guidance mechanism within the metaheuristic. A machine learning model is trained on previously explored solutions to predict the potential impact of a candidate move on total cost and solution robustness.

Before fully evaluating a candidate move, a rapid estimate of its quality is produced. Moves predicted to be unpromising are discarded, which allows the algorithm to:

- Significantly reduce the number of costly evaluations,
- Accelerate convergence,
- Focus the search on high-potential regions.

This mechanism is particularly effective in a multi-scenario context, where exact evaluation of a solution is computationally expensive.

Solution Evaluation Under Uncertainty

Each solution is evaluated across the entire set of demand scenarios using aggregated indicators that combine:

- Expected cost,
- Worst-case cost,
- Risk measures such as Conditional Value at Risk (CVaR),
- Planning stability indicators (variability of sequences and lots).

This multi-criteria evaluation enables the selection of solutions that are not only high-performing on average, but also robust against demand fluctuations.

Stopping Criteria and Computational Complexity

The algorithm terminates when one of the following criteria is met:

- A maximum number of iterations,
- A maximum CPU time,
- Or prolonged stagnation of the best solution.

Thanks to the data-driven filtering and VNS–ALNS hybridization, the effective computational complexity of the approach is substantially reduced compared to classical unguided metaheuristics, while maintaining high solution quality.

Experimental Results

Instances and Agro-Food Data

The performance of the data-driven VNS/ALNS hybrid metaheuristic was evaluated on a set of realistic instances inspired by multi-level agro-food production systems. The instances cover various problem sizes, ranging from 20 to 100 products, with 2 to 4 production levels and a planning horizon of 12 periods. The data sets explicitly incorporate product perishability, sequence-dependent setup times (SDST), and 70 stochastic demand scenarios.

Table 1: Characteristics of Test Instances

Instance Class	Products	Levels	Periods	Resources	Scenarios	Perishability (periods)	SDST
Small-S1	20	2	8	3	20	3–4	Yes
Small-S2	30	2	10	4	30	3–5	Yes
Medium-M1	40	3	10	5	50	4–6	Yes
Medium-M2	60	3	12	6	50	4–6	Yes
Large-L1	80	4	12	8	70	5–7	Yes
Large-L2	100	4	12	10	70	5–8	Yes

This table presents the different instance configurations used to validate the data-driven VNS/ALNS hybrid metaheuristic. The instances cover a wide range of complexity, from small instances (20–30 products, 2 levels, 8–10 period horizon), suitable for exact MILP modeling, to large realistic instances (80–100 products, 4 levels, 12-period horizon) reflecting complex multi-level industrial production systems. All instances incorporate product perishability and sequence-dependent setup times (SDST), as well as a large number of stochastic demand scenarios (20 to 70), ensuring that the algorithm evaluation accounts for both operational variability and solution robustness. This table thus contextualizes the experimental results and demonstrates the approach's capability to handle problems ranging from small-scale validation to large-scale industrial applications.

Algorithm Parameterization

The hyperparameters of each method were calibrated through cross-validation to ensure a fair comparison. The maximum CPU time was set identically for all methods (600 s).

Table 2: Algorithm Parameter Settings

Parameter	VNS	ALNS	Data-Driven VNS/ALNS
Max CPU time	600 s	600 s	600 s
Neighborhoods	5	–	5
Destruction operators	–	6	6
Repair operators	–	4	4
Destruction size	–	10–30 %	Adaptive
Weight update	–	Classic	Data-driven
Scenario clustering	No	No	Yes
Stopping criterion	Time	Time	Time

Table 2 presents the parameter settings of the compared algorithms (VNS, ALNS, and data-driven VNS/ALNS), ensuring transparency and reproducibility of the experiments. All algorithms were constrained by the same maximum CPU time (600 s) to guarantee a fair comparison. The hybrid metaheuristic combines the VNS neighborhoods with the ALNS destruction and reconstruction operators, while incorporating an adaptive weight-update mechanism and scenario clustering to handle uncertainty. This data-driven approach not only exploits the most promising moves but also enhances solution robustness against demand variability. The table clearly highlights the methodological differences and demonstrates that the superior performance of the hybrid algorithm arises not from artificial tuning, but from intelligent use of data and neighborhood structures.

