



HAL
open science

A robust data driven approach to supply planning.

Benoit Loger, Alexandre Dolgui, Fabien Lehuédé, Guillaume Massonnet

► To cite this version:

Benoit Loger, Alexandre Dolgui, Fabien Lehuédé, Guillaume Massonnet. A robust data driven approach to supply planning.. APMS 2021: IFIP International Conference on Advances in Production Management Systems, IMT Atlantique, Sep 2021, Nantes, France. pp.169-178, 10.1007/978-3-030-85910-7_18 . hal-03545577

HAL Id: hal-03545577

<https://hal.science/hal-03545577>

Submitted on 10 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

A robust data driven approach to supply planning

Benoit Loger ✉, Alexandre Dolgui, Fabien Lehuédé, and Guillaume Massonnet

IMT Atlantique, LS2N
4 rue Alfred Kastler - La Chantrerie - 44307 Nantes - France
`benoit.loger@imt-atlantique.fr`

Abstract. We develop two robust optimization models to plan the supply operations of an assembly line when the latter are subcontracted to an external service provider. The uncertainty sets are constructed from available information on picking times both in a classical budget-based robust approach and by using Support Vector Clustering. Numerical experiments are conducted on test instances derived from a practical case to illustrate the effectiveness of the proposed approach. The results show that the robust optimization approach is efficient to reduce the impact of picking time uncertainties on production and that the SVC-based model outperforms the classical budget-based model.

Keywords: Supply chain · Data-driven optimization · Robust optimization.

1 Introduction

We focus on a problem derived from a practical case encountered in the aircraft industry, in which a single assembly line manufactures several products in order to satisfy the needs that are planned over a discrete and finite planning horizon. Each final product has its own bill of material (BOM) that involves different components. The components are stored in a remote warehouse and their picking and delivery to the production line are handled by a third party logistic provider (TPL) following the manufacturer's orders. In a sense, this problem is closely related to the class of assemble-to-order systems, which are often encountered in the inventory control literature. The need to define both component replenishment and production policies makes assemble-to-order systems particularly challenging. Large surveys of this topic have been presented in [1] and more recently in [2] where authors established that the general case of multiple final products and multiple periods have not been solved to optimality yet. However, the recent work of [3] appear like a significant step forward towards this objective. Their method is based on a two stage stochastic program which decomposes the original problem considering the component replenishment problem as the first stage of decision and the assignment problem as second stage decisions.

In this paper, we investigate the planning of picking orders when the available information related to picking times at the TPL is uncertain. Several approaches

have been developed in the literature to deal with uncertainty in optimization problems (see [4] for a recent survey of these techniques). Stochastic programming and stochastic dynamic programming are two examples of optimization tools that are frequently applied when some parameters of the problem under study are random. These methods relies on the assumption that the random variables of the problem are characterized by a known probability distribution. The main issue with this paradigm is that in most real-life applications this knowledge is not available. To overcome the lack of accurate information, researchers from the operations research community have dedicated considerable efforts in the past decade to develop optimization methods that are robust to imprecision in the modeling assumptions.

The concept of robust optimization was introduced by [5] who considers that unknown parameters belong to a finite *uncertainty set* without making any assumption on their distribution. The objective is then to find a solution that is optimal when the uncertain parameters take their worst values in the uncertainty set. To avoid over-conservatism, [6] and [7] introduce an ellipsoidal uncertainty set controlled by a size parameter and derive a second-order cone robust program. In [8], the authors propose an alternative approach that constructs a polyhedral uncertainty set whose size is controlled by a so-called budget of uncertainty, which also limits over-conservatism. The major advantage of the latter is that when the original problem is formulated as a linear (or mixed integer) program, its robust counterpart remains of the same nature, which is convenient in terms of tractability. The design of the uncertainty sets have been in the center of researchers considerations since the early developments of robust optimization. One major restriction on these uncertainty sets is that they assume that uncertain parameters are symmetrically and independently distributed. Few works have proposed ways to reduce the effects of these drawbacks. For example, [9] modifies the definition of the polyhedral uncertainty set to improve their average performances and [10] integrates correlations between parameters to improve the accuracy of the robust models in this case.

