Stochastic Optimal Sizing of a Warehouse

Luca Ghezzi

Abstract The problem is considered of determining how many pieces to stock in a warehouse, for a multitude of stockable goods and accounting for random market demand and supply lead time. Classical reorder point theory is revisited, the underlying model is no longer linearized and the involved stochastic variables need not be normally distributed but, rather, are empirically deduced from historical data. Uncertainty propagation is carried out either by Monte Carlo method or by a Polynomial Chaos Expansion. A Quadratic Programming procedure is proposed to regularize data by filtering rare events out. Performance vs. cost curves are obtained and the global problem of choosing optimal points over them, subject to a global cost budget constraint, is set as a combinatorial, constrained optimization. The solution of a simplified version is attained by Linear Programming, while the full problem is addressed by Mixed Integer Linear Programming.

1 Introduction

Storage of physical goods is a practically unavoidable need in industry and commerce, common to most market sectors. As well-known, inventories allow decoupling raw materials and sub-components supply chain from production, as well as the latter from the commercial distribution of final products. Moreover, large factories employ internal sub-component warehouses (i.e., the *make-to-stock* manufacturing policy) in order to ease industrial operations by decoupling successive, serial phases, as well as to allow, if applicable, the *assemble-to-order* manufacturing policy at the very last segment of the productive chain. Similarly, commercial as well as service organizations are usually hierarchically structured, with tiers decoupled by warehouses. On the one side, stocked volumes allow for strategic, massive purchase deals, simpler factory management policies as well as high service levels in deliveries to final customers. Drawbacks include the high cost of stocked capital, the risk for stocked product obsolescence, as well as the operative cost for running the inventory.

L. Ghezzi (✉)
ABB Italy S.p.A., Mathematical Modeling & Numerical Simulations, Vittuone, MI, Italy
e-mail: luca.ghezzi@it.abb.com

L. Ghezzi et al. (eds.), *Math for the Digital Factory*, Mathematics in Industry 27,
DOI 10.1007/978-3-319-63957-4_4

The impact of inventories over operations or overall company's costs is clearly business-dependent; nonetheless, strong cost reduction initiatives pertaining to suitably sizing the inventories have been observed in last decades. A first trend is toward reducing stocks and increasing their rotations. A second trend is forecasting expected needs, to be scheduled so to have materials available only when necessary, and using stocks to cover deviations from forecasts. A third trend is shifting inventories upstream, possibly to suppliers (a fact clearly not without impact when negotiating supply conditions). A complex and multi-disciplinary scenario emerges, encompassing sales, planning, production, and logistic, where the functional relationship between warehouse benefits and costs is emphasized in order to quantitatively evaluate alternatives and draw conclusions. Clearly, such a central and transversal topic in factory dynamics may strongly benefit from the use of Mathematics to model its behavior and allow its suitable dimensioning.

1.1 Goals and Results

Modeling and predictive approaches thus find a sound economic motivation. Particularly, it is here of interest to address the problems of:

(A) Devising a performance vs. cost relationship for any stockable good, depending on the quantity being stocked;
(B) Determining the optimal quantities to stock for all stockable goods, given an overall budget constraint and an overall weighted average performance indicator, possibly lower bounded (dubbed *optimal budget expenditure problem* in what follows).

The kind of results obtained by solving problem (A) is exemplified by Fig. 6, showing the performance vs. cost curve for a sample stockable good, as obtained by the proposed approach and opposed to the one deduced by traditional approaches. The kind of results obtained by solving problem (B) is reported in Fig. 8, showing the fundamental Pareto front for the inventory, that is, the overall performance vs. cost curve. Thanks to the latter, the decision maker may define the global working point. Mathematics connect the latter to relevant working points for all stockable goods, that is, the positioning along individual performance vs. cost curves like Fig. 6. The combination of the two problems is relevant to global warehouse design as well as local warehouse revisions, and needs being solved periodically (say, quarterly) to keep the warehouse up to date with the current market scenario.

Inventories are a crucial ingredient of the "digital factory", because only their appropriate sizing allows a profitable and timely production flow. This means that, on the one hand, oversized stocks are completely inadmissible in a modern, lean production system and, on the other side, the large number of products, most of which with low volumes, make the traditional, deterministic or oversimplified stochastic approaches no longer adequate. It is precisely the stochastic nature of the problem which demands for a finer mathematical treatment than traditional variance propagation approaches.

