# A framework for multi-objective stochastic lot sizing with multiple decision stages

**Fabian Friese** · **Stefan Helber**

Received: date / Accepted: date

**Abstract** In stochastic lot sizing subject to dynamic and random demand, the minimization of operational costs is not the only conceivable objective. Minimizing the tardiness in customer demand satisfaction is no less important. Furthermore, the decision maker is interested in production plan stability. Therefore, we consider those three objectives simultaneously and propose a multi-objective model formulation and decision-making framework of the stochastic capacitated lot sizing problem (MO-SCLSP).

Demand is modeled via the Martingale Model of Forecast Evolution to allow gradual adaptations of the demand forecasts due to sequential market observations. We propose an interactive multi-objective optimization algorithm for solving the MO-SCLSP, that systematically takes prior demand realization information into account. In multiple decision stages, periodic re-optimizations are carried out, allowing to adjust the production plan to the actual demand realizations.

In each decision stage, methods from multi-objective optimization are applied to derive a set of Pareto-optimal solutions. These Pareto-optimal solutions outline the attainable objective space, thus supporting the decision maker in taking an informed and economically profound position between prioritizing low operational costs, high delivery reliability and low production plan nervousness.

Fabian Friese
Institute of Production Management, Leibniz University Hannover, Königsworther Platz 1, 30167 Hannover, Germany
Tel.: +49-511-7625650
Fax: +49-511-7624863
E-mail: fabianfriese@gmx.de
ORCID: 0000-0002-5131-5841

Stefan Helber
Institute of Production Management, Leibniz University Hannover, Königsworther Platz 1, 30167 Hannover, Germany
Tel.: +49-511-7625650
Fax: +49-511-7624863
E-mail: helber@prod.uni-hannover.de
ORCID: 0000-0002-9455-1828

**Keywords** Multi-objective lot sizing · Stochastic lot sizing · Multi-objective optimization · Multiple decision stages · System nervousness · Service levels

## 1 Motivation

Make-to-stock production environments are usually affected by random influences arising from uncertain customer demand. Deciding on production times and lot sizes in uncertain production environments requires a systematic consideration of this uncertainty. In the presence of demand uncertainty, it is not reasonable (and, in practice, even impossible) to warrant on-time delivery for all customer demands, as it is usually assumed in deterministic problem settings. Safety stocks are produced, which, on the one hand, buffer against unexpectedly high demands and therefore limit average tardiness in demand satisfaction (implying high delivery reliability), but on the other hand, their production and storage lead to increased operational costs. This causes a trade-off between low operational costs and high delivery reliability.

A production plan determined based on uncertain information can turn out to be deficient for a particular realization of random parameters, e.g., demand. This can motivate adjustments of the production plans after demand realization for a certain timespan. With production plan adjustments both the operational costs (e.g. holding costs, in the case of low demand realizations) and the tardiness (in the case of high demand) can be reduced. However, production plan adjustments also induce system nervousness, which can propagate to upstream process steps and induce the undesirable bullwhip-effect (see e.g. Lee et al. (1997)).

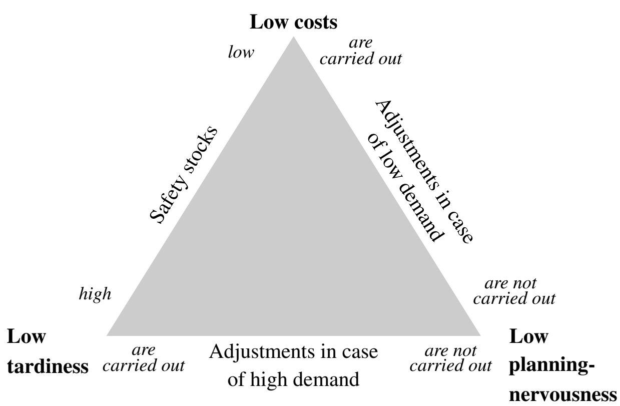



Fig. 1: Triangle of tension of stochastic lot sizing

Figure 1 illustrates the conflicting relationships between the three objectives of low operational costs, low tardiness in customer demand satisfaction and low system nervousness explained above. The precise connections between those objectives strongly depend on the parameters of the planning situation. Seeking a compromise between these conflicting objectives requires a statement of preferences.

The literature discusses many single-objective special cases of this problem. Usually a decision maker is required to decide a priori on an aspired level of delivery reliability (often expressed by relative service level formulations) and an accepted level of system nervousness (often selected from the seminal categorization scheme proposed by Bookbinder and Tan (1988)). Given this specific set of preferences, the remaining objective function of minimizing the operational costs is optimized.

Solving a single problem instance of a single-objective stochastic lot sizing problem with set aspiration levels for two of the three objective functions only reveals a single Pareto-optimal point of the attainable objective space. There might exist different Pareto-optimal solutions with substantially better values for some of the objective functions and only slightly poorer values for the worsening objective functions. Particularly good compromise solutions cannot be guaranteed by solving a single-objective lot sizing problem in which the other objectives have been transformed into constraints.

In practice, defining an aspired delivery reliability level and deciding how much nervousness to accept is difficult, to say the least. The economic consequences of selecting a specific set of aspiration levels are hard to predict. It takes a comprehensive overview of the objective space to take an informed and economically profound decision on a favourable compromise solution. In addition, the decision maker may not even be able to formally describe his or her complete preference functions. However, a multi-objective approach reveals multiple Pareto-optimal solutions on the Pareto front. Thus, the precise connections between the objective function metrics can be analysed and the favourite Pareto-optimal production plan can be found. Preferences among the objectives usually depend on the decision makers domain knowledge and can usually not be externalized as mathematical functions. Only by comparing different solutions is the decision maker able to reveal his or her preference order and to identify his or her favored compromise solution.

To the best of the authors' knowledge, there does not exist an approach that systematically evinces the relationships among the three aforementioned objectives in the literature. This research hence provides a multi-objective generalization with multiple decision stages of the stochastic capacitated lot-sizing problem (MO-SCLSP), which treats the three conceptional objectives as coequal objective functions. In each decision stage, the proposed approach for solving the MO-SCLSP takes the newly observed demand realization information into account to derive a set of Pareto-optimal solutions for different levels of system nervousness and different levels of delivery reliability. In an interactive approach, Pareto-optimal solutions are generated by determining the ideal-points (see Miettinen (1999)) and applying the augmented multidimensional $\varepsilon$-constraint method (see Chankong and Haimes (1983), Haimes et al. (1971) and Mavrotas (2009)).

The decision maker iteratively analyses the subset of the objective space outlined by the solutions determined so far and chooses which solutions to determine next until he or she has enough information to select a final solution. The contribution of the proposed methodology is threefold:

- It enables the decision maker to decide on an informed positioning between the conflicting objectives by giving a clear impression of the objective space.
- It ensures favourable solutions for each demand realization trajectory by allowing for adaptations of the production plan to make use of the observed demand realization information.
- It supports the decision maker in ascertaining his or her own preferences between the conflicting objectives.

The remainder of this paper is structured as follows: Section 2 discusses relevant literature. In Section 3, the problem setting is presented in detail and a conceptional model formulation for the MO-SCLSP for arbitrary demand distribution and unspecific objective function metrics is introduced. Applications for problems with discrete demand and with normally distributed demand are presented as well. In Section 4, re-optimizations of the production plans in multiple decision stages are discussed. Section 5 presents a multi-objective optimization model with specific objective functions, which is solved in each decision stage. An interactive approach to determine Pareto-optimal solutions for this model over multiple decision stages is presented in Section 6. Numerical studies are shown in Section 7. Section 8 concludes and outlines future research topics.

## 2 Relevant literature

### 2.1 Handling multiple objectives in stochastic lot sizing

Dynamic lot-sizing problems have been studied under a broad variety of differing assumptions. One stream of research assumes deterministic demand. Recent reviews on deterministic dynamic lot sizing can be found in Karimi et al. (2003), Jans and Degraeve (2008), Robinson et al. (2009) and Buschkühl et al. (2010).

With increasing interest, research on the stochastic counterpart of the dynamic lot sizing problem is carried out. A particular emphasis is put on models with uncertain demand. Reviews of lot sizing subject to stochastic demand can be found in Tempelmeier (2013), Winands et al. (2011) and Sox et al. (1999).

The literature on lot sizing considering *multiple objectives* is limited. Some studies on deterministic lot sizing consider different additional objectives. Azadnia et al. (2015) take environmental and social objectives into account, Hajipour et al. (2015) add the objectives to level the production volume in different production periods and to produce as close to a just-in-time level as possible, and Mehdizadeh et al. (2016) seek compromise solutions between costs and required storage space. Multi-objective formulations are also found in the literature on lot-sizing problems with integrated supplier selection: Rezaei and Davoodi (2011) and Rezaei et al. (2016) aim at finding balanced solutions leading to low costs, high product quality and high service levels.

Studies on stochastic lot sizing are limited to single-objective models, usually with a focus on optimizing operational costs. Tardiness is usually controlled with different service level constraints, thus treating it as "satisficing" objective rather than as optimizing objective (see Simon (1956)). Similarly, a rigid decision on how much system nervousness to accept is assumed.

In an influential publication, Bookbinder and Tan (1988) introduce three basic strategies to restrict system nervousness: static uncertainty, static-dynamic uncertainty, and dynamic uncertainty. In the following, this concept is applied to classify relevant publications.

According to the *static uncertainty* strategy, nervousness is fully eliminated by freezing the production plans at the beginning of the planning horizon, thus executing the frozen production plans irrespective of the realizations of uncertain demand. The static approach is applied by Tempelmeier and Herpers (2011) to an uncapacitated stochastic lot sizing problem with $\beta$-service-level constraints and solved with a solution procedure based on a modified shortest-path problem. In Tempelmeier (2011a) the capacitated counterpart of the stochastic lot sizing problem with $\beta$-service-level constraints is studied. Due to the increasing computational effort, he solves the problem heuristically by a combination of column generation and the $\mathrm{ABC}_\beta$-heuristic established in Tempelmeier (2010). Recent publications discussing static approaches include Helber et al. (2013) and Tempelmeier and Hilger (2015), who apply piecewise linearization to solve the stochastic optimization problem with $\delta$- or $\beta$-service-level constraints, respectively. Moreover, relevant extensions of the capacitated lot sizing problem are also transferred to the stochastic counterpart applying the static uncertainty strategy. Li and Song (2015) solve the multi-level capacitated lot sizing problem with stochastic demand. Sequence-dependent changeovers, originally proposed for a deterministic setting by Gopalakrishnan et al. (1995), were studied by De Smet et al. (2020) for a stochastic planning environment.

According to the *static-dynamic uncertainty* strategy, the decisions on the production times and the order-up-to-levels are taken at the beginning of the planning horizon as well. The actual lot sizes, however, are determined after the respective demand realizations have been observed. This strategy is applied to an uncapacitated lot sizing problem by Tarim and Kingsman (2004). While they consider $\alpha$-service-level constraints, the research of Tempelmeier (2007) widens the scope by also considering $\beta$-service-level constraints. Rossi et al. (2015) developed a MIP modeling approach based on piecewise linear approximation for stochastic lot sizing using the static-dynamic uncertainty strategy. They study various service level constraints as well as backorder penalty costs.

The *dynamic uncertainty* strategy allows for adaptations of both the production times and the production quantities after the demand realizations have been observed. According to Bookbinder and Tan (1988), the dynamic uncertainty strategy is usually not desired in practical applications, as it can cause exceptional high costs if the setup cost rate is considerably higher than the inventory cost rate. However, Tunc et al. (2013) present a simple approach to assess the costs induced by nervousness by comparing optimal solutions determined with the three different uncertainty strategies.

Other publications deal with different special cases that do not fit in the classification scheme proposed by Bookbinder and Tan (1988). Meistering and Stadtler (2017) propose the stabilized-cycle strategy. In contrast to static approaches, it takes realizations of uncertain demand into account. Low tardiness is prioritized over low nervousness, thus allowing for production plan adjustments only if an aspired service level cannot be met without adjustments. Adjustments of the production plans just to reduce operational costs are not allowed. Operational costs are thereby implicitly considered to be less important than the nervousness or tardiness. They are optimized, given that the aspired service level is satisfied. A focus on the trade-off between operational costs and quantity-oriented nervousness can be found in Koca et al. (2018). Nervousness is penalized in the objective function, thus optimizing a scalarized formulation of a bi-objective optimization problem considering operational costs and penalty costs for quantity-oriented system nervousness. Tavaghof-Gigloo and Minner (2021) compare models for stochastic lot sizing with cyclic $\beta$-service level constraints with and without re-optimization options with a sequential approach, where in a first step safety stocks are determined and in a second step a deterministic model considering those safety stocks is solved.

In the context of the publications discussed above, our main contribution is a multi-objective approach for a stochastic capacitated lot-sizing problem simultaneously considering the objectives of low operational costs, low tardiness in customer demand satisfaction, and low system nervousness, which can be considered a generalization of the aforementioned special cases.

### 2.2 Adaptation of production plans to updated demand information

Adaptations of production plans to newly observed demand information are studied in various fields of application. Bookbinder and H'ng (1986) and Bookbinder and Tan (1988) study the single-product uncapacitated probabilistic lot sizing problem with $\alpha$-service-level constraints in a rolling horizon approach. Oezer and Wei (2004) divide demand into an already observed and an unobserved part and call this concept *advance demand information*. Albey et al. (2015) apply the framework of the Martingale Model of Forecast Evolution (MMFE) proposed by Heath and Jackson (1994) to a production planning problem in semiconductor manufacturing. The same framework is applied in a context of stochastic dynamic programming for a single level, single item production planning problem by Claisse et al. (2016). The influence of conditional covariances in the MMFE is studied in Norouzi and Uzsoy (2014) with an application to inventory planning. Ziarnetzky et al. (2018) investigate the multiplicative variant of the MMFE incorporated into production planning of semiconductor wafers and show that forecast updates can lead to an increase in terms of both profit and service level. In a subsequent publication, Ziarnetzky et al. (2020) examine the nervousness of updated production plans. In a study on chance-constrained programs for a production planning problem with demand forecast evolution under rolling horizon planning, Albey et al. (2016) emphasize that forecast evolution is particularly advantageous in production systems with low capacity utilization. An application to lot sizing can be found in Rehman et al. (2019), who apply the MMFE framework to an autoregressive process to model demand for a multi-level, multi-stage lot-sizing problem. Forel and Grunow (2022) integrate the additive and multiplicative MMFE into stochastic lot sizing with backlog costs and rolling horizons.