The first application of RO to inventory management is due to [11], the authors apply the model of [8] to tackle uncertainty on the demand for a multi-period inventory management problem. On similar topics, [12] considered a production planning problem with uncertain costs and demands and [13] considered uncertain returns and demands. In [14], the authors proposed a robust model for an integrated production planning with order acceptance problem where customer demands is uncertain. The RO approach have been extended to uncertainty on raw material availability in [15] to schedule the production of a sawmill with uncertainty on products demand and raw material availability and in [16] where the authors proposed a robust approach to tackle lead-time uncertainty.

The growing volume of data available at the operational level recently became a focal point for researchers in order to better incorporate uncertainty into optimization methods. We now witness the emergence of the so-called *Data-Driven Robust Optimization* (DDRO) models in the literature, where uncertainty sets are designed directly from the data available. Among the different DDRO ap-

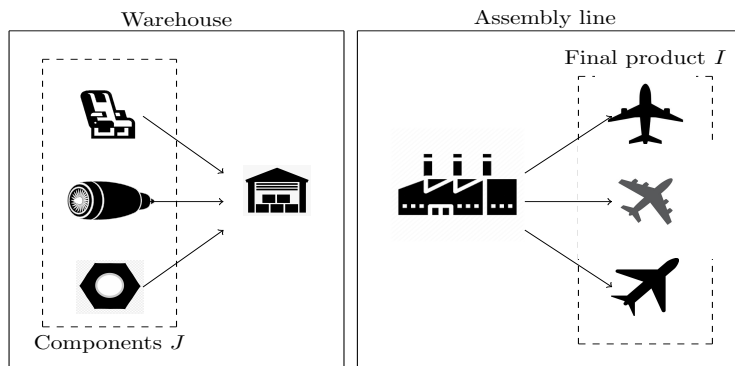


Fig. 1. Diagram of the problem

proaches, [17] and [18] use coherent risk measure to construct uncertainty sets for linear optimization problems. [19] proposed an uncertainty set based on a support vector clustering algorithm to avoid over conservatism. They introduced a piecewise linear kernel function to derive a polyhedral uncertainty set and a tractable reformulation of linear optimization problem. Their uncertainty set gives a better description of random variables in the case of asymmetric distribution and leads to less conservative solutions than traditional uncertainty sets. Their approach has been applied successfully to a multi-product inventory management problem by [20] and to energy system optimization by [21].

The remainder of this paper is organized as follows. Section 2 presents the supply planning problem that motivates this study, along with a deterministic mathematical formulation. In Section 3 we derive two tractable robust formulations of the problem from the budget-based and the SVC-based uncertainty sets, respectively. Finally, Section 4 presents the experimental protocol and the numerical results obtained with both robust formulations, while Section 5 contains our concluding remarks and some perspectives for future research directions.

2 Problem statement and deterministic formulation

We study the case of a single assembly line that combines different components into several final products over a finite planning horizon discretized into T periods. We denote I and J the set of products and components, respectively. In each period $t = 1, \dots, T$, the system faces a demand d_{it} for product $i \in I$ that is assembled on demand, i.e. it is impossible to manufacture a product in advance, then store it to satisfy a future demand. For all product $i \in I$ and component $j \in J$, let r_{ij} be the number of components j required to produce one unit of product i .