1.2 A Historical Perspective

The basic theory dates back to 1913, with the first simple models to connect production to inventories; see [8]. The *reorder point* (ROP) is a rule-based inventory control method consisting of issuing a supply reorder up to the feeder factory (more generally, upstream) any time the stock level falls below a predetermined threshold, the ROP itself. The threshold must be high enough to prevent from stock-out while the supply order is being worked out and products are asked for from the market (more generally, downstream). Assuming constant and predictable market demand and supply lead time, a deterministic contribution is first found. Then, a *safety stock* margin is supplemented, to cope with stochastic factors. The safety stock is computed by means of uncertainty propagation: the ROP formula is linearized, variance is propagated over the linear approximation and finally a normal distribution is assumed to introduce a confidence level in safety assessment. Such procedure ultimately leads to Hadley-Whitin formula (1963), very popular among logistic professionals; see [7]. The higher the quantity being ordered, the longer the time in between two consecutive reorders, the lower the number of reorders to be issued in a given time horizon, and the higher the average stock level, market demand and all other features being the same. Accounting for fixed reorder costs (monotonically decreasing with the reordered quantity) and stock costs (monotonically increasing with the reordered quantity), the *Economic Order Quantity* (EOQ) is determined as the global cost minimizer. As an alternative to ROP-EOQ, the *reorder time* (ROT) method compares the stock level with a reference level every fixed period of time, issuing an upstream supply order to fill the possible gap.

Most of later literature is about inventory dynamics and control. Standard approaches are known for inventory management, addressing the problem of when to issue replenishing orders and how large such orders need to be, for a warehouse with prescribed sizing; see, e.g., [18] and references therein for a comprehensive exposition of the state of the art.

Most notably, in 1964 Orlicky introduced *Material Requirement Planning* (MRP) as an inventory management strategy, coordinated with production planning, that quickly became an indispensable tool in industry; see [13]. With reference to purchase orders received from the market and estimated, and the level of all involved inventories, the MRP takes care of scheduling production, back-propagating suitable production or purchase orders upstream. O. Wright introduced in 1983 the *Manufacturing Resource Planning* (MRP II), significantly improving the MRP by, among other things, accounting for rough-cut production capacity; see [17]. Latest trends attempt to include market demand prediction by time series modeling. See [14] for a modern, retrospective review, along with the significant improvements and advanced features added in recent times and collected under the denomination of *Demand Driven MRP* (DDMRP).

Since the 1980s, the *pull* logic (i.e., supply orders are issued downstream to upstream, only when supply is needed) has been widely adopted, whenever

applicable for the specific business. In 1981, Kimura et al. modeled a celebrated, self-regulated implementation with the so-called *Just in Time* (JIT) paradigm; see [9]. Particularly, very small production lots are handled at all productive levels, with production order cards (termed *kanban*, in Japanese language) traveling upstream, whenever the representative lot is exhausted, and back, together with a new lot to be stocked. Frequent and small supply actions, resembling a continuous flux, characterize this scenario. Still, the need arises to determine the required stock levels, that is, equivalently, the number of lots, or of *kanban* cards.

MRP, or DDMRP, or also JIT, allow remarkably reducing the stock levels needed to run the business. As a limit condition, goods should simply transit in the suitable quantities and at the suitable times throughout logistic hubs (for the sake of transport and distribution chain optimization), and without being stocked for longer times than those strictly needed for their handling. Anyway, some residual inventory quota is still needed, along with classic theories like ROP or ROT modeling their dynamics. This is basically to ensure a short term market coverage, so to cope with: Discrepancies between forecast and current demand; Client orders' withdrawal, or postponement, or expediting; Deviations from planned delivery dates along the productive chain (late, missed or erroneous deliveries, unforeseen production downs or transport problems, and the like); Fluctuations in production capacity, discrepancy between reference (usually estimated considering runner codes) and actual (on a code-by-code basis) production capacity, when rough-cut over infinite capacity is carried out; Discrepancy between estimated and actual defective product rates, and other possible mishaps.

Dynamics based approaches could be useful to address problems (A) and (B), by checking a posteriori, typically by means of *Discrete Event Simulation* (DES), the adequacy of a given, prescribed inventory sizing and then moving towards a (usually pseudo-) optimal sizing or, more frequently, simply towards a better-than-actual sizing by means of: Either scenario analysis (that is, by partial enumeration of possible alternatives); Or some improving strategy, consisting of either down-hill moves (derivative free optimization, surrogate modeling, etc.) or some heuristics (genetic algorithms, particle swarms, etc.). Drawbacks of such approaches include being time consuming and being strongly dependent on a multitude of system describing input data of difficult availability and sometimes questionable reliability. Moreover, accounting for the stochastic nature of the problem is possible but emphasizes the above mentioned limits; see, e.g., [11] for a comparison between DES and classic inventory modeling.