We contribute to this stream of the literature by combining the Martingale Model of Forecast Evolution with multi-objective stochastic lot sizing with multiple decision stages, systematically controlling tardiness and nervousness. Based on updated demand forecasts, the production plans can be adjusted if a solution with a favorable combination of objective function values is found.

## 3 Conceptional model for the multi-objective stochastic capacitated lot sizing problem (MO-SCLSP)

### 3.1 Problem setting and the role of the decision maker

To keep the presentation as simple as possible and without loss of generality, we consider a single level production process with a single production resource producing multiple products to satisfy dynamic and random customer demand with arbitrary demand distribution function. Demand information is revealed continuously, gradually reducing the uncertainty over time. The final demand realization can be observed at the end of the respective period. Costly setup activities have to be performed prior to production in any period. The capacity available for production and setup activities is limited. It can be extended by (costly) overcapacity, which is limited as well. Processed products can be stored at a fixed holding cost rate. Due to the stochasticity of demand, a limited amount of backorders and backlogs is accepted. We refer to the back<u>orders</u> in a period as the quantity of unmet demand in this period, while the back<u>logs</u> constitute the cumulative backorders that have not caught up yet (see Gade and Küçükyavuz (2013)). Over the long run, the average production shall suffice to meet (at least) the average demand, thereby preventing systematic underproduction.

The planning task is to decide on the production plan in terms of the production times (setup pattern) and production quantities (lot sizes). Deciding on a production plan also implies a decision on the consumed overcapacity. The tentative production plan is revisited periodically, allowing for adaptations based on new demand information. The production plan should represent a compromise between low operational (setup and holding) costs[1], low tardiness in customer demand satisfaction, and a high level of long term production predictability.

Multiple competing solutions are evaluated based on a vector of multiple objective function values, rather than computing a single value of a common objective function, and several Pareto-optimal solutions exist. A solution is Pareto-optimal if one objective function value can only be improved at the cost of another one. The set of Pareto-optimal solutions constitutes the Pareto front. Selecting a solution from the Pareto front as the final solution reveals additional (non-explicit) information about preferences between the objectives. This selection is made by a decision maker, based on the objective space and according to preferences between the objective functions. These preferences depend on the decision makers domain knowledge and are assumed to be implicit. This means, that a priori, the decision maker is not able to externalize how to select a solution from a set of Pareto-optimal solutions. Only by inspecting a set of Pareto-optimal solutions can he or she identify the preferred final solution.

### 3.2 Conceptional optimization model for the MO-SCLSP

Respecting the problem setting described in the previous subsection, we now present the conceptional model (1) - (13) for the multi-objective stochastic capacitated lot sizing problem (MO-SCLSP) for arbitrary objective function metrics and arbitrary demand distribution functions based on the notation depicted in Table 1.

[1] In the following, we refer to the decision-relevant setup, holding and overtime costs as *operational costs*.

Table 1: Notation for the MO-SCLSP

| | |
|---|---|
| Indices and index sets: | |
| $t \in \mathscr{T}$ | periods |
| $k \in \mathscr{K}$ | products |
| Parameters: | |
| $b_t$ | capacity in period $t$ |
| $hc_k$ | holding cost rate of product $k$ |
| $M$ | big-M parameter |
| $o_t^{max}$ | maximum overtime in period $t$ |
| $oc$ | overtime cost rate |
| $sc_k$ | setup cost rate of product $k$ |
| $tp_k$ | processing time of product $k$ |
| $ts_k$ | setup time of product $k$ |
| $\overline{Q}_{kt}$ | selected lot sizes from the previous decision stage |
| $\overline{X}_{kt}$ | selected setup pattern from the previous decision stage |
| Random variables: | |
| $\mathbf{BL}_{kt}$ | backlog of product $k$ in period $t$ |
| $\mathbf{D}_{kt}$ | demand of product $k$ due in period $t$ |
| $\mathbf{Y}_{kt}$ | net inventory of product $k$ in period $t$ |
| $\mathbf{YP}_{kt}$ | physical inventory of product $k$ in period $t$ |
| Decision variables: | |
| $O_t \geq 0$ | utilized overtime in period $t$ |
| $Q_{kt} \geq 0$ | production quantity (lot size) of product $k$ in period $t$ |
| $X_{kt} \in \{0,1\}$ | binary setup variable of product $k$ in period $t$ |
| Abstract functions: | |
| $f_C$ | abstract function for a cost metric |
| $f_T$ | abstract function for a tardiness metric |
| $f_N$ | abstract function for a nervousness metric |

**MO-SCLSP**

$$\text{min costs} = f_C\ (X_{kt},\ \mathbf{YP}_{kt},\ O_t\ ; t = 1,...,T\ ; k = 1,...,K) \quad (1)$$

$$\text{min tardiness} = f_T\ (\mathbf{BL}_{kt}; t = 1,...,T\ ; k = 1,...,K) \quad (2)$$

$$\text{min nervousness} = f_N\ (\overline{X}_{kt} - X_{kt},\ \overline{Q}_{kt} - Q_{kt}\ ; t = 1,...,T\ ; k = 1,...,K) \quad (3)$$

s.t.

$$Q_{kt} \leq M \cdot X_{kt} \qquad \forall\, k,t \quad (4)$$

$$\sum_{k=1}^{K} tp_k \cdot Q_{kt} + ts_k \cdot X_{kt} \leq b_t + O_t \qquad \forall\, t \quad (5)$$

$$O_t \leq O_t^{max} \qquad \forall\, t \quad (6)$$

$$\mathbf{Y}_{k,t-1} + Q_{kt} - \mathbf{Y}_{kt} = \mathbf{D}_{kt} \qquad \forall\, k,t \quad (7)$$

$$\mathbf{YP}_{kt} = \max(0, \mathbf{Y}_{kt}) \qquad \forall\, k,t \quad (8)$$

$$\mathbf{BL}_{kt} = \max(0, -\mathbf{Y}_{kt}) \qquad \forall\, k,t \quad (9)$$

$$\sum_{t=1}^{T} Q_{kt} \geq \sum_{t=1}^{T} \mathrm{E}[\mathbf{D}_{kt}] \qquad \forall\, k \quad (10)$$

$$Q_{kt} \geq 0 \qquad \forall\, k,t \quad (11)$$

$$O_t \geq 0 \qquad \forall\, t \quad (12)$$

$$X_{kt} \in \{0,1\} \qquad \forall\, k,t \quad (13)$$

Equations (1) - (3) denote the different objective functions. In the general conceptional model, abstract functions are given, as different metrics for each objective can be evaluated for different fields of application. Specific options are discussed in Section 5. The first objective function (1) strives for the minimization of a metric of the operational costs comprising setup costs, inventory holding costs and costs for overcapacity. To comply with the aspired target tardiness level, dynamic and implicit safety stocks are planned.[2] Since demand $\mathbf{D}_{kt}$ is uncertain, so is physical inventory $\mathbf{YP}_{kt}$. Hence, the inventory holding costs and, therefore, the total operational costs are random variables. Equation (2) specifies the objective to minimize tardiness in demand satisfaction. The degree to which this objective is achieved can be measured by a function of the backlogs $\mathbf{BL}_{kt}$, which are also random variables depending on the demand. Finally, equation (3) models the minimization of system nervousness. A function to assess system nervousness has to take into account the deviations from the selected production plan from the last decision stage. The deviations can be quantified in terms of the setup pattern $(\overline{X}_{kt} - X_{kt})$ and in terms of the production quantities $(\overline{Q}_{kt} - Q_{kt})$.

Constraints (4) - (13) ensure feasible solutions in accordance with the problem setting. Setup constraints (4) force setup operations for each product $k$ produced in period $t$.[3] Capacity constraints (5) limit the capacity available for setup operations and production. It can be extended by overcapacity, which

[2] The production of safety stocks directly influences the expected backlogs (and therefore also the expected tardiness). The model does not directly output safety stocks. However, they can be derived from the production quantities. The safety stocks of a product $k$ in a specific period $t$ is the difference between the cumulated production quantities and the cumulated expected demand for this product $k$ up to period $t$ (see Helber et al. (2013), p. 83).

[3] In the implementation of the MO-SCLSP, we replace this constraint by computationally more efficient indicator constraints, which add the same mechanisms to the model formulation.

is in turn limited by (6). The inventories for each product in each period are computed with (7). Positive values of the net inventories $\mathbf{Y}_{kt}$ are assigned to physical inventories $\mathbf{YP}_{kt}$ by (8), and negative values to backlogs $\mathbf{BL}_{kt}$ by (9). Constraints (10) preclude systematic underproduction. The domains of the decision variables are defined by (11), (12) and (13).

This conceptual model is universally applicable for arbitrary demand distribution functions and objective function metrics. In the following subsection we present modelling techniques for discrete and normally distributed demand. In Section 5, we develop a model formulation for minimizing the expected value of the operational costs, the expected mean tardiness and discrete nervousness levels.

## 3.3 Determining expected inventories and backlogs for different demand distribution functions

### *3.3.1 Impact of the demand distribution function*

As demand is uncertain, also the physical inventories and backlogs depending on the uncertain demand are uncertain. Therefore, model formulations for stochastic lot sizing with uncertain demand often take into account the expected values for the physical inventories and the expected values for the backlogs. These values depend on the underlying distribution function of the demand. The conceptional model (1) - (13) is able to deal with arbitrary demand distribution functions. In the following subsections we present examples for discrete demand and for normally distributed demand. Applying sampling approaches, arbitrary distribution functions can be represented by a discrete distribution.

### *3.3.2 Stochastic lot sizing with discrete demand*

Considering discrete scenarios is a common approach to model uncertain parameters in stochastic optimization models. Applications on stochastic lot sizing with uncertain demand can be found, e.g., in Brandimarte (2006) and Helber et al. (2013). Uncertainty can be modeled by demand scenarios if demand is inherently discrete. This can be the case if the demand of each customer order is certain but the realization of the whole customer order is uncertain. With sampling strategies, an arbitrary demand distribution can be represented by discrete demand scenarios (see, e.g., Saliby (1990)). The representation is more accurate as more scenarios are taken into account. Each demand scenario represents a trajectory of demand realizations and all scenarios occur with identical probabilities.

Helber et al. (2013) describe an application of a scenario approach on stochastic lot sizing models. Applying the scenario approach, the stochastic demand variable $\mathbf{D}_{kt}$ can be replaced by a set of deterministic demand parameters $d^s_{kt}$ describing the demand for product $k$ in period $t$ in demand scenario $s$.

The expected value of demand can then be approximated by the mean over all demand scenarios (see Equation (14)) and the variance of demand with Equation (15).

$$\mathrm{E}[\mathbf{D}_{kt}] = \sum_{s=1}^{S} \frac{d_{kt}^{s}}{|S|} \tag{14}$$

$$\mathrm{VAR}[\mathbf{D}_{kt}] = \sum_{s=1}^{S} \frac{(d_{kt}^{s} - \mathrm{E}[\mathbf{D}_{kt}])^2}{|S|} \tag{15}$$

The inventory variables depend on the demand scenario and the expected values for the physical inventories and backlogs can be approximated by the mean values over all demand scenarios. With the notation in Table 2, constraints (16) - (18) can replace (7) - (9) in the MO-SCLSP.

Table 2: Additional notation for the scenario approximation

| | |
|---|---|
| Indices and index sets: | |
| $s \in \mathscr{S}$ | scenarios |
| Parameters: | |
| $d_{kt}^{s}$ | demand of product $k$ due in period $t$ in demand scenario $s$ |
| Decision variables: | |
| $BL_{kt}^{s}$ | backlog of product $k$ in period $t$ in demand scenario $s$ |
| $Y_{kt}^{s}$ | net inventory of product $k$ in period $t$ in demand scenario $s$ |
| $YP_{kt}^{s}$ | physical inventory of product $k$ in period $t$ in demand scenario $s$ |

$$Y_{k,t-1}^{s} + Q_{kt} - Y_{kt}^{s} = d_{kt}^{s} \qquad \forall\, k,t,s \tag{16}$$

$$\mathrm{E}[\mathbf{YP}_{kt}] = \sum_{s=1}^{S} \frac{\max(0, Y_{kt}^{s})}{|S|} \qquad \forall\, k,t \tag{17}$$

$$\mathrm{E}[\mathbf{BL}_{kt}] = \sum_{s=1}^{S} \frac{\max(0, -Y_{kt}^{s})}{|S|} \qquad \forall\, k,t \tag{18}$$

#### *3.3.3 Stochastic lot sizing with normally distributed demand*

In the case of normally distributed demand, determining the expected values of physical inventories $\mathrm{E}[\mathbf{YP}_{kt}]$ and backlogs $\mathrm{E}[\mathbf{BL}_{kt}]$ requires to evaluate the first order loss function (see Tempelmeier (2011b), p. 292). Helber et al. (2013) (p. 84) show, that the SCLSP is non-linear in the case of normally distributed demand. Therefore, they propose a piecewise linearization approach to approximate the expected values for physical inventories and backlogs.

Supporting points are defined for linear segments $l$ and for the cumulated production $cp_{ktl}$ at these supporting points, the corresponding physical inventories $eyp_{ktl}$ and the corresponding expected backlogs $ebl_{ktl}$ are determined in a preprocessing step. Between the supporting points, piecewise linear functions of the expected physical inventories and the expected backlogs are assumed. With the additional notation from Table 3, the expected values can be computed with constraints (19) - (22) replacing the non-linear set of constraints (7) - (9) in the conceptual version of the MO-SCLSP.

Table 3: Additional notation for the piecewise linearization

| | |
|---|---|
| Indices and index sets: | |
| $l \in \mathscr{L}$ | linear segments |
| Parameters: | |
| $cp_{ktl}$ | cumulated production of product $k$ in period $t$ at the supporting point at the end of linear segment $l$ |
| $eyp_{ktl}$ | expected physical inventories of product $k$ in period $t$ at the supporting point at the end of linear segment $l$ |
| $ebl_{ktl}$ | expected backlogs of product $k$ in period $t$ at the supporting point at the end of linear segment $l$ |
| Decision variables: | |
| $W_{ktl}$ | cumulated production quantities of product $k$ in period $t$ assigned to linear segment $l$ |

$$Q_{kt} \geq \sum_{l=1}^{L} W_{ktl} - \sum_{l=1}^{L} W_{k,t-1,l} \qquad \forall\, k,t \quad (19)$$

$$W_{ktl} \leq cp_{ktl} - cp_{kt,l-1} \qquad \forall\, k,t,l \quad (20)$$

$$\mathrm{E}[\mathbf{YP}_{kt}] = eyp_{kt0} + \sum_{l=1}^{L} \frac{eyp_{ktl} - eyp_{kt,l-1}}{cp_{ktl} - cp_{kt,l-1}} \cdot W_{ktl} \qquad \forall\, k,t \quad (21)$$

$$\mathrm{E}[\mathbf{BL}_{kt}] = ebl_{kt0} + \sum_{l=1}^{L} \frac{ebl_{ktl} - ebl_{kt,l-1}}{cp_{ktl} - cp_{kt,l-1}} \cdot W_{ktl} \qquad \forall\, k,t \quad (22)$$

Constraints (19) link the difference in the cumulated production quantities $W_{ktl}$ between two periods $t$ and $t-1$ with the production quantities in period $t$, $Q_{kt}$. Constraints (20) limit the production quantities that can be assigned to each linear segment $l$. Based on the parameters $eyp_{ktl}$ representing the expected physical inventories at the supporting points, the expected physical inventories for the respective production quantities are determined by (21) applying piecewise linear functions with the cumulated production quantities $W_{ktl}$ as variable. Analogously, the expected backlogs are determined with (22). Note that as we determine backlogs, and not backorders, we do not need to add additional binary auxiliary variables assuring the correct assignment of the production quantities to $W_{ktl}$ (see van Pelt and Fransoo (2017)).