All components are assumed to be available at all time in a remote warehouse managed by a TPL provider, who is in charge of delivering the assembly line based on its orders for components, as represented on Figure 1. For each picking operation for component $j \in J$, the quantity collected by an operator is

limited to a maximum batch size of m_j units. Note that several batches of the same components can be scheduled in the same period. The picking time for a particular batch of component j of size x_j can be divided into two main parts. First, we consider a fixed time p_j that corresponds to the travel time between the shipping point and the zone where components j are stored, which is spent regardless of the quantity that is transported. In addition, we consider a per-unit picking time τ_j . For all period $t = 1, \dots, T$, the total picking time spent by the TPL provider to collect all the components delivered to the assembly line in period t cannot exceed a maximum work capacity C_t .

Any demand for product i that is not satisfied immediately is backlogged until the corresponding product is assembled in a subsequent period. Each unit of product i incurs a backlogging penalty cost b_i during each period it is backlogged. Whenever a component j is available on the assembly line but is not immediately used to manufacture an end product, it disturbs the production process by interfering with people and other goods moving nearby. We model this situation with a per-unit, per-period obstruction cost o_j . The problem consists in planning the quantity of each component delivered to the assembly line by the TPL in each period such that the sum of the obstruction and backlogging costs is minimized. We declare decision variables, x_{jt} as the quantity of components $j \in J$ that should be brought to the assembly line during period t . These variables are closely related to variables v_{jt} representing the number of distinct picking operations of components $j \in J$ performed during period t . Variables s_{jt} represent the number of component $j \in J$ held on the border of the assembly line at the end of period t . Variables u_{it} denotes the number of final product $i \in I$ produced during period t and finally, variables P_t and O_t respectively denotes penalty costs caused by the backlogging of final product demand and the total obstruction cost for period t .

We formulate the problem with the following MIP:

$$\min \sum_{t=1}^T O_t + P_t \quad (1)$$

$$\text{s.t.} \quad s_{jt} = s_{j1} + \sum_{k=1}^{t-1} x_{jk} - \sum_{i \in I} r_{ij} u_{ik} \quad \forall j \in J, \forall t = 2, \dots, T \quad (2)$$

$$\sum_{i \in I} u_{it} r_{ij} \leq s_{jt} + x_{jt} \quad \forall j \in J, \forall t = 1, \dots, T \quad (3)$$

$$\sum_{k=1}^t u_{ik} \leq \sum_{k=1}^t d_{ik} \quad \forall i \in I, \forall t = 1, \dots, T \quad (4)$$

$$v_{jt} \geq x_{jt}/m_j \quad \forall j \in J, \forall t = 1, \dots, T \quad (5)$$

$$\sum_{j \in J} v_{jt} p_j + \tau_j x_{jt} \leq C_t \quad \forall t = 1, \dots, T \quad (6)$$

$$O_t \geq \sum_{j \in J} o_j \left(s_{jt} + x_{jt} - \sum_{i \in I} u_{it} r_{ij} \right) \quad \forall t = 1, \dots, T \quad (7)$$

$$P_t \geq \sum_{i \in I} b_i \left(\sum_{k=1}^t d_{ik} - u_{ik} \right) \quad \forall t = 1, \dots, T \quad (8)$$

$$x_{jt}, s_{jt}, P_t, O_t \in \mathbb{R}_+ \quad \forall j \in J, \forall t = 1, \dots, T \quad (9)$$

$$u_{it}, v_{jt} \in \mathbb{N}_+ \quad \forall i \in I, \forall j \in J, \forall t = 1, \dots, T \quad (10)$$

The objective (1) aims at minimizing the total cost incurred over the planning horizon. The inventory balance constraints (2) update the stock levels in each period. Constraints (3) ensure that the quantity of component $j \in J$ available in period t are sufficient to perform the planned assembly operations. Constraints (4) impose that the system never manufactures more units of product $i \in I$ than the expressed demand. The minimum number of picking batches corresponding to the quantity of components j ordered in each period t is defined in constraints (5). Constraints (6) ensure that the total picking time does not exceed the picking capacity of the TPL provider in any period t . Constraints (7) and (8) define the components holding costs and final product backlogging costs incurred in each period t , respectively. Finally, constraints (9) and (10) define the domain of the decision variables.