In order to avoid the mentioned drawbacks, we propose, in problem (A), to prescind from simulating the inventory evolution by means of some time-marching scheme but, rather, to focus onto the stochastic variable concept. Any artificial requirement of normally distributed probabilities is dropped. As an output, we get the performance vs. cost curves for all stockable goods. Secondly, by means of *Mathematical Programming* (MP) approaches, we deal with problem (B) as the combinatorial, constrained, optimization problem of finding the optimal positioning along the above mentioned curves for all stockable goods, possibly including the Boolean decision variable whether or not to stock some goods.

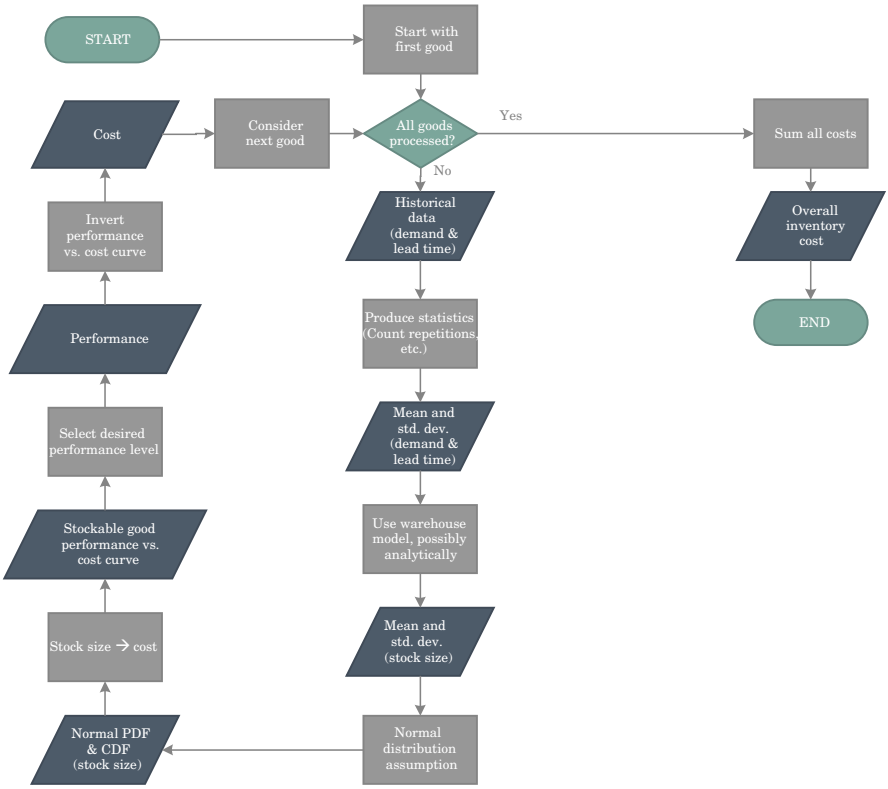

Fig. 1 Traditional warehouse sizing approach workflow

1.3 Traditional vs. Proposed Approach and Workflow

For the sake of a better comprehension, we provide a comparison between the traditional (see Fig. 1) and the proposed approach (see Fig. 2) to stock sizing. In both approaches, for each stockable good the required input data are the historical market demand and supply lead time time-series; see Sect. 2.2.

In the traditional approach to stock sizing, historical data are processed according to well-known statistical techniques, in order to obtain their mean value and standard deviation; see Sect. 2.2. Based on any applicable warehouse model (see, e.g., Sect. 2.1, or consider other models), the mean value and standard deviation for the required stock size are deduced, possibly analytically and usually with an underlying, implicit model linearization; see Sect. 2.3. Then, it is implicitly assumed that the required stock size be normally distributed, which needs not being the case, so to have available a probability density function (PDF), as well as a cumulated probability function (CDF). The latter represents the relationship between the stock size and the probability not to stock out. By converting the stock size into the