## 4 Solving the multi-objective stochastic lot sizing problem in multiple decision stages

### 4.1 Adapting production plans in multiple decision stages

As time passes, more accurate demand information is revealed. On the one hand, in the meantime the demand for some periods has been realized. Additionally, the demand realizations may indicate a changed customer behaviour resulting in updated demand forecasts. Based on the evolved demand information the tentative production plan might turn out to be inadequate and the decision maker might want to make production plan adjustments.

Therefore, we propose solving the multi-objective stochastic lot sizing problem in multiple decision stages, i.e., revisiting the production plan after some periods, taking newly observed demand realizations into account. In each decision stage, only a subset of decisions is fixed, while other parts of the production plan can be altered.

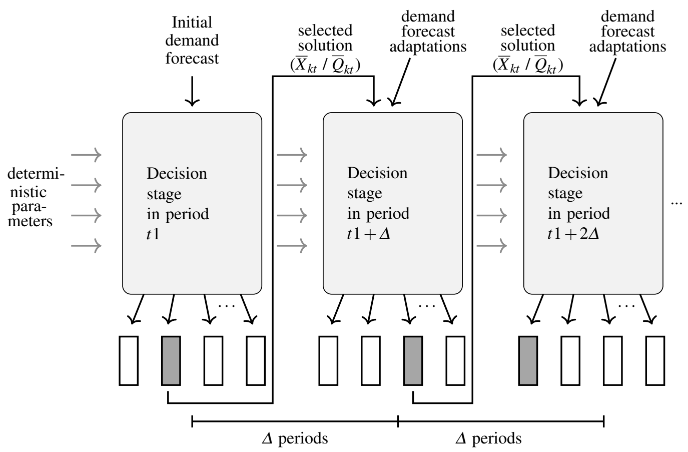



Fig. 2: Solving the MO-SCLSP in multiple decision stages

Figure 2 illustrates the process of solving the problem in multiple decision stages. It is modeled as a series of multi-objective problems, one for each decision stage. For each multi-objective problem, several Pareto-optimal solutions are determined from which one is selected by the decision maker. The corresponding production plan is executed until the next decision stage. The production plan also determines parameters of the problem for the subsequent decision stage. In the next decision stage after $\Delta$ periods, more accurate information on demand is available, and production plans can be adapted to this new demand information. In a similiar manner, the effect of uncertain production capacities or yield outcomes could be considered.

The following subsection explains how demand forecasts can evolve to make use of this additional demand information.

### 4.2 Demand forecast evolution in stochastic lot sizing with multiple decision stages

Demand is modeled as an independent random variable with neither cross-correlation between different products nor autocorrelation between different demand periods for the same product. Demand information is revealed gradually over the course of multiple periods, continuously reducing uncertainty as time advances. These demand forecast updates can be modeled in the framework of the additive Martingale Model of Forecast Evolution (MMFE) (see Heath and Jackson (1994)). Uncertainty is modeled by random adjustment steps of the expected values over the course of multiple periods. Starting with an initial forecast $\hat{d}_{kt}$ of the demand for product $k$ in period $t$, forecasts are updated over the course of $\theta$ periods. Thus, the first observation is made in period $\tau = t - \theta + 1$. The adjustment of the forecast of the demand for product $k$ in period $t$ made in period $\tau$ is modeled by the random step $\boldsymbol{\psi}_{kt}^{\tau}$ with expected value $\mathrm{E}[\boldsymbol{\psi}_{kt}^{\tau}] = 0$ and variance $\mathrm{VAR}[\boldsymbol{\psi}_{kt}^{\tau}]$. Figure 3 exemplarily illustrates the demand forecast evolution.

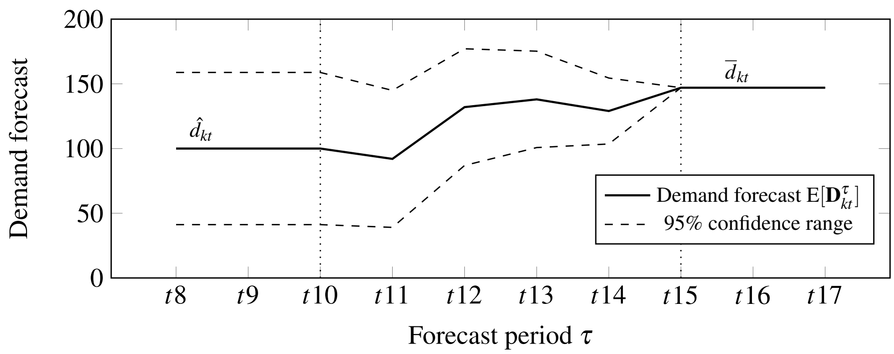



Fig. 3: Demand forecast adjustments over the course of 5 periods

In this example, the initial forecast for the demand in period 14 for a specific product is assumed to be $\hat{d}_{k,14} = 100$. Additional demand information is observed over the course of $\theta = 5$ periods, beginning with period $\tau = t14 - 4 = t10$. In each period, the demand forecast is adjusted by a random step $\boldsymbol{\psi}_{kt}^{\tau}$ to account for the newly observed demand information. These steps of demand forecast evolution are modeled as random variables with expected values of $\mathrm{E}[\boldsymbol{\psi}_{kt}^{\tau}] = 0$ and a variance of $\mathrm{VAR}[\boldsymbol{\psi}_{kt}^{\tau}] = 180$, $t - \theta < \tau \leq t$. Therefore, the closer to the respective demand period they are, the more precise are the forecasts for the demand based on the information in period $\tau$, $\mathbf{D}_{kt}^{\tau}$. As time advances, expected value $\mathrm{E}[\mathbf{D}_{kt}^{\tau}]$ approaches the actual demand realization $\overline{d}_{kt}$. Total variance $\text{VAR}[\mathbf{D}_{kt}^{\tau}] = \sum_{\tau=t-\theta+1}^{t} \text{VAR}[\boldsymbol{\psi}_{kt}^{\tau}]$ reduces over time, as the steps from fewer periods contribute to the uncertainty.

The random variables describing the forecasts of demand for product $k$ in period $t$ over the course of the forecast periods $\tau$ follow an AR(1) process, also known as a random walk (see Spitzer (1976)). The distribution of each step $\boldsymbol{\psi}_{kt}^{\tau}$ depends on the distribution of the total demand $\mathbf{D}_{kt}$. In the case of discrete demand distributions, the steps $\boldsymbol{\psi}_{kt}^{\tau}$ are discretely distributed as well. Analogously, the steps are normally distributed if the total demand is normally distributed.

Demand forecasts $\mathbf{D}_{kt}$ can be expressed as the sum of initial forecast $\hat{d}_{kt}$ and random steps $\boldsymbol{\psi}_{kt}^{\tau}$, as shown in Equation (23).

$$\mathbf{D}_{kt} = \hat{d}_{kt} + \sum_{\tau=t-\theta+1}^{t} \boldsymbol{\psi}_{kt}^{\tau} \tag{23}$$

As the expected values of all steps $\boldsymbol{\psi}_{kt}^{\tau}$ are $\text{E}[\boldsymbol{\psi}_{kt}^{\tau}] = 0$, the expected value $\text{E}[\mathbf{D}_{kt}]$ equals $\hat{d}_{kt}$, as shown in Equation (24). The variance $\text{VAR}[\mathbf{D}_{kt}]$ can be computed following Equation (25).

$$\text{E}[\mathbf{D}_{kt}] = \hat{d}_{kt} + \sum_{\tau=t-\theta+1}^{t} \text{E}[\boldsymbol{\psi}_{kt}^{\tau}] = \hat{d}_{kt} \tag{24}$$

$$\text{VAR}[\mathbf{D}_{kt}] = \sum_{\tau=t-\theta+1}^{t} \text{VAR}[\boldsymbol{\psi}_{kt}^{\tau}] \tag{25}$$

Assuming the final demand realizations are revealed at the end of the respective period, in a decision stage in period $p$ the demand realization for all periods $t < p$ are already known. Furthermore, based on the increased market observations, the demand forecast can be adjusted. In the current period $p$, the realizations $\overline{\psi}_{kt}^{\tau}$ for the steps in periods $\tau < p$ are already known. Based on this information, the forecasts can be updated. The current demand forecast in period $p$ is denoted by $\mathbf{D}_{kt}^{p}$. Equation (26) shows that expected value $\text{E}[\mathbf{D}_{kt}^{p}]$ is shifted by the already realized steps $\overline{\psi}_{kt}^{\tau}$, and Equation (27) shows how variance $\text{VAR}[\mathbf{D}_{kt}^{p}]$ reduces as the number of steps contributing to the uncertainty reduces.

$$\text{E}[\mathbf{D}_{kt}^{p}] = \hat{d}_{kt} + \underbrace{\sum_{\tau=t-\theta+1}^{p-1} \overline{\psi}_{kt}^{\tau}}_{\text{observed part}} + \underbrace{\sum_{\tau=p}^{t} \text{E}[\boldsymbol{\psi}_{kt}^{\tau}]}_{\text{unknown part}} = \hat{d}_{kt} + \sum_{\tau=t-\theta+1}^{p-1} \overline{\psi}_{kt}^{\tau} \tag{26}$$

$$\text{VAR}[\mathbf{D}_{kt}^{p}] = \sum_{\tau=p}^{t} \text{VAR}[\boldsymbol{\psi}_{kt}^{\tau}] \tag{27}$$

Modeling demand forecast updates allows for utilizing the already observed realization information for updates of the production plans and for making the final decision under a lower level of uncertainty. This reduces the need for safety stocks. As a result, the production plans are better adapted to the actual demand situation. The proposed modeling approach is quite flexible. The number of periods in which forecast adjustments are made can be chosen according to the requirements of a specific application based on the nature of the demand realization process.

## 5 A model formulation for re-optimizing the production plan in multiple decision stages

### 5.1 Modeling tardiness by computing the disaggregated mean expected tardiness

High delivery reliability is achieved by satisfying customer demand as close to the demand period as possible, thus limiting backlogs and ensuring low tardiness. Although the minimization of total backlogs would lead to minimal tardiness, often, relative service level formulations are evaluated, which allow for an interpretation of the results (see for example Gruson et al. (2018) for an overview). To avoid the disadvantages of those service level formulations, as pointed out in, for example, Tempelmeier (2011b), in this research, the mean expected tardiness is determined instead, as this criterion portrays a clear and comparable interpretation of the results. The mean expected tardiness $\mathrm{E}[MT]$ equals the mean expected waiting time of demand to be satisfied. Assuming that long-run production meets long-run demand, the mean expected demand has to equal the mean throughput of the production system. Applying Little's Law (see Little (1961)) the mean tardiness can then be determined as the quotient of the total expected backlogs and the total expected demand. With the additional notation in Table 4, Equation (28) defines the mean expected tardiness for a given product $k$.[4]

Table 4: Additional notation for calculating mean expected tardiness

| | |
|---|---|
| Random variables: | |
| $\mathbf{MT}_k$ | product-specific mean tardiness of demand satisfaction |
| Decision variables: | |
| $MT^{agg}$ | aggregated mean tardiness of demand satisfaction |
| $MT^{dis}$ | disaggregated mean tardiness of demand satisfaction |
| Parameters: | |
| $\zeta$ | period lenght |

$$\mathrm{E}[\mathbf{MT}_k] = \frac{\sum_{\tau=1}^{T} \mathrm{E}[\mathbf{BL}_{k\tau}]}{\sum_{\tau=1}^{T} \mathrm{E}[\mathbf{D}_{k\tau}]} \cdot \zeta \tag{28}$$

[4] Section 3.3 provides examples of determining $\mathrm{E}[\mathbf{BL}_{k\tau}]$ for different demand distribution functions

In multi-product problem instances, either the aggregated mean tardiness $MT^{agg}$ or the disaggregated mean tardiness $MT^{dis}$ can be taken into account. In an aggregated formulation, the mean tardiness is averaged over all considered products (see Sereshti et al. (2020)). This allows for the compensation of a particularly bad mean tardiness of a specific product with a particularly good mean tardiness of another product. In a disaggregated formulation, in contrast, the goal to reach an aspired mean tardiness is formulated for each product individually.

$$MT^{dis} \geq \mathrm{E}[\mathbf{MT}_k] \qquad \forall\, k \quad (29)$$

To implement the disaggregated mean tardiness $MT^{dis}$, it is necessary to determine the product-specific expected value of the mean tardiness $\mathrm{E}[\mathbf{MT}_k]$ with Constraints (28) as a first step. Constraints (29) couple $MT^{dis}$ with the largest product-specific expected value of mean tardiness $\mathrm{E}[\mathbf{MT}_k]$, thus allowing for the direct minimization of $MT^{dis}$.

### 5.2 Modeling system nervousness via discrete nervousness levels

Nervousness occurs when an already decided production plan is altered after demand realizations have been observed. Nervousness can stem from adjustments of the setup pattern or the lot sizes (see Tunc et al. (2013)). Particularly in supply chains, those adjustments should be limited to avoid the propagation of nervousness to upstream production resources.

While operational costs and mean tardiness can be measured, interpreted and compared on a ratio scale (see Stevens (1946)), nervousness caused by different adjustments of the production plans is more difficult to assess. Measuring system nervousness on a ratio scale would require the definition of weights to assess the significance of nervousness caused by different adjustments of the production plans to calculate some kind of nervousness metric. This would allow for flexible adjustments, but defining those weights can hardly be done in practice.