3 Robust models for uncertain setup times

In practice, the manufacturer often has incomplete or imprecise knowledge on the picking times. As a consequence, some combinations of her orders may exceed the picking capacity of the TPL provider, forcing the latter to postpone some operations to subsequent periods. This delay in some components delivery induces two types of inefficiencies on the assembly line. Indeed, missing components (i) prevent the manufacturer to assemble some of the products, leading to backlogging penalty costs and (ii) leave the other components in the BOM on the border of the line, disturbing other production operations. This double effect strongly incentivizes the planner to protect her decisions against uncertain setup times.

In what follows, we assume that we have at our disposal a set of historical setup times $\mathcal{P} = \{\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(N)}\}$ and we investigate the performance of two robust optimization models in this context.

3.1 Interval based robust model.

As a first robust approach, we consider the classical uncertainty set of [8] defined by :

$$\mathcal{U}_t = \left\{ \mathbf{p} \left| p_{jt} = \bar{p}_j + \hat{p}_j z_{jt}, \sum_j |z_{jt}| \leq \Gamma_t, \forall j \in J \right. \right\} \quad (11)$$

where \bar{p}_j is called the nominal value of p_j and \hat{p}_j is a maximal deviation. The scaled deviation coefficient $z_{jt} = (p_j - \bar{p}_j)/\hat{p}_j \in [-1, 1]$ ensures that $p_{jt} \in [\bar{p}_j - \hat{p}_j, \bar{p}_j + \hat{p}_j]$ for all components $j \in J$ and period t and that the total scaled deviation of \mathbf{p} never exceeds the budgets of uncertainty Γ . The robust counterpart of the capacity constraint is then obtained thanks to duality theory (see [8] for details), where constraints (6) are replaced with:

$$\sum_{j \in J} v_{jt} \bar{p}_j + x_{jt} \tau_j + q_t \Gamma_t + \sum_{j \in J} w_{jt} \leq C_t \quad \forall t = 1, \dots, T \quad (12)$$

$$q_t + w_{jt} \geq \hat{p}_j v_{jt} \quad \forall j \in J, \forall t = 1, \dots, T \quad (13)$$

where $\mathbf{q} \geq 0$ and $\mathbf{w} \geq 0$ are additional real variables obtained from the dual formulation of the left-hand side of constraints (6).

3.2 Support Vector Clustering based model.

We now consider an alternative, data-driven approach introduced in [19] to construct uncertainty sets directly from \mathcal{P} . This method rely on the Support Vector Clustering algorithm introduced in [22] and results in a polyhedral uncertainty set that gives a more precise description of \mathcal{P} . By using the SVC algorithm we obtain a vector $\alpha = \{\alpha_k, k = 1, \dots, N\}$ that defines whether a given sample $\mathbf{p} \in \mathcal{P}$ resides in the uncertainty set \mathcal{U} . More precisely, samples with $\alpha_k = 0$ are strictly contained in \mathcal{U} . Samples $\mathbf{p}^{(k)}$ with positive α_k are called *support vectors (SV)*. Points for which $0 < \alpha_k < 1/N\nu$ lie exactly on the boundary of \mathcal{U} and are referred to as *boundary support vectors (BSV)*. In addition, the samples contained in $SV \setminus \{BSV\}$ with $\alpha_k = 1/N\nu$ are considered as outliers and are located outside of \mathcal{U} . The regularization parameter $\nu \in [0, 1]$ is used to control the portion of samples in \mathcal{P} covered by \mathcal{U} . In practice when $N \rightarrow \infty$, the latter portion tend to be $1 - \nu$ (see [19] and [22] for details).