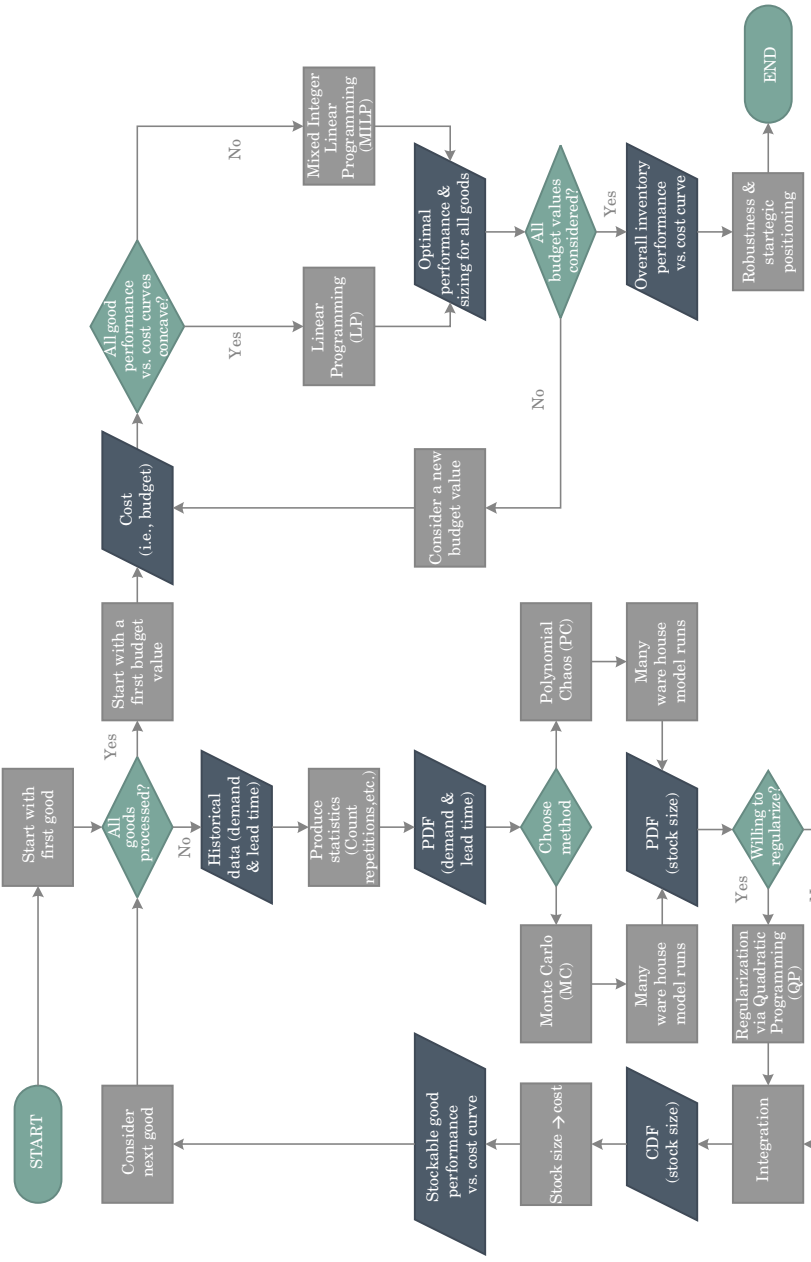

Fig. 2 Proposed warehouse sizing approach workflow

corresponding cost, a performance vs. cost curve is obtained for the stockable good at hand; see Sect. 4.1. Then, after a prescribed performance level has been selected, either arbitrarily or according to some suitable argument, from inverting the performance vs. cost curve the corresponding cost is found. Finally, the global inventory cost is obtained from the summation over all stockable goods.

The normality assumption is proven false by the finer approach proposed hereafter. This introduces both a theoretical and practical bias in the performance vs. cost curves, as well as in the computations following downstream. Moreover, there is no guarantee, a priori, that the global inventory cost be less than the available budget. In case not, the performance levels need being lowered, usually according to some trial-and-error scheme. Similarly, no guarantee is provided for optimal budget expenditure, meaning that if the global inventory cost is lower than the available budget, then the prescribed performance levels have to increased (and also in this case there is no direct indication on how much) so to spend the additional available resources.

In the proposed approach to stock sizing, historical data are processed according to well-known statistical techniques, in order to obtain their probability density function (PDF), not just their mean value and standard deviation; see Sect. 2.2. Then, either with Monte Carlo (MC) method (see Sect. 2.4) or with Polynomial Chaos (PC) expansion method (see Sect. 2.5), the probability density function for the required stock size is deduced numerically, in both cases by means of a number of runs of a suitable warehouse model (see, e.g., Sect. 2.1, or consider other models). Then, a Quadratic Programming (QP) based, mathematical regularization technique is strongly advised, in order to clean the PDF from rare events and outliers; see Sect. 3.2. By PDF integration, the relevant CDF is obtained for the stock size. The regularization attempts to (and usually manages to) produce a concave CDF. The latter represents the relationship between the stock size and the probability not to stock out. By converting the stock size into the corresponding cost, a performance vs. cost curve is obtained for the stockable good at hand; see Sect. 4.1.