Therefore, we propose the modeling of different discrete nervousness levels $n$ on an ordinal scale. These nervousness levels define characteristics describing to which extend the production plan can be adjusted and therefore represent both the accepted nervousness and the production plan flexibility. The nervousness level may restrict the following criteria:

- frozen horizon for the setup pattern fixation
- frozen horizon for the lot size fixation
- maximum allowed relative/absolute adjustment of planned lot sizes
- maximum allowed additional/cancelled/shifted setups per product
- maximum postponement/preponement of a scheduled setup activity

Which criteria the nervousness level restrict should be decided based on the specific field of application and on the economic consequences of the respective criteria. For this study we assume that the nervousness levels determine the frozen horizons for both the setup pattern fixation and the lot size fixation. These nervousness levels are chosen in such a way, that the lower the nervousness level is, the higher is the number of periods for which the decisions regarding the production plan are preserved. With the notation presented in Table 5, this kind of nervousness levels can be integrated in a linear model formulation with Constraints (30) - (34), forcing production plan fixations depending on the chosen nervousness level $N^*$.

Table 5: Additional notation for modeling discrete nervousness levels

| | |
|---|---|
| Indices and index sets: | |
| $n \in \mathcal{N}$ | Discrete nervousness levels |
| $p \in \mathcal{T}$ | Current period in respective decision stage |
| $\Delta$ | Re-optimization interval |
| Parameters: | |
| $h_n^{fix_Q}$ | frozen horizon for the lot size fixation in nervousness level $n$ |
| $h_n^{fix_X}$ | frozen horizon for the setup pattern fixation in nervousness level $n$ |
| Decision variables: | |
| $N^* \in \mathbb{N}_0$ | Chosen nervousness level |
| $V_n \in \{0,1\}$ | Binary auxiliary variable: 1 if $N^* \geq n$ |

$$X_{kt} \leq \overline{X}_{kt} + V_n \qquad \forall\, k,n,t < p-\Delta+h_n^{fix_X} \tag{30}$$

$$X_{kt} \geq \overline{X}_{kt} - V_n \qquad \forall\, k,n,t < p-\Delta+h_n^{fix_X} \tag{31}$$

$$Q_{kt} \leq \overline{Q}_{kt} + V_n \cdot M \qquad \forall\, k,n,t < p-\Delta+h_n^{fix_Q} \tag{32}$$

$$Q_{kt} \geq \overline{Q}_{kt} - V_n \cdot M \qquad \forall\, k,n,t < p-\Delta+h_n^{fix_Q} \tag{33}$$

$$N^* = \sum_{n=1}^{N} V_n \tag{34}$$

The fixation of the setup pattern is implemented with Constraints (30) and (31). The production times are fixed according to the corresponding nervousness level $n$ for $h_n^{fix_X}$ periods on the values from the selected production plan from the last decision stage $\overline{X}_{kt}$. In a decision stage at the beginning of period $p$, the previous decision was made in period $p-\Delta$. Thus, the frozen horizon begins with period $p-\Delta$ and ends with $p-\Delta+h_n^{fix_X}-1$. If the setup decision $X_{kt}$ deviates from the decision from the last decision stage $\overline{X}_{kt}$ in a period controlled by nervousness level $n$, then the auxiliary variable $V_n$ has to be 1 to obtain a feasible solution.

Analogously, Constraints (32) and (33) fix the production quantities $Q_{kt}$ for $h_n^{fixQ}$ periods on the values from the selected production plan from the previous decision stage $\overline{Q}_{kt}$, unless auxiliary variable $V_n$ is set to 1. Provided that $V_n$ is set to 1, Constraints (32) and (33) do not restrict possible values of $Q_{kt}$ due to the multiplication of $V_n$ with a sufficiently large number $M$. Finally, Constraint (34) determines the chosen nervousness level $N^*$ as the sum of the auxiliary variables $V_n$.

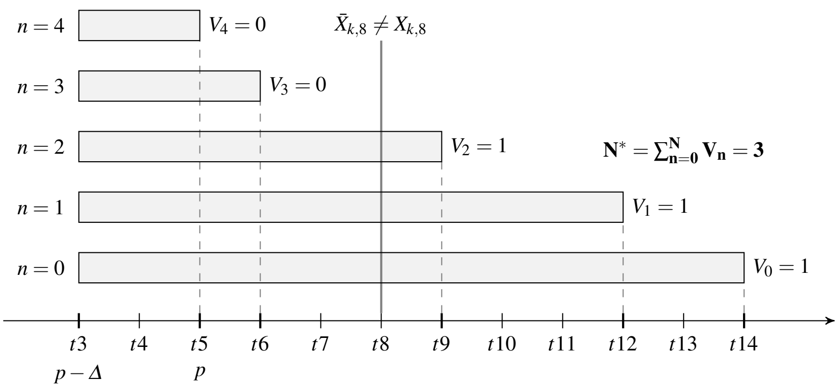



Fig. 4: Example of how the selected nervousness level $N^*$ is computed with Constraints (30) - (34)

Figure 4 illustrates the principle of Constraint (34). For the sake of simplicity the presented example is limited to adjustments of the setup pattern. In this example, in the decision stage at the beginning of period 5 an additional setup in period 8 shall be added. In the figure, gray bars depict the fixation of the setup pattern starting in period $p-\Delta=3$ for different nervousness levels $n$. The grey bars show that nervousness levels 0, 1 and 2 force fixations in this period. Therefore, the corresponding auxiliary variables $V_n$ have to be set to 1. However, nervousness levels 3 and 4 allow for adjustments in this period. Thus, for those nervousness levels, $V_3=V_4=0$ is feasible. The chosen nervousness level can be determined according to Constraint (34) as $N^*=\sum_{n=0}^{N}V_n=3$. Thus, nervousness level 3 corresponds to this adjustments.

Special cases of the proposed nervousness levels include the well-known uncertainty strategies proposed by Bookbinder and Tan (1988). By fixing both the setup pattern and the production quantities for the whole planning horizon ($h^{fixQ}=h^{fixX}=T$) after the first decision stage, the static uncertainty strategy can be depicted. Fixing the setup pattern for the whole planning horizon ($h^{fixX}=T$) but the production quantities only until the next decision stage ($h^{fixQ}=\Delta$) represents the static-dynamic uncertainty strategy. Finally, the dynamic uncertainty strategy can be rebuild by fixing all decision variables only until the next decision stage ($h^{fixQ}=h^{fixX}=\Delta$). Therefore, the proposed discrete nervousness levels can be understood as a generalization of the classification proposed by Bookbinder and Tan (1988).

### 5.3 Model formulation for re-optimizing production plans in a specific decision stage

Based on the modeling concepts presented in the previous subsections, we now present the model formulation for the re-optimization of the production plan in a specific decision stage. Tardiness is modeled by the disaggregated mean expected tardiness and nervousness levels by discrete nervousness levels. We therefore refer to this model as MO-SCLSP$^{\text{stage}}_{\text{MT},N^*}$. The setup pattern and the production quantities are fixed according to the selected nervousness level. Beginning with the current period $p$, the look-ahead in each optimization stage is limited to the next $h^{look}$ periods, as the cumulated uncertainty for later periods is still high and decisions for those periods are not yet necessary. Limiting the number of periods to $h^{look}$ therefore reduces the problem size and therefore the computational effort. Figure 5 illustrates the sub-horizons in each decision stage.

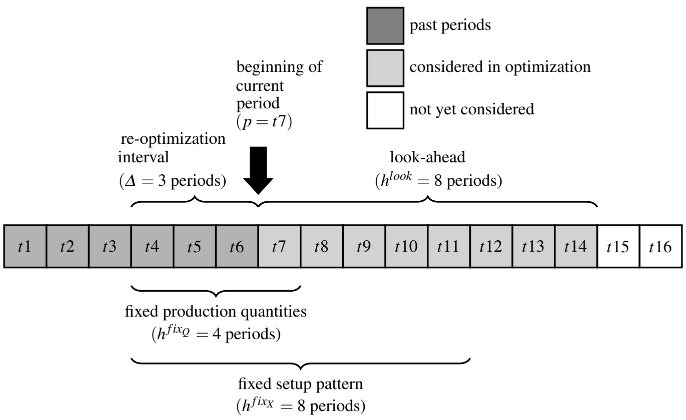



Fig. 5: Dividing the planning horizon in subhorizons

**MO-SCLSP$^{\text{stage}}_{\text{MT},N^*}$**

$$\min \; \{\mathrm{E}[COST],\, MT^{dis},\, N^*\} \tag{35}$$

s.t.

$$\mathrm{E}[COST] = \sum_{k=1}^{K} \sum_{t=1}^{p+h^{look}-1} \Big[ sc_k \cdot X_{kt} + hc_k \cdot \mathrm{E}[\mathbf{YP}_{kt}] \Big] + \sum_{t=1}^{p+h^{look}-1} oc \cdot O_t \tag{36}$$

$$\mathrm{E}[\mathbf{MT}_k] = \frac{\sum_{t=1}^{p+h^{look}-1} \mathrm{E}[\mathbf{BL}_{kt}]}{\sum_{t=1}^{p+h^{look}-1} \mathrm{E}[\mathbf{D}_{kt}]} \cdot \zeta \qquad \forall\, k \quad (37)$$

$$MT^{dis} \geq \mathrm{E}[\mathbf{MT}_k] \qquad \forall\, k \quad (38)$$

$$N^* = \sum_{n=1}^{N} V_n \qquad (39)$$

$$Q_{kt} \leq M \cdot X_{kt} \qquad \forall\, k, t < p + h^{look} \quad (40)$$

$$\sum_{k=1}^{K} \left[ tp_k \cdot Q_{kt} + ts_k \cdot X_{kt} \right] \leq b_t + O_t \qquad \forall\, t < p + h^{look} \quad (41)$$

$$O_t \leq o_t^{max} \qquad \forall\, t < p + h^{look} \quad (42)$$

$$\sum_{t=1}^{p+h^{look}-1} Q_{kt} \geq \sum_{t=1}^{p+h^{look}-1} \mathrm{E}[\mathbf{D}_{kt}] \qquad \forall\, k \quad (43)$$

$$\mathbf{Y}_{k,t-1} + Q_{kt} - \mathbf{Y}_{kt} = \mathbf{D}_{kt} \qquad \forall\, k, t \quad (44)$$

$$\mathbf{YP}_{kt} = \max(0, \mathbf{Y}_{kt}) \qquad \forall\, k, t \quad (45)$$

$$\mathbf{BL}_{kt} = \max(0, -\mathbf{Y}_{kt}) \qquad \forall\, k, t \quad (46)$$

$$X_{kt} \leq \overline{X}_{kt} + V_n \qquad \forall\, k, n, t < p - \Delta + h_n^{fixX} \quad (47)$$

$$X_{kt} \geq \overline{X}_{kt} - V_n \qquad \forall\, k, n, t < p - \Delta + h_n^{fixX} \quad (48)$$

$$Q_{kt} \leq \overline{Q}_{kt} + V_n \cdot M \qquad \forall\, k, n, t < p - \Delta + h_n^{fixQ} \quad (49)$$

$$Q_{kt} \geq \overline{Q}_{kt} - V_n \cdot M \qquad \forall\, k, n, t < p - \Delta + h_n^{fixQ} \quad (50)$$

$$Q_{kt}, W_{ktl}, O_t \geq 0 \qquad \forall\, k, t < p + h^{look}, l \quad (51)$$

$$X_{kt} \in \{0; 1\} \qquad \forall\, k, t < p + h^{look} \quad (52)$$

The multi-criteria objective function (35) aims at minimizing the expected operational costs, the disaggregate mean tardiness and the discrete nervousness level. The expected operational costs, comprising holding costs, setup costs and overcapacity costs, are determined in (36) while limiting the look-ahead to periods $t < p + h^{look}$. The determined value incorporates already realized costs for periods $t < p$ and expected costs for periods $p \leq t < p + h^{look}$. The product-specific expected mean tardiness is derived in (37) from the expected backlogs and the expected demand. Constraints (38) couple the highest product-specific expected mean tardiness with the disaggregate mean tardiness $MT^{dis}$ to prevent balancing effects between different products. Finally, (39) determines the discrete nervousness level $N^*$ as the sum of the auxiliary variables $V_n$.

Constraints (40) - (43) ensure feasible solutions with respect to the problem setting described in Subsection 3.1. Setup operations are forced for each product $k$ produced in period $t$ with positive production $Q_{kt}$ via constraint (40). The production capacity available for production and setup activities is limited by capacity constraint (41). Overcapacity is limited by (42). Constraints (43) prevent systematic underproduction. Net inventories are calculated with (44) and assigned to the corresponding variables for the physical inventories and backlogs by (45) and (46), respectively.

Finally, Constraints (47) - (50) control the accepted system nervousness depending on the accepted nervousness level $N^*$, as presented in Subsection 5.2. The constraints ensure production plan stability in compliance with the nervousness levels $n \geq N^*$.

The presented model is linear and has tangible objective functions. In the following section, an interactive approach for deriving Pareto-optimal solution vectors for the multi-objective problem is proposed.

## 6 An interactive framework for determining Pareto-optimal solutions for a specific decision stage of the MO-SCLSP

### 6.1 Advantages of an interactive approach for multi-objective lot-sizing

The multi-objective optimization problems in each decision stage are solved with an interactive approach (see Luque et al. (2011)): The decision maker interactively specifies areas of the objective space that he or she wants to explore and decides how new Pareto-optimal solutions shall be determined. With each additional computed set of solutions, the boundaries of the objective space become clearer. The decision maker makes use of this information to iteratively refine his or her search for Pareto-optimal solutions until he or she is eventually able to select a final solution. Compared to defining aspiration levels *a priori*, the proposed interactive approach has three major advantages:

- The Pareto front gives a clear outline of the objective space. Thus, the decision maker recognizes which combinations of objective function values are attainable.
- The Pareto front describes details of the conflicting relationships between the objective functions, giving a clear understanding of the consequences of claiming a certain aspiration level.
- The interactive approach allows to quickly find the area of the objective space in which the decision maker is interested, thus concentrating the computational effort on the relevant area.

In the proposed interactive approach, the determination of Pareto-optimal ideal-points is combined with the determination of further Pareto-optimal solutions with the augmented $\varepsilon$-constraint method and controlled by the decision maker according to his or her preferences. Figure 6 gives an overview of the proposed interactive approach for multi-objective lot sizing in multiple decision stages.