Similar to the robust framework of [8] the capacity constraint can be reformulated as

$$\sum_{k \in SV} (\boldsymbol{\mu}_{kt} - \boldsymbol{\lambda}_{kt})^T \mathbf{Q} \mathbf{p}^{(k)} + \eta \theta + \sum_{j \in J} x_{jt} \tau_j \leq C_t, \quad \forall t = 1, \dots, T \quad (14)$$

with the additional constraints :

$$\sum_{k \in SV} \mathbf{Q} (\boldsymbol{\lambda}_{kt} - \boldsymbol{\mu}_{kt}) + \mathbf{v}_t = \mathbf{0} \quad (15)$$

$$\boldsymbol{\lambda}_{kt} + \boldsymbol{\mu}_{kt} = \eta \alpha_k \mathbf{1} \quad \forall k \in SV \quad (16)$$

where \mathbf{Q} is called the weighting matrix and is defined by $\mathbf{Q} = \Sigma^{-\frac{1}{2}}$ with Σ the covariance matrix of \mathbf{p} , $\theta = \min_{k' \in BSV} \left(\sum_{k \in SV} \alpha_k \|\mathbf{Q}(\mathbf{p}^{(k')} - \mathbf{p}^{(k)})\|_1 \right)$ and $\boldsymbol{\lambda}_{kt} > 0$, $\boldsymbol{\mu}_{kt} > 0$, $\eta > 0$ are additional real variables obtained from the dual formulation of the left-hand side of constraints (6) (see [19] for details).

4 Experimental results

We first describe several rules that we apply to generate the instances we use to evaluate the performances of the different models presented above. The specific assumptions we consider are listed below, with the objective reflect the framework of application cases: r_{ij} is chosen randomly from an interval $[r_j, \bar{r}_j]$, the smallest interval is $[2,4]$ and the widest is $[10,25]$. Each product i requires a subset $J_i \subset J$ of components. o_j is chosen randomly in $[0, 1]$, $b_i = \rho \sum_{j \in J} r_{ij} o_j$, where ρ represents a ratio between the penalty incurred for backlogging a demand for product i versus the maximum obstruction cost of its components. In our numerical experiments, we use $\rho = 5$, but this value can be adjusted depending on the context. In order to primarily focus on the impact of the setup times uncertainty, we disregard the linear part of picking times and set $\tau_j = 0$ and $m_j = 3 \max_{i \in I} r_{ij}$ for all $j \in J$. Demands are deterministic with values drawn from a normal distribution whose parameters vary among products. We compute the capacity of the TPL C_t from the average time needed to pick the required components in each period, that we increase by a given percentage to ensure that it is not always saturated.

4.1 Evaluation of the solution

To conclude on the performance of the models we need to evaluate how these solutions performs in the face of uncertainty. To do that we first generate a set \mathcal{P} of $N = 400$ historical setup times to define the models. We then generate a set of 10 000 setup time scenarios \mathcal{S} following the same distribution than \mathcal{P} . Each scenario is composed of one setup time for each component $j \in J$ and each period $t = 1, \dots, T$. Finally we evaluate the real cost of each scenario for the given solution in three steps. For each scenario: we simulate the picking operations to obtain the realized quantity of each components delivered in each period; we compute the optimal production plan for the quantities delivered; and we evaluate the cost of the obtained production plan.

4.2 Setup time distribution

We assume that the TPL provider organizes the storage of the components in order to optimize the picking operations. Specifically, components are stored in such a way that their accessibility improves with their order frequency. We consider three types of components and separate them based on their storage area. We assume that both the mean and the variability of the setup times decrease with component accessibility. In our instances, we thus we generate setup times using Gamma distributions with three different shape parameters, one for each type of component.

In practice, setting the budget parameter I_t is hard and the performance obtained for a given parameter value heavily depends on the characteristics of the instance. In order to compare both models, we set I_t to cover a certain portion of \mathcal{P} , by iteratively increasing its value until we reach the desired coverage.