We stress that, contrarily to the traditional approach, the performance vs. cost curve is not the outcome of a forced, and usually false, normality distribution assumption, neither for input data (market demand and supply lead time) nor for the output data (required stock size). Rather, all PDFs and CDFs are now either empirically deduced or produced throughout a mathematical model that attempts to describe the real course of events. The finer the model, the more reliable the probability distributions. The fallacies introduced by the traditional approach may be observed in Fig. 6, in the case of the same warehouse model (ROP).

Also the problem of optimal resource allocation, that is, optimal budget expenditure, is now addressed in a radically different and mathematically structured manner. After all stockable goods have been dealt with as above, a first global inventory budget is considered, for instance the actually available one. Depending on whether all stockable good performance vs. cost curves be concave or not, the optimal budget expenditure problem is addressed by either Linear Programming (LP; see Sect. 4.2) or Mixed Integer Linear Programming (MILP; see Sect. 4.3), respectively.

In both cases, the optimal cost and service level for each stockable goods is obtained, along with the overall weighted inventory performance and overall cost. Differently from the traditional approach, now the overall inventory cost is guaranteed to exactly equal the available budget (in case of feasible problems, for, otherwise, side constraints need being relaxed, as shown in Sect. 4).

A yet further advantage of the proposed approach is that the optimal budget expenditure problem is suggested to be solved for other, different budget values, in such a way to deduce, pointwise, an overall performance vs. cost curve for the whole inventory. This allows gaining a measure of the robustness of the problem (see whether or not little budget perturbations result into little performance perturbations). Robustness analysis then plays a key role in the strategic choice of global inventory budget allocation, because the collocation in a sharp bending point of the performance vs. cost curve may be very alluring.

It is important to notice that the particular choice of the warehouse model is largely immaterial for the proposed approach to stock sizing, meaning that the latter is general and may embed any other model with the same output. Also the case of different input data (number and kind) may be handled with straightforward modifications of inessential aspects. We mentioned the two simplest models historically and commonly used in engineering practice, developed the exemplification with reference to the one currently adopted in the industrial ABB case at hand, and finally proposed a simple extension that may apply to both, in order to address the MRP management policy. This has been done both for the sake of simplicity and in order to more strictly adhere to the traditional approach in all inessential aspects.

1.4 Chapter Outline

The approach proposed in this chapter may be outlined as follows. In Sect. 2 we address problem (A). First, in Sect. 2.1, basic ROP theory is recalled and given a formal dress compatible with standard probability theory. An additional contribution is considered that accounts for time discrete inventory control or, which is equivalent, for large market requests. Then, in Sect. 2.2, input data probability distributions are estimated, mainly based on historical data. Three alternative approaches are discussed to propagate uncertainty through ROP model and deduce the probability not to stock-out, namely: The traditional variance propagation approach, in Sect. 2.3; *Monte Carlo* (MC) method, in Sect. 2.4; *Polynomial Chaos Expansion* (PCE, or PC), in Sect. 2.5. In Sect. 3, the output from problem (A), that is, performance vs. cost curves, are improved before being used as input to problem (B). A suitable, *Quadratic Programming* (QP) based regularization technique is introduced, aimed at enforcing concavity in the performance vs. cost curves portion of interest. In Sect. 4 we address problem (B) as a combinatorial, constrained optimization problem. Two alternative approaches are discussed, namely: A simplified, and computationally very slim, *Linear Programming* (LP)

based optimization, in Sect. 4.2; A more refined and versatile, but computationally more challenging, *Mixed Integer Linear Programming* (LP) based optimization, in Sect. 4.3. The application of the procedure to the industrial case of ABB, the results obtained, the possible extension of the very same machinery to other contexts of industrial operations, including but not limited to production management, as well as possibilities for further research are finally discussed, in Sect. 5.

2 The Local Stochastic Problem

In this section we deduce, for each single good, the stochastic relationship, i.e., a probability density function (PDF), relating the probability not to stock-out with the amount of items to be stocked; see Sect. 1, problem (A). Item amounts are large when mass produced goods are at hand, allowing for real numbers to describe what would rigorously require integer numbers.