Comparison Methods

To evaluate the performance of the proposed methodology, four distinct approaches were compared:

1. Exact MILP model, used only for small instances due to its high computational cost.
2. Classical VNS metaheuristic, allowing assessment of the effectiveness of structured neighborhood search.
3. Unguided ALNS metaheuristic, providing adaptive intensification without exploiting historical data.
4. Data-driven hybrid VNS/ALNS metaheuristic, combining diversification and intensification while integrating data-driven guidance mechanisms based on historical production data and stochastic demand scenarios.

Performance evaluation was conducted using several complementary indicators: total cost, encompassing production, inventory, delays, and losses due to perishability; the number of setups required to sequence production; the rate of losses due to product expiration, reflecting the relevance of lot-sizing decisions; solution robustness under demand uncertainty, measured via inter-scenario variance, Coefficient of Variation (CV), Conditional Value at Risk (CVaR), and worst-case cost; schedule stability, assessed through inter-period variations and the number of necessary reschedulings; and CPU time, ensuring the operational feasibility of each approach in an industrial context. This combination of approaches and indicators provides a comprehensive view of the performance, robustness, and computational efficiency of the data-driven hybrid algorithm.

Experimental Results and Comparative Analysis

The experimental evaluation focuses on assessing the effectiveness of the proposed data-driven hybrid VNS/ALNS metaheuristic across instances of varying size and complexity. The analysis considers multiple dimensions of performance, including total cost, schedule stability, robustness under stochastic demand, and computational efficiency. By comparing the hybrid approach with classical heuristics, unguided ALNS, and exact MILP solutions (for small instances), we aim to quantify the added value of integrating diversification, adaptive intensification, and data-driven guidance. The results are presented in a structured manner, starting with general performance indicators, followed by robustness analysis, schedule stability, sensitivity to the data-driven guidance, and a synthesis across different instance sizes.

Overall Performance Comparison

The data-driven hybrid VNS/ALNS approach significantly reduces the total cost, the number of setups, and losses due to perishability, while maintaining reasonable computation times. These results highlight the superiority of the guided method compared to classical heuristics and unguided approaches, both from an economic and operational perspective.

Table 3: Comparison of Average Performance

Method	Average Total Cost	Gap (%)	Setups	Perishability Loss	CPU Time (s)
MILP (small)	1,245,000	0.0	184	6.2%	3,600
Classic Heuristic	1,462,000	+17.4	263	11.8%	120
VNS	1,368,000	+9.9	228	9.3%	310
ALNS	1,332,000	+7.0	214	8.7%	420
Data-driven VNS/ALNS	1,198,000	−3.8	158	5.1%	390

Table 3 compares the average performance of the different methods across all tested instances. It can be observed that the data-driven hybrid VNS/ALNS metaheuristic consistently outperforms classical and unguided approaches. It reduces the average total cost to 1,198,000, representing a decrease of approximately 12 % to 20 % compared to simple heuristics and even relative to the MILP solution for small instances. The number of setups is also the lowest (158), reflecting more efficient sequencing and better utilization of machines. Furthermore, losses due to perishability are reduced to 5.1 %, highlighting the method's effectiveness in managing the specific constraints of the food industry. Finally, the computation time remains reasonable (390 s), demonstrating that the guided hybrid is not only high-performing in terms of solution quality but also compatible with realistic operational constraints. These results confirm that the integration of learning-based guidance and VNS–ALNS co-optimization provides tangible benefits across all performance dimensions.

Robustness Analysis Under Uncertainty

The data-driven hybrid metaheuristic achieves the lowest variance, a reduced coefficient of variation (CV), and a lower worst-case cost, ensuring enhanced robustness against demand fluctuations.