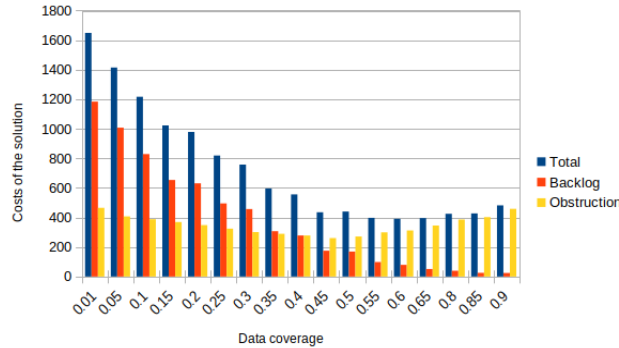


Fig. 2. Evolution of the average costs for different data coverage

4.3 Results

Figure 2 shows the mean value of the total, obstruction and backlogging costs obtained by the SVC based model with different data coverage. The instance considered is composed of 8 components and 4 products and the ratio between the obstruction and backlogging cost parameters o_j and b_i are set to reflect a high preference to avoid backlogging. We observe that the backlogging cost strictly decreases as the data coverage increases which corresponds to the preference based on the costs. The obstruction cost decreases when the data coverage is lower than 0.45 and increases otherwise. The total cost decreases to reach its lowest value for 60% of data coverage. Beyond that point, the reduction in backlogging costs do not compensate the increase of obstruction costs, which leads to a slow increase in the total cost. We also observe that the obstruction costs are nearly the same for 1% and 90% of data coverage while the backlogging costs decrease from 1195 to 26 in this interval.

Figure 3 presents the average backlogging and obstruction costs obtained by the classical and SVC based robust models. The considered instance is composed of 8 components and 4 products, the planning horizon is 5 periods. Each point corresponds to a different value of data coverage in $[1\%, 50\%]$. We observe that for the same backlogging cost, the SVC based model leads to a smaller obstruction cost than the classical robust model. For example, when both backlogging costs are equal to 288, the obstruction cost of the SVC based solution is 259 whereas it is 317 for the classical robust solution. We also observe that the SVC based model can lead at the same time to a lower obstruction cost and a lower backlogging cost. For example, the SVC based model leads to a backlogging cost of 251 with an obstruction cost of 273 whereas the classical robust model leads to a backlogging cost of 260 with an obstruction cost of 346.

5 Conclusion

In this paper, we considered a data driven robust optimization approach to plan the picking operations of an assembly line when the latter are subcontracted to an outside service provider. Experimental results show that the proposed method

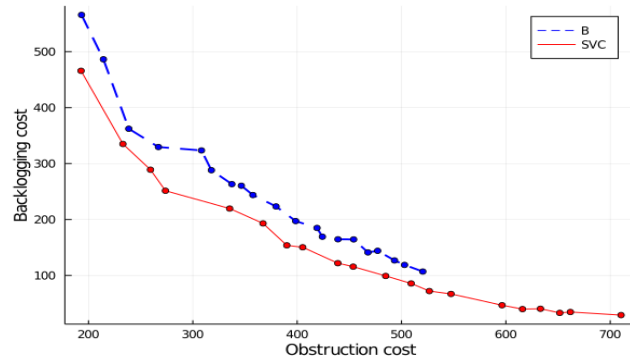


Fig. 3. Comparison of backlogging and obstruction costs obtained with different data coverage

efficiently reduces the impact of uncertainties on the performance of the assembly line. By controlling the level of robustness of the model, good tradeoffs can be found between reducing the backlog of assembly operations and minimizing the quantity of components stored on the border of the assembly line. In comparison with a classical budget based robust optimization model, the data driven model yields better performance on the test instances. The proposed robust approach aims at respecting the maximum picking time capacity of the service provider and ignore the consequences of robustness on the solution cost. Hence, it is difficult to define the robustness level in order to find the optimal trade-off. Future research may consider a multi-stage data-driven robust optimization approach to overcome this drawback. However, such methods generally lead to more complex formulations and make their application to large-scale industrial cases impossible.