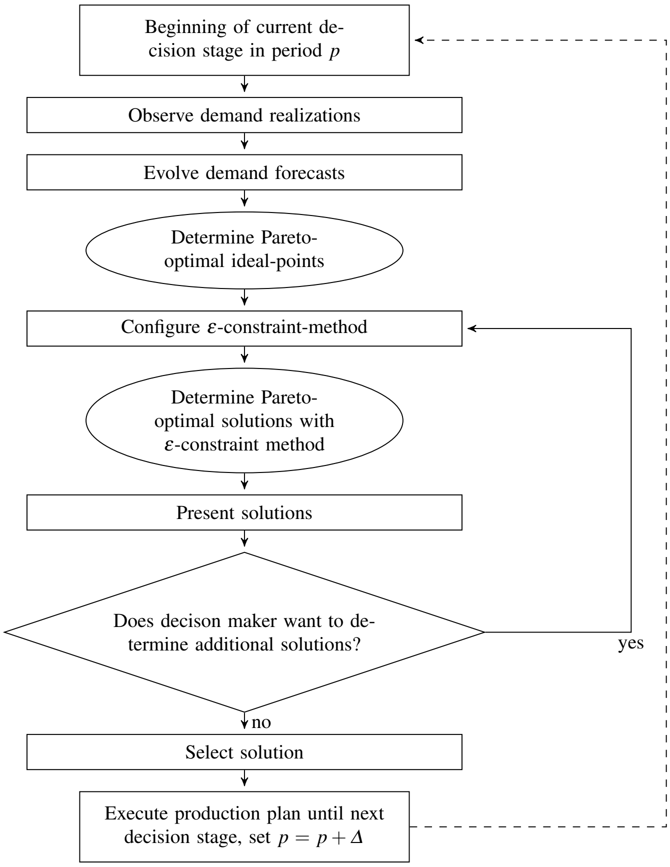



Fig. 6: An interactive approach for selecting a Pareto-optimal solution in a specific decision stage of the MO-SCLSP

In each decision stage, at the beginning of the respective period $p$ new demand information is observed. Based on this new information the demand forecasts are updated. Then, bounds of the Pareto-optimal space are determined by determining Pareto-optimal ideal-points (see Subsection 6.2). Based on the ideal-points the decision maker can configure the $\varepsilon$-constraint-method to determine further Pareto-optimal solutions based on his or her preferences (see Subsection 6.3). The determined objective vectors, constituting an outline of the Pareto front, are presented to the decision maker. With this impression of the objective space and the dependencies between the objectives, the decision maker can refine the search for Pareto-optimal solutions. This process is applied iteratively until the decision maker decides to select a final solution.

### 6.2 Determining Pareto-optimal ideal-points

The idea behind determining ideal-points is to find lower bounds of the objective space and upper bounds for the objective function values of potential Pareto-optimal solutions. Figure 7 illustrates the following explanations.

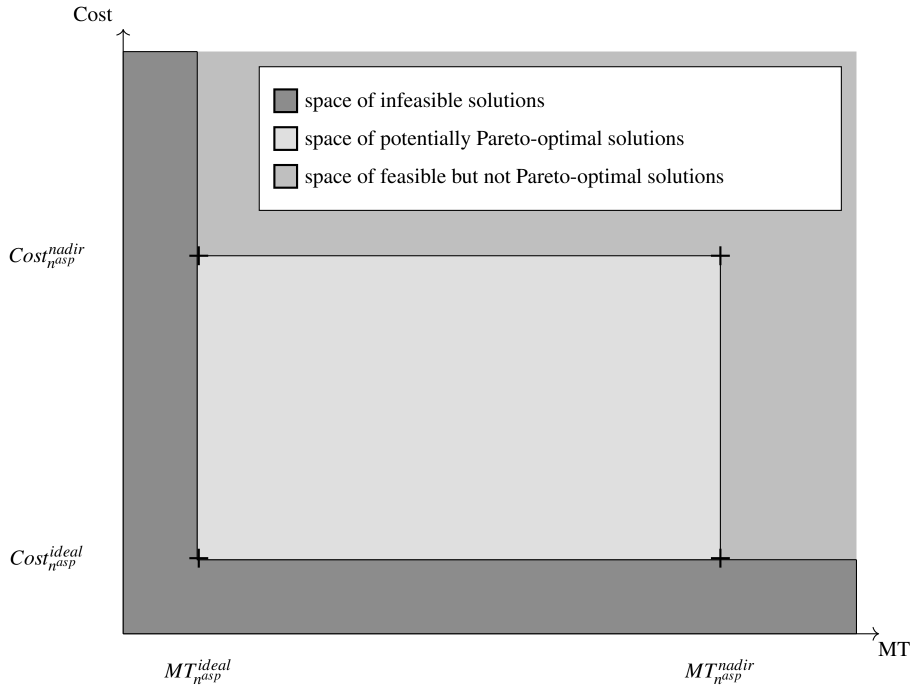



Fig. 7: ideal-points framing the bounds of the objective function values of Pareto-optimal solutions

In general, an *ideal objective function value* $Z^{ideal}$ is defined as the optimal value of an objective function $Z$ of a single-objective counterpart of a multi-objective problem, neglecting all other objective functions except for $Z$ (see Miettinen (1999)). Therefore, no solution can exist with any value for $Z$ better than $Z^{ideal}$, and the corresponding area of the objective space does not contain feasible solutions.

The optimal objective function value for an objective function $Y$ of a bi-objective problem that can be obtained under the condition that the other objective function $Z$ attains its ideal value $Z^{ideal}$ is called *nadir value* $Y^{nadir}$ (see Deb (2001)). Thus, any solution containing a value of $Y$ worse than $Y^{nadir}$ can not be Pareto-optimal, as $Y$ can be improved without any deterioration of the other objective function values.

In multi-objective stochastic lot sizing, keeping the current production plan from the previous stage of sequential decision-making is always feasible (this solution might incur particularly bad values for the operational costs and the mean tardiness, though). Thus, a feasible solution without nervousness can always be found, and the ideal nervousness level is always $n^{ideal} = 0$. Therefore, we rather compute the ideal-points for bi-objective cross-sections of the objective space with defined aspired nervousness level $n^{asp}$.

The ideal-points of the remaining bi-objective problem can be determined with lexicographic optimization (see Isermann (1982), pp. 223 and Ehrgott (2005), pp. 128). To determine $COST_{n^{asp}}^{ideal}$, in a first step, the operational costs are optimized accepting arbitrary values of mean tardiness. $MT_{n^{asp}}^{nadir}$ is determined in a second step by solving the model again, subject to the condition that the operational costs take their ideal value ($\mathrm{E}[COST] = COST_{n^{asp}}^{ideal}$). In (53), the lexicographic optimization is expressed by the operator *lexmin*, and the order of the arguments expresses the hierarchy of the objective functions (see Ehrgott (2005), p. 129).

**MO-SCLSP$_{\mathbf{COST^{ideal}},\mathbf{n^{asp}}}$**

$$\mathrm{lexmin}(\mathrm{E}[COST],\ MT^{dis}) \tag{53}$$

s.t. (36) - (38), (40) - (46), (51), (52) and

$$X_{kt} = \overline{X}_{kt} \qquad \forall\ k, t < p - \Delta + h_{n^{asp}}^{fix_X} \tag{54}$$

$$Q_{kt} = \overline{Q}_{kt} \qquad \forall\ k, t < p - \Delta + h_{n^{asp}}^{fix_Q} \tag{55}$$

Analogously, solving the following bi-objective problem with lexicographic optimization and inverse hierarchical order obtains a Pareto-optimal solution with objective function values $MT_{n^{asp}}^{ideal}$ and $COST_{n^{asp}}^{nadir}$.

**MO-SCLSP$_{\mathbf{MT^{ideal}},\mathbf{n^{asp}}}$**

$$\mathrm{lexmin}(MT^{dis},\ \mathrm{E}[COST]) \tag{56}$$

s.t. (36) - (38), (40) - (46), (51), (52), (54) and (55)

As the nadir-values for the non-prioritized objective functions usually are unacceptably bad, the ideal-points probably do not contain desired compromise solutions. However, determining the ideal-values and nadir-values reveals the bounds of the space containing potentially Pareto-optimal solutions. Only solutions with objective functions values of $Z$ better than $Z^{nadir}$ and worse than $Z^{ideal}$ are candidates for further Pareto-optimal solutions (see Figure 7). This information can be applied to narrow down the search for Pareto-optimal solutions with the multidimensional $\varepsilon$-constraint method, which is presented in the following subsection.

## 6.3 Determining Pareto-optimal solutions with an augmented multidimensional $\varepsilon$-constraint method

The augmented $\varepsilon$-constraint method (see Chankong and Haimes (1983), Haimes et al. (1971) and Mavrotas (2009)) can be applied to determine a set of Pareto-optimal solutions for a multi-objective optimization problem by solving a sequence of constrained single-objective problems. All objective functions but one are transferred into constraints, assuring aspiration levels specified by the decision maker. Different combinations of aspiration levels are systematically inspected and the resulting single-objective problem is solved. Then the aspiration level is adjusted by a fixed increment $\varepsilon$. The combination of the optimized objective function value and the set aspiration levels for the objective functions transferred into constraints constitutes a Pareto-optimal solution.

For the application in multi-objective stochastic lot sizing, we choose to minimize expected operational costs, limiting both mean expected tardiness and nervousness with constraints. The expected operational costs are optimized under the condition that the mean expected tardiness does not exceed the aspiration level $MT^{asp}$ and that nervousness does not exceed the aspired level $n^{asp}$. Applying the notation in Table 6, this results in the following model:

Table 6: Additional notation for the augmented $\varepsilon$-constraint method

| | |
|---|---|
| Parameters: | |
| $\varepsilon$ | step size for determining Pareto-optimal solutions with the $\varepsilon$-constraint method |
| $\rho_{MT}$ | small value |
| Decision variables: | |
| $S_{MT}$ | slack of the constraint limiting $MT^{dis}$ |

**MO-SCLSP**$_{\varepsilon-aug}$

$$\min \ \mathrm{E}[COST] = \sum_{k=1}^{K} \sum_{t=1}^{p+h^{look}-1} \Big[ sc_k \cdot X_{kt} + hc_k \cdot \mathrm{E}[\mathbf{YP}_{kt}] \Big] + \sum_{t=1}^{p+h^{look}-1} oc \cdot O_t - \rho_{MT} \cdot S_{MT} \quad (57)$$

s.t. (37), (38), (40) - (46), (51), (52) and

$$MT^{dis} \leq MT^{asp} - S_{MT} \quad (58)$$

$$X_{kt} = \overline{X}_{kt} \qquad \forall\, k, t < p - \Delta + h^{fix_X}_{n^{asp}} \quad (59)$$

$$Q_{kt} = \overline{Q}_{kt} \qquad \forall\, k, t < p - \Delta + h^{fix_Q}_{n^{asp}} \quad (60)$$

$$S_{MT} \geq 0 \quad (61)$$

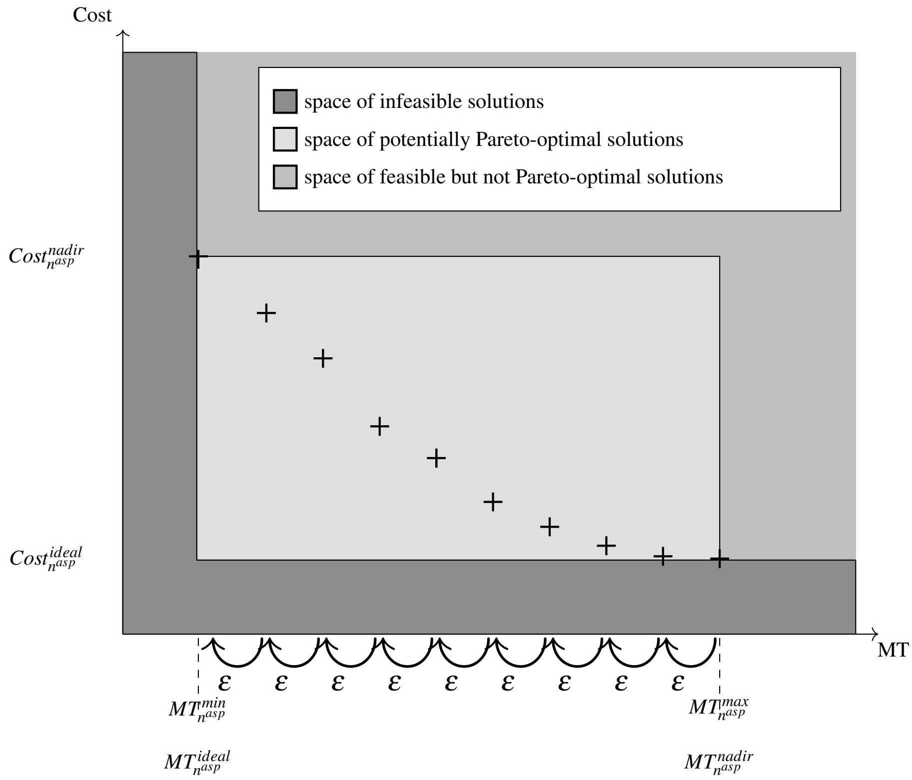



Fig. 8: Finding Pareto-optimal solutions in a bi-objective cross-section with the $\varepsilon$-constraint method

In Constraint (58), variable $S_{MT}$ is the slack measuring by how much the expected mean tardiness $MT^{dis}$ remains under the aspired mean tardiness variable $MT^{asp}$. The augmentation term (see Mavrotas (2009)) $\rho_{MT} \cdot S_{MT}$ is subtracted from the objective function value. If an aspired level of mean tardiness $MT^{asp}$ can be undercut without an increase of expected operational costs, then this term leads to the highest possible value of $S_{MT}$, thus ensuring Pareto-optimality of the found solution. The parameter $\rho_{MT}$ should be chosen small enough not to influence the operational costs in the optimal solution.

Algorithm 1 shows how Pareto-optimal solutions can be generated by solving this model for different values of $MT^{asp}$ and $n^{asp}$. An example of solutions generated for a bi-objective cross-section with given nervousness level $n^{asp}$ is presented in Figure 8.