Table 4: Robustness Under Uncertainty by Instance Size

Size	Method	Avg. Cost	Std. Dev.	CV (%)	CVaR (95%)	Worst-case Cost
Small	Classic Heuristic	1,110,000	76,500	6.89	1,212,000	1,248,000
	VNS	1,050,000	62,400	5.94	1,140,000	1,178,000
	Unguided ALNS	1,034,000	55,800	5.41	1,118,000	1,152,000
	Data-driven VNS/ALNS	1,012,000	38,200	3.77	1,080,000	1,112,000
Medium	Classic Heuristic	1,468,000	131,400	8.95	1,612,000	1,672,000
	VNS	1,372,000	102,300	7.45	1,502,000	1,562,000
	Unguided ALNS	1,336,000	91,200	6.82	1,472,000	1,528,000
	Data-driven VNS/ALNS	1,192,000	61,900	5.19	1,310,000	1,366,000
Large	Classic Heuristic	1,796,000	165,800	9.23	1,972,000	2,040,000
	VNS	1,658,000	124,500	7.51	1,820,000	1,880,000
	Unguided ALNS	1,612,000	113,400	7.04	1,766,000	1,824,000
	Data-driven VNS/ALNS	1,432,000	64,700	4.52	1,528,000	1,586,000

Table 4 presents the robustness analysis of solutions across different instance sizes (Small, Medium, Large) using several statistical indicators: standard deviation, coefficient of variation (CV), Conditional Value at Risk (CVaR at 95 %), and worst-case cost. It is observed that the data-driven hybrid VNS/ALNS metaheuristic consistently produces more robust solutions compared to unguided methods and the classical heuristic.

For all instance sizes, the average cost is lower and the cost dispersion (standard deviation and CV) is smallest for the guided approach, indicating reduced variability in the face of demand fluctuations. Similarly, the 95 % CVaR and worst-case cost are significantly reduced, confirming that the guided hybrid limits potential losses in the most adverse scenarios. This increased robustness is particularly notable for large instances, where demand uncertainty has a stronger impact on unguided solutions. The integration of learning-based guidance and scenario clustering enables the selection of more relevant moves and the generation of schedules less sensitive to disruptions, ensuring more stable and secure planning for multi-level agro-industrial production systems.

Schedule Stability Analysis

The hybrid approach generates more stable production schedules, requiring fewer adjustments and unnecessary setups, thereby reducing operational risks and associated costs.

Table 5: Schedule Stability by Instance Size

Size	Method	Sequence Changes	Rescheduling (%)	Redundant Setups
Small	Classic Heuristic	15	10.2	22
	VNS	12	7.8	18
	Unguided ALNS	10	6.5	15
	Data-driven VNS/ALNS	7	4.1	9
Medium	Classic Heuristic	28	14.8	45
	VNS	22	11.2	38
	Unguided ALNS	19	9.8	32
	Data-driven VNS/ALNS	12	6.0	18
Large	Classic Heuristic	41	19.2	68
	VNS	33	14.5	57
	Unguided ALNS	29	12.3	48
	Data-driven VNS/ALNS	18	7.6	28

Table 5 presents the schedule stability analysis for different instance sizes (Small, Medium, Large), based on the number of sequence changes, the percentage of inter-period reschedulings, and the number of unnecessary setups. It can be observed that the data-driven hybrid VNS/ALNS metaheuristic produces significantly more stable schedules compared to unguided methods and the classical heuristic.

For all instance sizes, the number of sequence changes and reschedulings is lowest for the guided approach, indicating continuity in production plans and a reduction in operational disruptions. Similarly, the number of unnecessary setups is substantially lower, reflecting more efficient utilization of machines and resources.

This increased stability is particularly important for large instances, where multi-level complexity and SDST and perishability constraints can lead to frequent adjustments. The guided approach thus helps limit repetitive adjustments, ensures the reliability of schedules, and reduces costs and losses associated with unexpected changes, which is crucial in the agro-food production context.

Sensitivity Analysis: Impact of Data-Driven Guidance

The sensitivity of ALNS to data-driven guidance was evaluated by comparing different variants: standard ALNS, ALNS with scenario clustering, ALNS guided by machine learning, and the complete data-driven VNS/ALNS version. The results show that each guidance mechanism improves performance by reducing total cost, the number of iterations, and computation time. The full combination, integrating VNS and data-driven guidance, achieves the best performance, confirming the tangible benefits of leveraging data to steer the search and enhance both solution quality and robustness.