References

1. Song, J.S., Zipkin, P.: Supply chain operations: Assemble-to-order systems. Handbooks in Operations Research and Management Science **11**, 561–596 (2003)
2. Atan, Z., Ahmadi, T., Stegehuis, C., de Kok, T., Adan, I.: Assemble-to-order systems: A review. European Journal of Operational Research **261**(3), 866–879 (2017)
3. Reiman, M.I., Wang, Q.: Asymptotically optimal inventory control for assemble-to-order systems with identical lead times. Operations Research **66**(3), 716–732 (2015)
4. Keith, A., Ahner, D.: A survey of decision making and optimization under uncertainty. Annals of Operations Research **300**, 1–35 (2019)
5. Soyster, A.L.: Technical note - convex programming with set-inclusive constraints and applications to inexact linear programming. Operations Research **21**(5), 1154–1157 (1973)
6. L. El Ghaoui, F.O., Lebret, H.: Robust solutions to uncertain semidefinite programs. SIAM Journal on Optimization **9**(1), 33–52 (1998)
7. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Mathematics of Operations Research **23**(4), 769–805 (1998)
8. D. Bertsimas, M.S.: The price of robustness. Operations Research **52**(1), 35–53 (2004)

9. Poss, M.: Robust combinatorial optimization with variable budgeted uncertainty. *4OR* **11**(1), 75–92 (2013)
10. Jalilvand-Nejad, A., Shafaei, R., Shahriari, H.: Robust optimization under correlated polyhedral uncertainty set. *Computers & Industrial Engineering* **92**, 82–94 (2016)
11. Bertsimas, D., Thiele, A.: A robust optimization approach to inventory theory. *Operations Research* **54**(1), 150–168 (2006)
12. José Alem, D., Morabito, R.: Production planning in furniture settings via robust optimization. *Computers & Operations Research* **39**(2), 139–150 (2012)
13. Wei, C., Li, Y., Cai, X.: Robust optimal policies of production and inventory with uncertain returns and demand. *International Journal of Production Economics* **134**(2), 357–367 (2011)
14. Aouam, T., Brahimi, N.: Integrated production planning and order acceptance under uncertainty: A robust optimization approach. *European Journal of Operational Research* **228**, 504–515 (2013)
15. Varas, M., Maturana, S., Pascual, R., Vargas, I., Vera, J.: Scheduling production for a sawmill: A robust optimization approach. *International Journal of Production Economics* **150**, 37–51 (2014)
16. Thorsen, A., Yao, T.: Robust inventory control under demand and lead time uncertainty. *Annals of Operations Research* **257**, 207–236 (2017)
17. Bertsimas, D., Brown, D.: Constructing uncertainty sets for robust linear optimization. *Operations Research* **57**(6), 1483–1495 (2009)
18. Bertsimas, D., Thiele, A.: Robust and data-driven optimization: Modern decision making under uncertainty. In: *Tutorial on Operations Research* pp. 95–122. INFORMS (2014)
19. C. Shang, X.H., You, F.: Data-driven robust optimization based on kernel learning. *Computers & Chemical Engineering* **106**, 464–479 (2017)
20. Qiu, R., Sun, Y., Shu, P., Sun, M.: Robust multi-product inventory optimization under support vector clustering-based data-driven demand uncertainty set. *Soft Computing* **24**, 6259–6275 (2019)
21. Shen, F., Zhao, L., Du, W., Zhong, W., Qian, F.: Large-scale industrial energy systems optimization under uncertainty: A data-driven robust optimization approach. *Applied Energy* **259**, 114199 (2020)
22. Ben-Hur, A., Horn, D., Siegelmann, H., Vapnik, V.: Support vector clustering. *Journal of Machine Learning Research* **2**, 125–137 (2001)