The decision maker chooses a list of aspiration levels of nervousness $\hat{N} \in \mathscr{N}$, a range of values of aspired mean expected tardiness defined by $MT_{n^{asp}}^{min}$ and $MT_{n^{asp}}^{max}$ for each aspiration level of nervousness and an increment $\varepsilon$. If the ideal-points are known, then $MT_{n^{asp}}^{min} = MT_{n^{asp}}^{ideal}$ and $MT_{n^{asp}}^{max} = MT_{n^{asp}}^{nadir}$ constitute an appropriate choice for the inspected range, as it covers the whole space

**Algorithm 1:** Augmented $\varepsilon$-constraint method for multi-objective stochastic lot sizing

```
Let Cost(MT^asp, n^asp) be the optimal operational costs of the single-objective problem with
  mean expected tardiness bound MT^asp and limiting nervousness to level n^asp;
Choose increment ε;
Set objective space ω = ∅;
for each n ∈ N̂ do
    set n^asp = n;
    set MT^asp = MT^max_{n^asp};
    while MT^asp > MT^min_{n^asp} do
        determine x = Cost(MT^asp, n^asp);
        set ω = ω ∪ {x};
        set MT^asp = MT^asp − ε;
return ω
```

of potentially Pareto-optimal solutions, and for any given aspiration level in this range, a feasible solution can be found. For each chosen aspiration level of nervousness $n^{asp} \in \hat{N}$, the following loop is executed (see Algorithm 1): The first single-objective problem for each $n^{asp}$ is solved calling for an aspired mean expected tardiness of $MT^{asp} = MT^{max}_{n^{asp}}$. Iteratively, $MT^{asp}$ is reduced by $\varepsilon$, and the corresponding single-objective optimization problem is solved. As soon as the adjusted $MT^{asp}$ falls below $MT^{min}_{n^{asp}}$, the iteration ends, and the next iteration for the next $n^{asp} \in \hat{N}$ begins.

By applying the multidimensional $\varepsilon$-constraint method, e.g. iteratively solving instances of the model MO-SCLSP$_{\varepsilon-aug}$, a set of Pareto-optimal solutions is found for each inspected nervousness level with a fixed distance between two aspired values of mean expected tardiness. Thereby, the maximum diversity of the found solutions is achieved.

## 7 Numerical studies

### 7.1 Interaction between decision maker and planning algorithm in a single exemplary decision stage

*Experimental setup*

The following example shall demonstrate how the decision maker could interact with the planning algorithm. A small problem instance with $K = 2$ products and a look-ahead horizon of $h^{look} = 15$ periods is chosen, which allows for the generation of a large number of Pareto-optimal solutions and for the thorough studying of the solutions. The instance generator parameters presented in Table 7 are used to randomly generate problem instances.

Normally distributed demand is assumed and demand information is observed evenly over a time span of $\theta = 5$ periods. Piecewise linearization of the inventory functions is performed with 11 supporting points resulting in 10 linear segments.

Table 7: Instance generator parameterization

| | | | | |
|---|---|---|---|---|
| Expected value of demand | $\mathrm{E}[\mathbf{D}_{kt}]$ | 100 | $\cdot\, \mathscr{U}\{0.9, 1.1\}$ | QU |
| Coefficient of variation of demand | $\mathrm{VC}[\mathbf{D}_{kt}]$ | 0.3 | | |
| Holding cost rate | $hc_k$ | 1 | $\cdot\, \mathscr{U}\{0.9, 1.1\}$ | MU / (QU · period) |
| Setup cost rate | $sc_k$ | 450 | $\cdot\, \mathscr{U}\{0.9, 1.1\}$ | MU / setup |
| Processing time | $tp_k$ | 1 | $\cdot\, \mathscr{U}\{0.9, 1.1\}$ | CU |
| Setup time | $ts_k$ | 40 | $\cdot\, \mathscr{U}\{0.9, 1.1\}$ | CU |
| Regular capacity | $b_t$ | 350 | $\cdot\, \mathscr{U}\{0.9, 1.1\}$ | CU |
| Maximum overcapacity | $o^{max}$ | 100 | $\cdot\, \mathscr{U}\{0.9, 1.1\}$ | CU |
| Cost rate for overcapacity | $oc$ | 10 | | MU / CU |

In this example, the decision stage in period $p = 5$ is considered. A re-optimization interval of $\Delta = 2$ is assumed. In the previous decision stage in period 3, a production plan is assumed to have been selected that results in a disaggregated mean expected tardiness of 0.3 periods. This production plan is stored in parameters $\overline{X}_{kt}$ and $\overline{Q}_{kt}$ and revisited in the current decision stage. Six discrete nervousness levels, according to Table 8, are taken into account.

Table 8: Overview over the nervousness levels in the numerical example

| Nervousness level $n$ | $h_n^{fixX}$ | $h_n^{fixQ}$ |
|---|---|---|
| 0 | $h^{look}$ | $h^{look}$ |
| 1 | 9 | 9 |
| 2 | 9 | 3 |
| 3 | 3 | 3 |
| 4 | 3 | $\Delta$ |
| 5 | $\Delta$ | $\Delta$ |

Nervousness level 0 conserves all decisions made in the previous decision stage. Thus, the production plan can only be decided for the last $\Delta$ periods, which have not been taken into account in the previous decision stage. As this approach is usually not reasonable from a economic point of view, these results can be considered to be a borderline case. In nervousness level 1, the entire production plan is fixed 9 periods upfront. In this example, this is considered to be early enough to avoid planning nervousness and therefore treated as the base nervousness level. The remaining nervousness levels 2 - 5 fix the decisions for shorter time spans. The higher the nervousness level, the shorter the time spans of fixed decisions are.

*Evolution of demand forecasts and adjustment of the piecewise linearization*

The decisions in the previous decision stage at the beginning of period 3 were made based on the demand information available at that time. At the end of periods 3 and 4, additional demand informations $\boldsymbol{\psi}_{kt}^{\tau}$, $\forall\, k, \tau \in \{3,4\}, t < \tau + \theta$ were observed, which shall now be taken into account to adjust the production plan. Table 9 exemplarily illustrates the adjustments for product 1 for the relevant demand periods.

Table 9: Demand forecast evolution for product 1 from period 3 to period 5

| t | $\mathrm{E}[\mathbf{D}_{k=1,t}^{p=3}]$ | $\boldsymbol{\psi}_{k=1,t}^{\tau=3}$ | $\boldsymbol{\psi}_{k=1,t}^{\tau=4}$ | $\mathrm{E}[\mathbf{D}_{k=1,t}^{p=5}]$ | $\sigma_{\mathbf{D}_{k=1,t}^{p=3}}$ | $\sigma_{\mathbf{D}_{k=1,t}^{p=5}}$ |
|---|---|---|---|---|---|---|
| 2 | 126.84 | 0 | 0 | 126.84 | 0 | 0 |
| 3 | 153.46 | -7.67 | 0 | 145.79 | 13.42 | 0 |
| 4 | 86.69 | 8.59 | -10.55 | 84.73 | 18.97 | 0 |
| 5 | 89.69 | 10.12 | -9.89 | 89.92 | 23.24 | 13.42 |
| 6 | 98.56 | -14.60 | -7.72 | 76.24 | 26.83 | 18.97 |
| 7 | 106.74 | -23.57 | 11.33 | 94.50 | 30 | 23.24 |
| 8 | 90.28 | 0 | 17.58 | 107.86 | 30 | 26.83 |
| 9 | 90.91 | 0 | 0 | 90.91 | 30 | 30 |

Periods 1 and 2 were already in the past in the previous decision stage in period 3. Thus, for those demand periods, the demand realizations were already known. For the demand in period 3, most of the adjustments $\boldsymbol{\psi}_{k=1,t=3}^{\tau}$ are already known. The final adjustment step $\boldsymbol{\psi}_{k=1,t=3}^{\tau=3}$ is revealed at the end of period 3. Hence, in the beginning of period 4, the final demand realization is known. For the following demand periods, additional demand information is revealed both at the end of period 3 and period 4. For demand period 8, the first adjustment step is observed at the end of period 4, and for all subsequent demand periods, no additional information was observed between the two periods considered in this example. Based on the adjusted demand forecasts ($\mathrm{E}[\mathbf{D}_{kt}^{p=5}]$ and $\sigma_{\mathbf{D}_{kt}^{p=5}}$), parameters $cp_{ktl}$, $ebl_{ktl}$ and $eyp_{klt}$ of the piecewise linearization are updated.

*Generating boundaries of the Pareto-front*

To obtain a first impression of the feasible objective space, the ideal values $MT_{n^{asp}}^{ideal}$ and $COST_{n^{asp}}^{ideal}$ as well as the corresponding nadir values $MT_{n^{asp}}^{nadir}$ and $COST_{n^{asp}}^{nadir}$, respectively, for all considered nervousness levels are determined. The results are shown in Figure 9 with two markers for each nervousness level $n$.

Figure 9 (and the following) portray *decision maps* (see Meisel (1973)). Decision maps illustrate three-dimensional data of the multi-objective objective space as series of bi-objective cross-sections of the Pareto-front: the third dimension describing the nervousness level is portrayed by choosing different markers for different values.

In nervousness level 0, decisions can only be made for periods 18 and 19, which had not been considered in the previous decision stage, and the production plan for the remaining periods is fixed to the values from the previous decision stage already. As demand was rather overestimated, without adjustments, relatively high inventories build up, which result in high operational costs and low mean expected tardiness. The objective space attainable with nervousness level 1 is rather small as well: Pareto-optimal solutions with a mean tardiness $MT_{n^{asp}}^{asp} < 0.1013$ or $MT_{n^{asp}}^{asp} > 0.5695$ are only attainable with $n^{asp} \geq 2$. The higher the accepted nervousness, the greater the va-

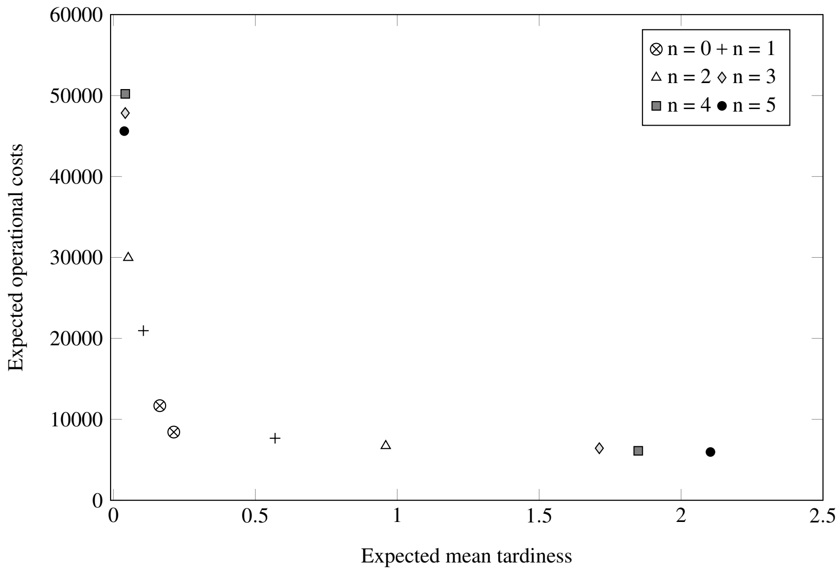



Fig. 9: Decision map with Pareto-optimal ideal solutions

riety of attainable Pareto-optimal solutions. The solutions corresponding to ideal values entail particularly bad objective function values for the nonprioritized objective functions. Therefore, in the next step shown in Figure 10, Pareto-optimal compromise solutions are determined by applying the augmented $\varepsilon$-constraint method for each nervousness level in the interval from $MT^{min}_{n^{asp}} = 0.1$ to $MT^{max}_{n^{asp}} = 1$ with an increment of $\varepsilon = 0.1$. As the ideal values have been determined before, all problem instances of the MO-SCLSP$_\varepsilon$ with $MT^{asp}_{n^{asp}} < MT^{ideal}_{n^{asp}}$ or $MT^{asp}_{n^{asp}} > MT^{nadir}_{n^{asp}}$ are skipped, as it is already known that they are infeasible ($MT^{asp}_{n^{asp}} < MT^{ideal}_{n^{asp}}$) or do not generate additional Pareto-optimal solutions ($MT^{asp}_{n^{asp}} > MT^{nadir}_{n^{asp}}$). The results from the augmented $\varepsilon$-constraint method are depicted in Figure 10.

Figure 10 illustrates the conflicts between the operational costs, mean tardiness and planning nervousness. Particularly for achieving higher deviations from the aspired tardiness value from the last decision stage, accepting a higher level of nervousness can reduce the corresponding operational costs. In this example, particularly high cost differences can be observed comparing nervousness levels 1 and 2, while in the case of nervousness levels 4 and 5, the differences are rather small. In some cases, the aspired level of mean tardiness $MT^{asp}_{n^{asp}}$ is undershot in the optimal solution. In this example in Figure 10, this can be observed, for example, for nervousness level 2 with $MT^{asp}_{n^{asp}=2} = 0.8$, $MT^{asp}_{n^{asp}=2} = 0.9$ and $MT^{asp}_{n^{asp}=2} = 1$. This happens if no Pareto-optimal solution with $MT^{asp}_{n^{asp}}$ exists: a solution with $MT < MT^{asp}_{n^{asp}}$ dominates the optimal solution forcing $MT = MT^{asp}_{n^{asp}}$.

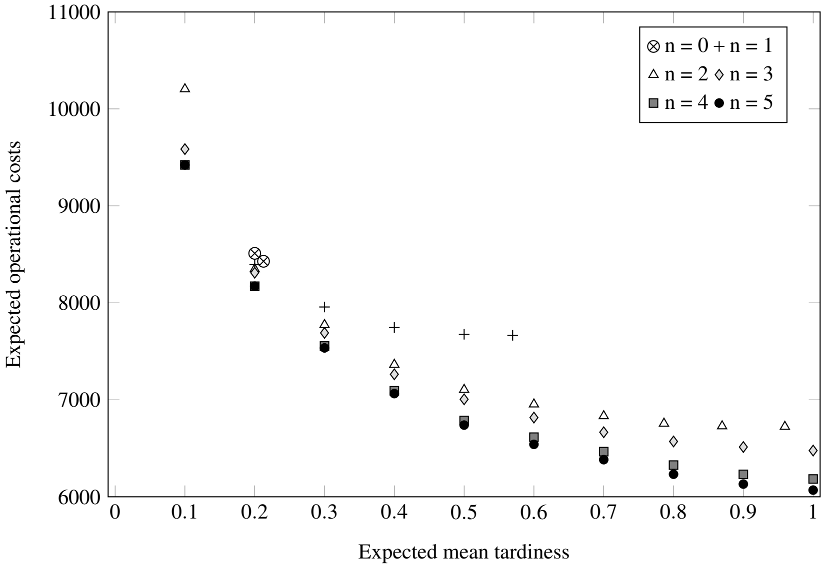



Fig. 10: Decision map after applying the $\varepsilon$-constraint method

*Generating additional Pareto-optimal solutions based on the preferences of the decision maker*
Based on the Pareto-optimal solutions determined so far, the decision maker might identify the area of the objective space with a mean tardiness of 0.2 - 0.4 periods as interesting for potential compromise solutions between expected mean tardiness and operational costs. Cost differences can particularly be observed between nervousness levels 1 and 2. If the cost reduction attainable when higher nervousness levels would be accepted does not balance out the increased system nervousness, the decision maker could exclude nervousness levels 3, 4 and 5 from further investigation. For the remaining nervousness levels, a detailed analysis by applying the augmented $\varepsilon$-constraint method is carried out with $\hat{N} = \{1,2\}$, $MT_{n^{asp}}^{min} = 0.2$, $MT_{n^{asp}}^{max} = 0.4$ and $\varepsilon = 0.02$. The results are presented in Figure 11. Based on this information, the decision maker might select a final solution.