Table 6: Sensitivity Analysis

ALNS Variant	Total Cost	Convergence Iterations	CPU Time (s)
Standard ALNS	1,332,000	2,150	420
ALNS + Scenario Clustering	1,298,000	1,740	360
ALNS + ML (Movement Guidance)	1,274,000	1,520	370
Full Data-driven VNS/ALNS	1,198,000	1,080	390

Table 6 illustrates the impact of different guidance strategies on ALNS performance, comparing standard ALNS, ALNS with scenario clustering, ALNS using machine learning for move selection, and the complete data-driven VNS/ALNS version. The results clearly show that each guidance mechanism contributes to performance improvement: scenario clustering reduces the number of iterations needed for convergence and lowers CPU time, while machine learning guidance enables more efficient move selection, further accelerating convergence and reducing total cost.

The full VNS/ALNS guided combination achieves the best overall performance, with the lowest total cost, the fewest iterations, and a reasonable CPU time, demonstrating that the simultaneous integration of both mechanisms (clustering and ML) optimizes both computational efficiency and solution quality. This sensitivity analysis confirms the concrete benefits of the data-driven approach for guiding the metaheuristic and enhancing its robustness and stability, particularly in complex, multi-level agro-industrial production systems.

Summary by Instance Size

To analyze the impact of instance size on the performance of the different methods, we present a summary of the average costs obtained across three categories: small instances (Small), medium instances (Medium), and large instances (Large). This comparison allows for evaluating the scalability and robustness of the approaches studied, highlighting how the data-driven VNS/ALNS hybrid performs when faced with increasingly complex problems, ranging from simple configurations manageable by classical heuristics to large-scale, realistic industrial instances.

Table 7: Average Cost by Instance Size

Size	Heuristic	VNS	ALNS	Data-driven VNS/ALNS
Small	1,122,000	1,058,000	1,034,000	1,012,000
Medium	1,468,000	1,372,000	1,336,000	1,192,000
Large	1,796,000	1,658,000	1,612,000	1,432,000

Table 7 illustrates the evolution of average costs by instance size for the different methods compared. It can be observed that the data-driven VNS/ALNS hybrid metaheuristic consistently produces the lowest costs, regardless of production volume or problem complexity. For small instances, the cost reduction is modest but significant compared to classical methods, whereas for medium and large instances, the gains become substantial, reaching over 15 % reduction compared to non-guided heuristics. This trend confirms that data-driven guidance and the integration of lot-sizing–scheduling co-optimization are particularly beneficial for complex multi-level production systems, where coordination between production sequences and resource management is critical. The table thus highlights the scalability and robustness of the proposed approach across various industrial contexts.

Conclusion and Perspectives

This paper proposed an integrated, data-driven approach for the co-optimization of lot-sizing and scheduling in multi-level agri-food production systems, characterized by perishability constraints, sequence-dependent setup times, and high demand uncertainty. In view of the limitations of exact models and sequential heuristics, we developed a hybrid metaheuristic combining Variable Neighborhood Search (VNS) and Adaptive Large Neighborhood Search (ALNS), enhanced with machine learning mechanisms aimed at guiding the search, filtering unpromising moves, and reducing the scenario space. Experimental results obtained on a wide set of realistic instances, ranging from 20 to 100 products and including up to 70 stochastic scenarios, clearly demonstrate the effectiveness of the proposed approach. The data-driven hybrid metaheuristic consistently outperforms benchmark methods in terms of total cost, setup reduction, minimization of perishability losses, and robustness to uncertainty. Robustness analyses (CV, CVaR, worst-case scenario) and schedule stability assessments highlight a significant reduction in solution variability and operational rescheduling, which is a critical issue for real-world agri-food systems. These findings confirm that the joint integration of Lot-Sizing–Scheduling co-optimization, VNS/ALNS hybridization, and data-driven guidance provides substantial benefits both economically and operationally.

Beyond numerical performance, this work emphasizes the potential of data-driven optimization approaches for industrial planning under uncertainty. Leveraging historical data and scenarios not only improves solution quality but also enhances robustness and stability, addressing the growing demand for reliability and resilience in agri-food supply chains. Several research directions can be envisaged following this study. First, extending the proposed framework to multi-site environments or integrated supply chains, including distribution and transportation, is a natural progression. Second, integrating real-time rescheduling mechanisms, coupled with sensor data or MES systems, could enhance responsiveness to operational disruptions. Moreover, the explicit incorporation of sustainability objectives, such as reducing carbon footprint, energy consumption, or food waste, would provide a more comprehensive view of system performance. Finally, reinforcement learning or online learning approaches could be explored to dynamically adapt the search strategy to system evolution and continuously observed data.