2.1 Reorder Point and Other Models

A classical inventory management policy for *make-to-stock* (MTS) goods is the so-called *reorder point* method, modeling stock level dynamics according to a saw-tooth behavior; see Fig. 3. Starting from an arbitrary, high value, the stock level for a given item is progressively lowered by *market demand D*, until a *reorder point*

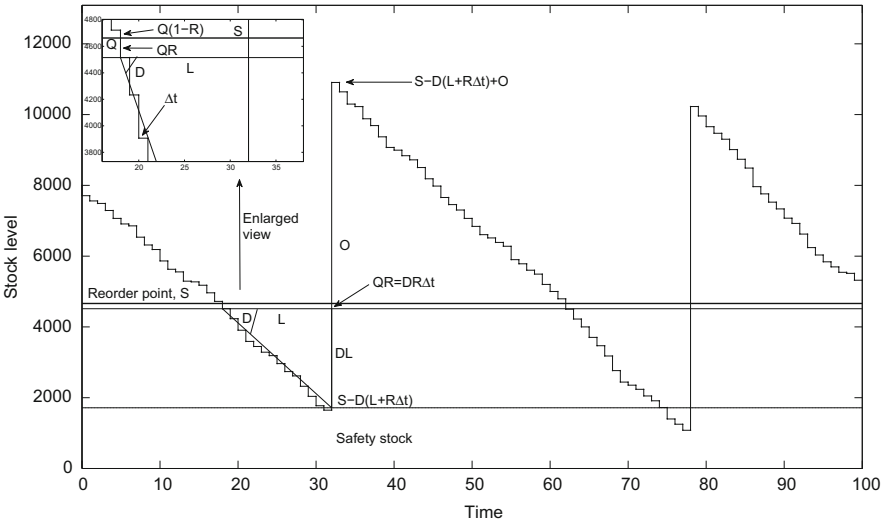

Fig. 3 Typical saw-tooth dynamics of a reorder point ruled inventory

threshold S is reached. A supply order is then issued upstream, and ordered goods eventually enter the warehouse after a *supply lead time* L during which time the stock level continues to go down below the reorder point, due to the last market requests received. Stock levels and reorders are dealt with at discrete time, i.e., they are spaced by a prescribed period of time Δt (a day, or an hour, etc., depending on the inventory type and dynamics). If Δt is non negligible compared to L, or equivalently, if market orders size may be large compared to S, then the last ordered quantity $Q = D\Delta t$ may result into a non negligible portion RQ below the safety stock, where $R \in [0, 1]$ is a proportionality coefficient. Elementary geometrical considerations (see Fig. 3, where the *safety stock* has to be initially neglected; as the name suggests, this additional quota is later introduced in order to raise the reorder point and cope with the stochastic nature of the problem) relate the reorder point to its contributors, reading

$$S = DL + RQ = D(L + R\Delta t). \tag{1}$$

In a deterministic frame, inventory dynamics would be known and repeatable, leading to the determination of the "exact", sharp reorder point S directly by plugging sharp D, L and R values into (1). Actually, all involved quantities are beyond our full control and may be considered stochastic variables. Market demand may be associated (written $D \sim \phi^D(\xi^D)$) to a *probability density function* (PDF) $\phi^D(\xi^D)$, where ξ^D is any admissible value for D, while $\phi^D(\xi^D)$ is the probability density associated with such value. Let $\Phi^D(\xi^D) = \int_{-\infty}^{\xi^D} \phi(\zeta)\,\mathrm{d}\zeta$ be the probability to draw values of D up to ξ^D, so that the relevant function is the *cumulated density function* (CDF). Similarly, $L \sim \phi^L(\xi^L)$, $R \sim \phi^R(\xi^R)$ and $S \sim \phi^S(\xi^S)$, with relevant CDFs. As a final result, the quantity S to be stocked is to be estimated as the value ξ^S such that

$$\Phi^S(\xi^S) = \int_{-\infty}^{\xi^s} \phi^S(\zeta)\,\mathrm{d}\zeta = 1 - \alpha, \tag{2}$$

that is, such that the reorder point S will be sufficient to cope with inventory dynamics with confidence level $1 - \alpha$. Equivalently, the probability to stock-out will be no higher than $\alpha \in [0, 1]$ (typically, α is a number of the order of 5–10%, or even less). In what follows, a mathematical method is proposed to derive the actual expression of (2), in the case above as well as in the similar conditions introduced hereafter.