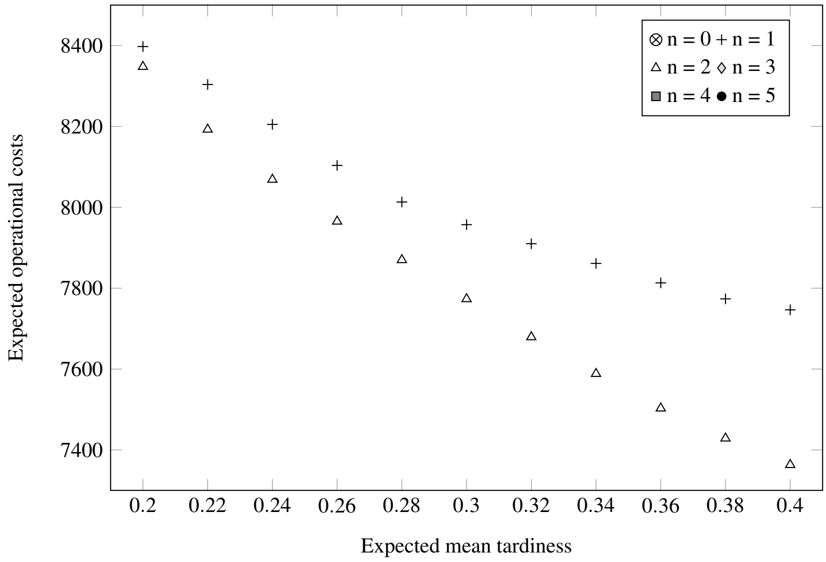



Fig. 11: Decision map after applying the $\varepsilon$-constraint method with finer step size

7.2 Comparing solutions from the interactive approach with various benchmarks

In Subsection 7.1, we exemplarily showed how a decision maker could select a Pareto-optimal solution in a single decision stage. This implies a certain degree of manual interaction of the decision maker. To study a larger number of problem instances with different demand trajectories based on the problem setting presented in Subsection 7.1, but with a single product, we now replace the implicit preferences of the decision maker and the individual decisions and assume that the usage of linear decision rules can mimic the decision-makers behaviour over time. Linear relationships between the preferences between the three objective functions are therefore assumed. This allows for the determination of a single scalarized objective function value by assigning some cost based assessment to each objective function metric. Replacing the decision maker with a simple linear rule is a substantial simplification, which in general is not sufficient to account for the complexity of finding compromise solutions between the conflicting objectives. However, it allows for the systematic selection of Pareto-optimal solutions and therefore for the analysis of a large number of problem instances without manual (and subjective) interaction of a decision maker and thus shows how decisions made over time could determine the system behaviour.

After Pareto-optimal solutions have been determined in a given decision stage, a solution can be selected by solving model (62) - (64) with the notation presented in Table 10.

Table 10: Notation for the rule-based solution selection

| | |
|---|---|
| Indices and index sets: | |
| $i \in \{1,...,I\}$ | Pareto-optimal solutions |
| Parameter: | |
| $oc_i$ | operational costs for solution $i$ |
| $tc_i$ | cost-based assessment of the mean tardiness for solution $i$ |
| $nc_i$ | cost-based assessment of the planning nervousness for solution $i$ |
| Decision variables: | |
| $\Gamma_i$ | binary selection variable: 1 if solution $i$ is selected |

**Rule-based solution selection**

$$\min\ \tilde{Z} = \sum_{i=1}^{I} \Big( oc_i + tc_i + nc_i \Big) \cdot \Gamma_i \tag{62}$$

s.t.

$$\sum_{i=1}^{I} \Gamma_i = 1 \tag{63}$$

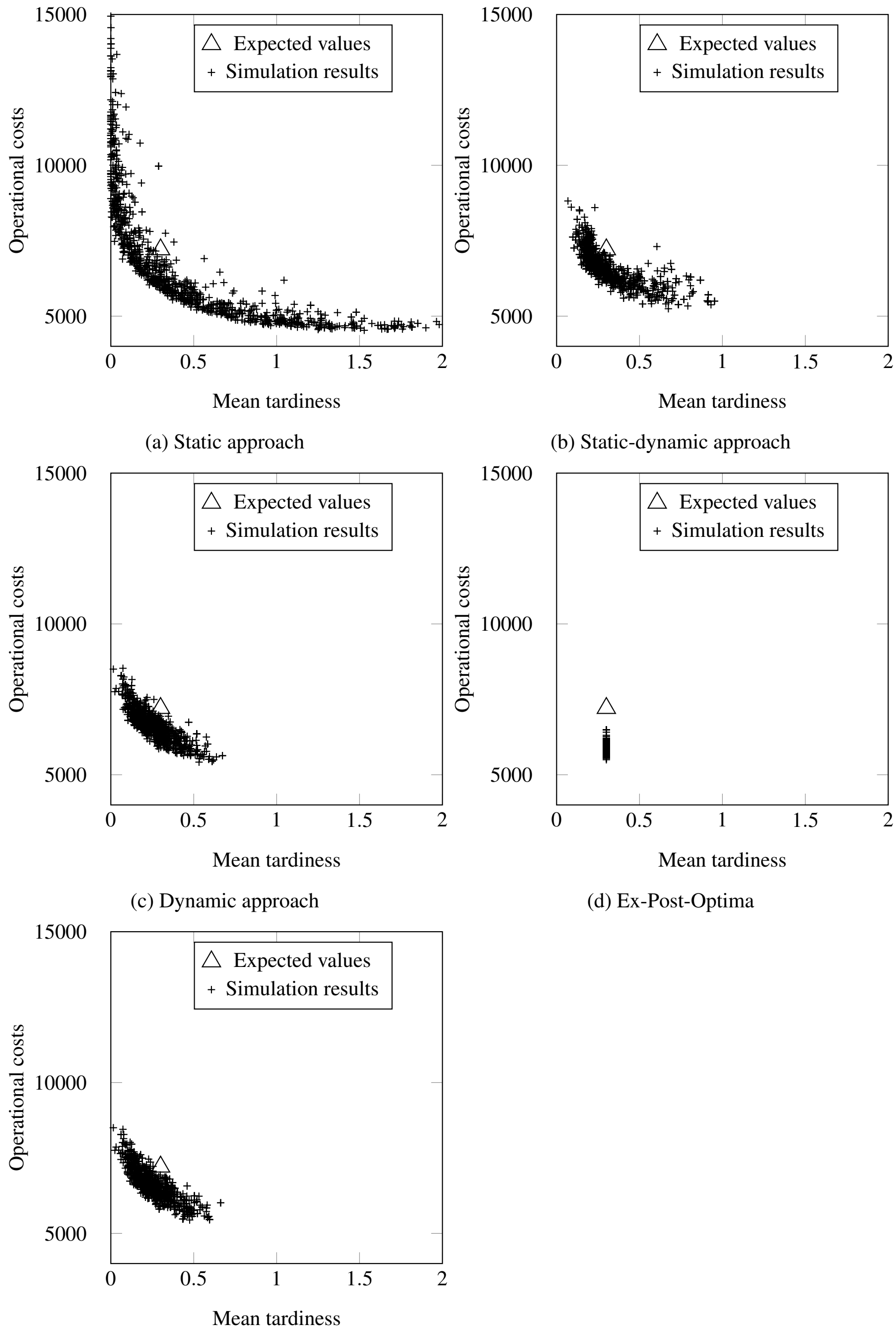
$$\Gamma_i \in \{0,1\} \tag{64}$$

A solution shall be selected from the set of Pareto-optimal solutions $i$ (63), minimizing the sum of operational costs and cost-based assessments of the mean tardiness and the production plan nervousness (62).

Preliminary numerical studies showed that for the problem setting at hand, a conversion rate for assessing the mean tardiness of 5500 GE per period of mean tardiness and a conversion rate for planning nervousness of 120 GE per nervousness level results in selecting economic compromise solutions, often dominating the expected values. This information could have been derived from comparing the selected final solution with other Pareto-optimal solutions over the course of many decision stages, as this gives an insight into the inherent internal preferences of the decision maker. By analyzing the decisions over a longer horizon, some explicit understanding of the implicit preferences can be derived.

To generate Pareto-optimal solutions, the augmented $\varepsilon$-constraint-method is applied twice with different parameters: a broad overview is generated with $\hat{N} = \{0,1,2,3,4,5\}$, $MT_{n^{asp}}^{min} = 0.1$, $MT_{n^{asp}}^{max} = 1$ and $\varepsilon = 0.1$. A more detailed search is conducted for all nervousness levels for the range between $MT_{n^{asp}}^{min} = 0.2$ and $MT_{n^{asp}}^{max} = 0.4$ with an increment of $\varepsilon = 0.02$. The production plan is revisited every $\Delta = 2$ periods.

The final realized objective function values derived with the interactive approach with rule-based solution selection are compared with the objective function values resulting from executing the ex ante optimal (static) production plan without any adjustments and the static-dynamic and the dynamic approach proposed by Bookbinder and Tan (1988). Additionally, for a theoretical benchmark of the solution quality, the ex post optimal solution is determined as well. Note that the ex post optimal solution could only be attained if all the uncertain demand realizations were known at the beginning of the planning horizon and thus deviations from the ex-post optimum are expected even in high quality solutions. One thousand random demand trajectories are drawn from the demand distribution functions. For each problem instance, the interactive approach with rule-based solution selection is applied and compared to the three approaches proposed by Bookbinder and Tan (1988) as well as the ex post optimum. Figure 12 shows the scatterplots of the resulting combinations of the realized objective function values.




(a) Static approach

(b) Static-dynamic approach

(c) Dynamic approach

(d) Ex-Post-Optima

(e) Interactive approach

Fig. 12: Comparison of results from the interactive approach and various benchmarks

Figure 12 (a) shows results derived with the static approach without any adjustments of the production plan. The results clearly show that applying the static approach, the resulting objective function values are random with high variance. In many realizations they deviate from the expected values by more than 100 %. With the static-dynamic approach depicted in Figure 12 (b) the deviations from the expected values can be reduced as frequent adjustments of the production quantities are allowed. Additionally, allowing adjustments of the setup pattern results in further reduction of the range of the realized values, as Figure 12 (c) portrays. As a theoretical benchmark, the ex-post-optima of the 1000 studied problem instances are depicted in Figure 12 (d). Each ex-post-optimal production plan exactly leads to the target value of mean tardiness.

The final results derived with the suggested interactive approach with rule based solution selection applying model (62) - (64) are shown in Figure 12 (e). The objective function realizations appear quite similar to those derived with the dynamic approach (Figure 12 (c)). However, potentially the results are of higher solution quality in terms of production plan nervousness. Figure 13 explores the production plan nervousness induced by applying the interactive approach.

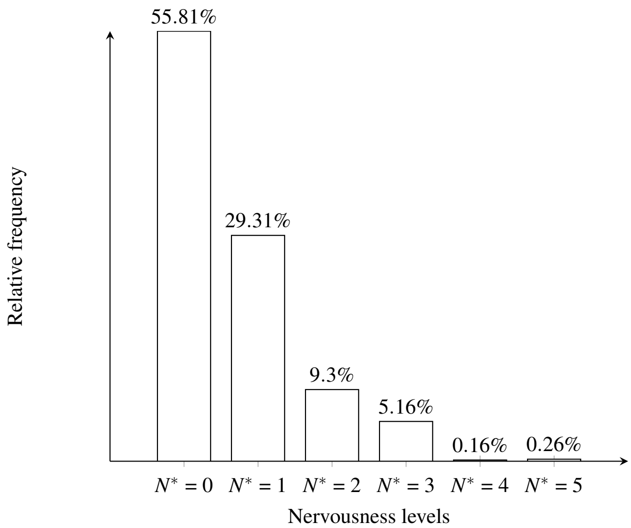



Fig. 13: Histogram of selected nervousness levels

Figure 13 shows the relative frequency of the nervousness levels of the solutions selected by the selection rules. Each of the 1000 studied problem instances involved 14 decision stages with possible adjustments of the production plan. In 29.31 % of these decisions (4103 out of 14000), a solution complying with nervousness level $N^* = 1$ was selected. This nervousness level is interpreted as basis level, and it implies production plan fixations for the next 9 periods. In 7813 decisions (55.81 %), nervousness level $N^* = 0$ was selected, fixing each decision from the last production plan. A nervousness level $N^* = 2$ was accepted in 9.3 % of the decisions and $N^* = 3$ in 5.16 % of the decision stages. Summarizing the results so far, in 85.12 % of the decisions the whole production plan remained unchanged for at least the following 9 periods and in 99.58 % of the decisions for at least the following 3 periods. Nervousness levels $N^* = 4$ and $N^* = 5$ were only selected sporadically. In the vast majority of the decisions, the production plan could be retained for (at least) the next nine periods. Production plan adjustments in one of the nine next periods were on average made about twice per problem instance with a planning horizon of 30 periods. Planning nervousness was therefore accepted systematically in those cases, in which accepting nervousness meant substantially better objective function values for the other objective functions.

Table 11: Comparison of statistical characteristics of the objective function values applying the interactive approach and different benchmarks

| | Static | Stat-dyn | Dynamic | Ex-Post | Interactive |
|---|---|---|---|---|---|
| Mean [operational costs] [MU] | 6806.0 | 6640.6 | 6650.7 | 5836.7 | 6626.5 |
| Coeff. of variation [operational costs] | 0.311 | 0.089 | 0.080 | 0.022 | 0.080 |
| Mean [tardiness] [periods] | 0.507 | 0.321 | 0.260 | 0.3 | 0.257 |
| Coeff. of variation [tardiness] | 1.059 | 0.522 | 0.445 | 0 | 0.436 |

Table 11 compares the means and coefficients of variation of both operational costs and mean tardiness for the five planning approaches. The dispersion of the objective function values can be considerably reduced by applying the proposed interactive approach. The sample coefficient of variation of the mean tardiness applying the interactive approach is only about one-quarter of the sample coefficient of variation applying the static approach. Similarly, the sample mean of the mean tardiness could be reduced from 1.059 to 0.436 by applying the interactive approach. The interactive approach leads to slightly better mean values of operational costs and tardiness compared to the static-dynamic and the dynamic approach. As this also comes with a reduced level of nervousness (see Figure 13), the interactive approach dominates the other approaches in terms of the mean values of all three considered objectives.