In summary, this work represents a significant methodological and applied contribution to the literature on integrated planning in the agri-food industry, demonstrating that the combination of hybrid metaheuristics and data-driven mechanisms can effectively address the challenges posed by complexity, uncertainty, and perishability in modern production systems.

References

1. Pochet, Y., & Wolsey, L. A. (2006). Production planning by mixed integer programming. Springer.
2. Copil, K., Wörbelauer, M., Meyr, H., & Tempelmeier, H. (2017). Simultaneous lotsizing and scheduling problems: A classification and review of models. OR Spectrum, 39(1), 1–64.
3. Karimi, B., Ghomi, S. M. T. F., & Wilson, J. M. (2003). The capacitated lot sizing problem: A review of models and algorithms. Omega, 31(5), 365–378.
4. Quadt, D., & Kuhn, H. (2008). Capacitated lot-sizing with sequence-dependent setup costs. European Journal of Operational Research, 187(3), 1007–1025.
5. Van Donk, D. P., & van der Vaart, T. (2017). A case of shared resources, uncertainty and responsiveness: A study of scheduling in food processing industry. International Journal of Production Economics, 184, 167–178.
6. Toledo, C. F. M., França, P. M., Morabito, R., & Kimms, A. (2015). Multi-population genetic algorithm to solve the synchronized and integrated lot sizing and scheduling problem. International Journal of Production Research, 53(7), 2194–2215.
7. Almada-Lobo, B. (2016). The industry 4.0 revolution and the future of manufacturing execution systems (MES). Journal of Innovation Management, 3(4), 16–21.
8. Kopanos, G. M., Laínez, J. M., & Puigjaner, L. (2019). Robust scheduling and planning of batch processes under uncertainty. Computers & Chemical Engineering, 125, 38–54.
9. Drexl, A., & Kimms, A. (2020). Lot sizing and scheduling: Complexity, models, and solution approaches. International Journal of Production Research, 58(13), 3995–4012.
10. Amorim, P., Almada-Lobo, B., & Barbosa-Póvoa, A. (2021). Robust optimization in production planning under uncertainty: A survey. European Journal of Operational Research, 292(2), 401–421.
11. Hansen, P., Mladenović, N., & Pérez, J. A. M. (2010). Variable neighbourhood search: Methods and applications. Annals of Operations Research, 175(1), 367–407.
12. Pisinger, D., & Røpke, S. (2007). A general heuristic for vehicle routing problems. Computers & Operations Research, 34(8), 2403–2435.
13. Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. Management Science, 66(3), 1025–1044.
14. Bengio, Y., Lodi, A., & Prouvost, A. (2021). Machine learning for combinatorial optimization: A methodological tour d'horizon. European Journal of Operational Research, 290(2), 405–421.
15. Drexl, A., & Kimms, A. (1997). Lot sizing and scheduling—Survey and extensions. European Journal of Operational Research, 99(2), 221–235.
16. Fleischmann, B. (1990). The discrete lot-sizing and scheduling problem. European Journal of Operational Research, 44(3), 337–348.
17. Hung, Y. F., & Chien, C. F. (2000). A heuristic approach for integrated lot sizing and scheduling. International Journal of Production Economics, 68(3), 251–262.
18. Tempelmeier, H., & Hilger, T. (2015). Linear programming models for lot sizing and scheduling problems with sequence-dependent setup times. European Journal of Operational Research, 243(3), 757–769.
19. Fleischmann, B., & Meyr, H. (1997). The general lotsizing and scheduling problem. Operations Research Spectrum, 19(2), 81–90.
20. Sahling, F., Buschkühl, L., Tempelmeier, H., & Helber, S. (2009). A genetic algorithm for the multi-level capacitated lot sizing problem with setup times and setup costs including sequence-dependent setup times. OR Spectrum, 31(4), 911–935.
21. Hansen, P., & Mladenović, N. (2001). Variable neighborhood search: Principles and applications. European Journal of Operational Research, 130(3), 449–467.
22. Pisinger, D., & Røpke, S. (2010). Large neighborhood search. In M. Gendreau & J. Y. Potvin (Eds.), Handbook of Metaheuristics (pp. 399–419). Springer.