Generalizations or other, different models are also available. It is similarly deduced that, for goods that follow the ROT policy,

$$S = D(L + T + R\Delta t), \tag{3}$$

where T is the (prescribed) *reorder time*, i.e., the fixed period of time after which the current stock level is compared with the reference value S and a reorder quantity is possibly requested upstream, so to restore level S. Most goods are nowadays

handled by MRP. To estimate the residual safety inventory quota that is practically still needed, we propose to use the model

$$S = \max\{0, (D - E[D])(L + R\Delta t)\} \tag{4}$$

or

$$S = \max\{0, (D - E[D])(L + T + R\Delta t)\} \tag{5}$$

instead of (1) or (3), resp., with market demand netted of its expected value $E[D]$ (i.e., its mean value). Netting demand D with $E[D]$, rather than with a different quantity, is an arbitrary but reasonable choice in absence of more precise information and in order to express any deviation from the reference scenario, for which nothing more than average values is accessible in practice. Since negative reorder point values would make no sense, the max operator is used as a lower cut-off with zero. For the sake of simplicity, we shall develop the following with reference to (1). It remains understood that methods and conclusions are to be rewritten, with suitable modifications, with reference to (3), (4) or (5) in relevant cases.

2.2 Input Probability Densities

Stock sizing is generally accounted for in a stochastic frame, to a different extent of mathematical subtlety and reliability, as later discussed. In any case, knowledge is needed about input variables probability distributions. The estimation of PDFs is generally not easy and requires care. As for R, owing to the random time sequencing of market purchase orders as long as the reorder point threshold is randomly crossed, there is no conceptual reason to privilege some values compared to others and one may assume a uniform distribution for ϕ^R over the interval [0, 1].

Supply lead time PDF ϕ^L could be estimated either by means of some suitable model describing its internal dynamics and policies (not covered here), or phenomenologically by counting the frequencies of supply within the terms of multiples of a suitably defined time unit like, e.g., days (or half days, or even less, depending on the reactivity of the supply structure). For instance, if the time unit is fixed to be the single day, then one may consider a sufficiently long historical series, covering, say, 6 months or a whole year, and then count how many times the supply lead time did not exceed 1 day, how many times it was in between 1 and 2 days, how many times in between 2 and 3 days, and the like.

In absence of more refined information, market demand PDF ϕ^D may be deduced from a rich enough historical data base of purchase orders received. A possible, proposed approach consists of considering three different time scales. First, we define a long enough period of time compared to inventory dynamics time constant like, e.g., 1 year or more (*long time scale*). It could be not advisable to consider longer durations, because markets are mutable and one wants to focus on present

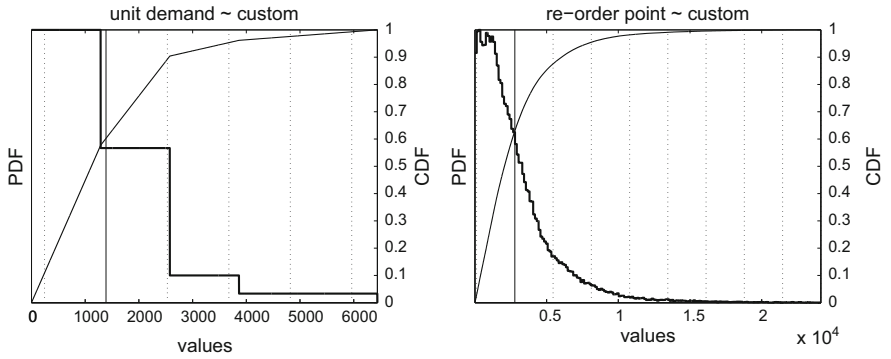

Fig. 4 Probability distributions for a sample product code in ABB inventory test case. *Left* Demand D, PDF ϕ^D (*bold*) and CDF Φ^D (*lite*), as empirically deduced from historical data records. *Right* Reorder point S, PDF ϕ^S (*bold*) and CDF Φ^S (*lite*), as computed by means of PC and fast MC on polynomial surrogate model. A remarkable difference is observed from the traditionally adopted Gaussian assumption: the actual PDF is single tailed and starts high-valued on the *left side*

trends. Second, the long period is subdivided into time buckets like, e.g., months, or weeks, or the like, so that their number is large enough to allow for a reasonable PDF resolution, as later discussed, and so that each is not too short to contain just a few orders (*middle time scale*). Finally, the third time scale is the time constant of the inventory dynamics like, e.g., the day or the like (*short time scale*). The latter can be the time scale with time constant Δt, and must be definitely smaller than the middle time scale. Market demand PDF ϕ^D requires cumulating all orders received over any single time bucket of the middle time scale and then counting the recurrence frequencies of ordered quantities, i.e., how many times the whole market required a quantity of pieces lying in any bin. See Fig. 4 (left) for the illustration of a sample product code in later discussed ABB inventory test case.