It should be highlighted that a selection of Pareto-optimal solutions based on simple linear decision rules does not suffice to account for the complexity of finding an economically profound compromise solution. However, the study shows that the static approach can already be outperformed with simple rules. By manual decision based on the domain knowledge of the decision maker, the adjustments to production plans can be customized even better to the actual demand situation, which eventually results in a better solution quality and reduced dispersion of the objective function values.

## 8 Conclusions, managerial insights and outlook

In this paper, we proposed an interactive framework for multi-objective stochastic lot sizing and demonstrated its utility. The approach explicitly addresses the conflicting relationships between low operational costs, low tardiness, and low system nervousness. In multiple decision stages, the production plans are revisited periodically to adapt to demand realizations. Demand forecasts are updated periodically as demand information is revealed gradually over time. Demand is modeled as a linear combination of independent random variables for different observation periods. To facilitate solving the problem as mixed-integer problem, the expected values of the random variables for physical inventories and backlogs are approximated with piecewise linear functions. In an interactive optimization approach, a set of Pareto-optimal solutions is generated by determining ideal-points and applying a multidimensional augmented $\varepsilon$-constraint method with different aspiration levels according to specifications made by a decision maker. Numerical examples show that the relationships between the analyzed objective metrics depend considerably on the planning situation and demand realizations observed so far. In many situations, a significantly better objective function value can be achieved for one objective function if slightly poorer objective function values are accepted for the other objective functions.

Different important managerial insights can be derived from our numerical studies: specifying objective bounds for the delivery reliability and system nervousness unaware of the objective space does not in general lead to good compromise solutions. A multi-objective approach evinces the relationship of the objectives, thus allowing for a reasoned positioning according to the preferences of the decision maker. Our numerical studies also show that focusing on expected values in a static approach, without taking demand realizations into account, leads to random outcomes. Multiple decision stages allow for adaptations to production plans and give the decision maker control over the advancement of the objective value metrics. The proposed interactive approach can finally be used to derive some explicit insight into the implicit preferences of the decision maker between the objectives. By analyzing selections over the course of many optimizations, this information can be used to improve the multi-objective approach by focusing on the relevant area of the objective space.

Future research should consider the application of heuristics to approximate the Pareto front to tackle larger problem instances. Another research direction should address alternative metrics for the objectives. Finally, a more sophisticated approach to externalize the preferences of the decision maker could be developed, as this would allow for the selection of a Pareto-optimal solution based on the preferences of the decision maker, without manual interaction.

## Conflict of interest

The authors declare that they have no conflict of interest.


## Funds

No funds, grants, or other support was received.


## References


Albey E, Norouzi A, Kempf K, Uzsoy R (2015) Demand modeling with forecast evolution: An application to production planning. Semiconductor Manufacturing, IEEE Transactions on 28:374–384, DOI 10.1109/TSM.2015.2453792

Albey E, Uzsoy R, Kempf KG (2016) A chance constraint based multi-item production planning model using simulation optimization. In: 2016 Winter Simulation Conference (WSC), pp 2719–2730, DOI 10.1109/WSC.2016.7822309

Azadnia AH, Saman MZM, Wong KY (2015) Sustainable supplier selection and order lot-sizing: an integrated multi-objective decision-making process. International Journal of Production Research 53(2):383–408, DOI 10.1080/00207543.2014.935827

Bookbinder J, Tan JY (1988) Strategies for the probabilistic lot-sizing problem with service-level constraints. Management Science 34:1096–1108, DOI 10.1287/mnsc.34.9.1096

Bookbinder JH, H'ng BT (1986) Rolling horizon production planning for probabilistic time-varying demands. International Journal of Production Research 24(6):1439–1458, DOI 10.1080/00207548608919814

Brandimarte P (2006) Multi-item capacitated lot-sizing with demand uncertainty. International Journal of Production Research 44(15):2997–3022, DOI 10.1080/00207540500435116, `https://doi.org/10.1080/00207540500435116`

Buschkühl L, Sahling F, Helber S, Tempelmeier H (2010) Dynamic capacitated lot-sizing problems: a classification and review of solution approaches. OR Spectrum 32(2):231–261, DOI 10.1007/s00291-008-0150-7

Chankong V, Haimes YY (1983) Multiobjective Decision Making: Theory and Methodology. Dover Publications Inc.

Claisse M, Chu C, Jemai Z (2016) Production planning under uncertainties and forecasts updates. In: 2016 International Conference on Control, Decision and Information Technologies (CoDIT), pp 490–495

De Smet NDS, Minner S, Aghezzaf EH, Desmet B (2020) A linearisation approach to the stochastic dynamic capacitated lotsizing problem with sequence-dependent changeovers. International Journal of Production Research 58(16):4980–5005, DOI 10.1080/00207543.2020.1736722

Deb K (2001) Multi-Objective Optimization Using Evolutionary Algorithms. John Wiley & Sons, Inc., USA

Ehrgott M (2005) Multicriteria Optimization (2. ed.). Springer, DOI 10.1007/3-540-27659-9

Forel A, Grunow M (2022) Dynamic stochastic lot sizing with forecast evolution in rolling-horizon planning. Working Paper

Gade D, Küçükyavuz S (2013) Formulations for dynamic lot sizing with service levels. Naval Research Logistics (NRL) 60(2):87–101, DOI https://doi.org/10.1002/nav.21519

Gopalakrishnan M, Miller DM, Schmidt CP (1995) A framework for modelling setup carryover in the capacitated lot sizing problem. International Journal of Production Research 33(7):1973–1988, DOI 10.1080/00207549508904793

Gruson M, Cordeau JF, Jans R (2018) The impact of service level constraints in deterministic lot sizing with backlogging. Omega 79:91 – 103, DOI https://doi.org/10.1016/j.omega.2017.08.003

Haimes YY, Lasdon LS, Wismer DA (1971) On a bicriterion formation of the problems of integrated system identification and system optimization. IEEE Transactions on Systems, Man and Cybernetics 1(3):296–297

Hajipour V, Kheirkhah A, Tavana M, Absi N (2015) Novel pareto-based meta-heuristics for solving multi-objective multi-item capacitated lot-sizing problems. The International Journal of Advanced Manufacturing Technology 80(1):31–45, DOI 10.1007/s00170-015-6993-6

Heath DC, Jackson PL (1994) Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. IIE Transactions 26(3):17–30, DOI 10.1080/07408179408966604

Helber S, Sahling F, Schimmelpfeng K (2013) Dynamic capacitated lot sizing with random demand and dynamic safety stocks. OR Spectrum 35(1):75–105, DOI 10.1007/s00291-012-0283-6

Isermann H (1982) Linear lexicographic optimization. Operations-Research-Spektrum 4(4):223–228, DOI 10.1007/BF01782758

Jans R, Degraeve Z (2008) Modeling industrial lot sizing problems: a review. International Journal of Production Research 46(6):1619–1643, DOI 10.1080/00207540600902262

Karimi B, Fatemi Ghomi S, Wilson J (2003) The capacitated lot sizing problem: a review of models and algorithms. Omega 31(5):365 – 378, DOI https://doi.org/10.1016/S0305-0483(03)00059-8

Koca E, Yaman H, Akturk MS (2018) Stochastic lot sizing problem with nervousness considerations. Computers & Operations Research 94:23–37, DOI {10.1016/j.cor.2018.01.021}

Lee HL, Padmanabhan V, Whang S (1997) Information distortion in a supply chain: The bullwhip effect. Management Science 43(4):546–558, DOI 10.1287/mnsc.43.4.546

Li L, Song S (2015) Solving a multi-level capacitated lot sizing problem with random demand via a fix-and-optimize heuristic. In: 2015 IEEE Congress on Evolutionary Computation (CEC), pp 2721–2728, DOI 10.1109/CEC.

2015.7257226
Little JDC (1961) A proof for the queuing formula: L = $\lambda$w. Operations Research 9(3):383–387, DOI 10.1287/opre.9.3.383
Luque M, Ruiz F, Miettinen K (2011) Global formulation for interactive multiobjective optimization. OR Spectrum 33(1):27–48, DOI 10.1007/s00291-008-0154-3
Mavrotas G (2009) Effective implementation of the $\varepsilon$-constraint method in multi-objective mathematical programming problems. Applied Mathematics and Computation 213(2):455–465, DOI https://doi.org/10.1016/j.amc.2009.03.037
Mehdizadeh E, Hajipour V, Mohammadizadeh MR (2016) A bi-objective multi-item capacitated lot-sizing model: Two pareto-based meta-heuristic algorithms. International Journal of Management Science and Engineering Management 11(4):279–293, DOI 10.1080/17509653.2015.1086965
Meisel WS (1973) Tradeoff decision in multiple criteria decision making. In: Cochrane J, Zeleny M (eds) Multiple Criteria Decision Making, University of South Carolina Press, pp 461–476
Meistering M, Stadtler H (2017) Stabilized-cycle strategy for capacitated lot sizing with multiple products: Fill-rate constraints in rolling schedules. Production and Operations Management 26(12):2247–2265, DOI 10.1111/poms.12757
Miettinen K (1999) Nonlinear Multiobjective Optimization. Kluwer Academic Publishers
Norouzi A, Uzsoy R (2014) Modeling the evolution of dependency between demands, with application to inventory planning. IIE Transactions 46(1):55–66, DOI 10.1080/0740817X.2013.803637
Oezer O, Wei W (2004) Inventory control with limited capacity and advance demand information. Operations Research 52(6):988–1000, DOI 10.1287/opre.1040.0126
van Pelt TD, Fransoo J (2017) A note on “linear programming models for a stochastic dynamic capacitated lot sizing problem”. Computers & Operations Research 89, DOI 10.1016/j.cor.2017.06.015
Rehman HU, Wan G, Zhan Y (2019) Multi-level, multi-stage lot-sizing and scheduling in the flexible flow shop with demand information updating. International Transactions in Operational Research 28(Advance online publication), DOI 10.1111/itor.12645
Rezaei J, Davoodi M (2011) Multi-objective models for lot-sizing with supplier selection. International Journal of Production Economics 130(1):77 – 86, DOI https://doi.org/10.1016/j.ijpe.2010.11.017
Rezaei J, Davoodi M, Tavasszy L, Davarynejad M (2016) A multi-objective model for lot-sizing with supplier selection for an assembly system. International Journal of Logistics Research and Applications 19(2):125–142, DOI 10.1080/13675567.2015.1059411
Robinson P, Narayanan A, Sahin F (2009) Coordinated deterministic dynamic demand lot-sizing problem: A review of models and algorithms. Omega 37(1):3–15, DOI https://doi.org/10.1016/j.omega.2006.11.004

Rossi R, Kilic OA, Tarim SA (2015) Piecewise linear approximations for the static–dynamic uncertainty strategy in stochastic lot-sizing. Omega 50:126 – 140, DOI https://doi.org/10.1016/j.omega.2014.08.003

Saliby E (1990) Descriptive sampling: A better approach to monte carlo simulation. Journal of the Operational Research Society 41(12):1133–1142

Sereshti N, Adulyasak Y, Jans R (2020) The value of aggregate service levels in stochastic lot sizing problems. Omega p 102335, DOI https://doi.org/10.1016/j.omega.2020.102335

Simon HA (1956) Rational choice and the structure of the environment. Psychological Review 63:129–138

Sox C, Jackson P, Bowman A, Muckstadt J (1999) A review of the stochastic lot scheduling problem. International Journal of Production Economics 62(3):181–200, DOI {10.1016/S0925-5273(98)00247-3}

Spitzer FL (1976) Principles of random walk / Frank Spitzer, 2nd edn. Springer-Verlag New York

Stevens SS (1946) On the theory of scales of measurement. Science 103:677 – 680, DOI https://doi.org/10.1126/science.103.2684.677

Tarim S, Kingsman BG (2004) The stochastic dynamic production/inventory lot-sizing problem with service-level constraints. International Journal of Production Economics 88(1):105 – 119, DOI https://doi.org/10.1016/S0925-5273(03)00182-8

Tavaghof-Gigloo D, Minner S (2021) Planning approaches for stochastic capacitated lot-sizing with service level constraints. International Journal of Production Research 59(17):5087–5107, DOI 10.1080/00207543.2020.1773003

Tempelmeier H (2007) On the stochastic uncapacitated dynamic single-item lotsizing problem with service level constraints. European Journal of Operational Research 181(1):184 – 194, DOI https://doi.org/10.1016/j.ejor.2006.06.009

Tempelmeier H (2010) A column generation heuristic for dynamic capacitated lot sizing with random demand under a fill rate constraint. Omega 39:627–633, DOI 10.1016/j.omega.2011.01.003

Tempelmeier H (2011a) A column generation heuristic for dynamic capacitated lot sizing with random demand under a fill rate constraint. Omega 39(6):627 – 633, DOI https://doi.org/10.1016/j.omega.2011.01.003

Tempelmeier H (2011b) Inventory Management in Supply Networks. Books on demand

Tempelmeier H (2013) Stochastic Lot Sizing Problems. In: Smith JM, Tan B (eds) Handbook of Stochastic Models and Analysis of Manufacturing System Operations, International Series in Operations Research & Management Science, Springer, chap 0, pp 313–344, DOI 10.1007/978-1-4614-6777-9

Tempelmeier H, Herpers S (2011) Dynamic uncapacitated lot sizing with random demand under a fillrate constraint. European Journal of Operational Research 212(3):497 – 507, DOI https://doi.org/10.1016/j.ejor.2011.02.012

Tempelmeier H, Hilger T (2015) Linear programming models for a stochastic dynamic capacitated lot sizing problem. Computers & Operations Research 59:119 – 125, DOI https://doi.org/10.1016/j.cor.2015.01.007

Tunc H, Kilic O, Tarim S, Eksioglu B (2013) A simple approach for assessing the cost of system nervousness. International Journal of Production Economics 141(2):619–625, DOI 10.1016/j.ijpe.2012.09.022

Winands E, Adan I, Houtum GJ (2011) The stochastic economic lot scheduling problem: A survey. European Journal of Operational Research 210:1–9, DOI 10.1016/j.ejor.2010.06.011

Ziarnetzky T, Mönch L, Uzsoy R (2018) Rolling horizon, multi-product production planning with chance constraints and forecast evolution for wafer fabs. International Journal of Production Research 56(18):6112–6134, DOI 10.1080/00207543.2018.1478461

Ziarnetzky T, Mönch L, Uzsoy R (2020) Simulation-based performance assessment of production planning models with safety stock and forecast evolution in semiconductor wafer fabrication. IEEE Transactions on Semiconductor Manufacturing 33(1):1–12, DOI 10.1109/TSM.2019.2958526