In all cases, histograms are obtained and PDFs are readily produced by normalization. Standard techniques must be applied in bin sizing, possibly considering variable and adaptive bin sizing, in order to guarantee statistical soundness. The approach above is admittedly poor, and more refined statistical methods could be conceived to enhance PDFs' quality. It must be anyway remembered that many other uncertainties and approximations affect the problem, so that extreme accuracy could be unmotivated and practically unattainable.

Large purchase orders, like, e.g., periodic and contract negotiated wholesalers' refills, need being detected and translated into the stemming multitude of smaller order actions spread over a long time horizon, as usually dealt with by logistic planning. Actually, since the latter actions are planned and thus known in advance, they should be handled in a deterministic way and removed from the stochastic sizing of the safety stock.

Some items may have such a rare and/or volatile demand, that the concept itself of probability density in the above sense would not properly hold. Such

and similar cases should not be considered in this modeling frame, nor stocked according to the logics of recurrent and abundant production and resell, i.e., they should not enter inventories. Rather, such special items should be more profitably supplied to the market according to suitably defined policies, and with terms and conditions negotiated with clients, like, e.g., *assemble-to-order* (ATO) or *make-to-order* (MTO).

2.3 Output Probabilities: Traditional Approach

The traditional approach for safety stock stochastic estimation is based on the linearization $S = (E[L] + \Delta t/2)D + E[D]L + E[D]R\Delta t - E[D](E[L] + \Delta t/2)$ of (1), obtained after straightforward simplifications from the relevant Taylor series expansion truncated to degree 1 terms and centered on point $(E[D], E[L], E[R] = 1/2)$, corresponding to the mean values of D, L and R. After the linearization, the mean value follows with a standard change of variables and reads

$$E[S] = \int_{\mathbb{R}} \xi^S \phi^S(\xi^S)\, \mathrm{d}\xi^S = E[D](E[L] + \Delta t/2). \tag{6}$$

Variance follows similarly, yielding the well-known variance propagation formula that in the specific case reads

$$\begin{aligned} \sigma_S^2 &= \int_{\mathbb{R}} (\xi^S - E[S])^2 \phi^S(\xi^S)\, \mathrm{d}\xi^S \\ &= (E[L] + \Delta t/2)^2 \sigma_D^2 + E[D]^2 \sigma_L^2 + (E[D]\Delta t)^2 \sigma_R^2, \end{aligned} \tag{7}$$

where D, L and R have been assumed to be mutually uncorrelated and σ_D, σ_L and $\sigma_R = 1/\sqrt{12}$, resp., are their standard deviations. When $\Delta t \to 0$ (i.e., Δt is small compared to L), the term $RD\Delta t$ is frequently neglected compared to the term DL in (1). Under such assumption, the mean value (6) reduces to $E[S] = E[D]E[L]$ while variance reduces to $\sigma_S^2 = E[L]^2\sigma_D^2 + E[D]^2\sigma_L^2$. This latter expression is to be compared with $\sigma_S^2 = E[L]\sigma_D^2 + E[D]^2\sigma_L^2$, as deduced by Hadley-Whitin [7]. When supply from the feeder factory is governed by reliable policies such as prescribed and easily attainable delivery dates, then supply lead time is considered deterministic and known equal to a fixed value L (here intended as a deterministic variable, no longer as a stochastic variable), furtherly simplifying the mean to $E[S] = LE[D]$ and variance to $\sigma_S^2 = L^2\sigma_D^2$, or to $\sigma_S^2 = L\sigma_D^2$ in the Hadley-Whitin form.

In the traditional approaches discussed above, the reorder point is estimated as

$$S = E[S] + c(\alpha)\sigma_S, \tag{8}$$

where $c(\alpha)$ is a suitable multiplicator of standard deviation σ_S, corresponding to confidence level $1 - \alpha$. The $E[S]$ contribution may be considered as a sort of theoretical deterministic prediction, conservatively supplemented by the $c(\alpha)\sigma_S$ contribution that accounts for stochastic uncertainties. The latter contribution is frequently termed *safety stock*.

The actual probability distribution ϕ^S of S is never introduced nor used. If, on the one hand, this makes the mathematical developments extremely simple, on the other hand $c(\alpha)$ is questionably computed by implicitly assuming that S be normally distributed. There is no theoretical reason leading to Gaussian ϕ^S (actually, theory clearly tells this cannot be the case, because ϕ^S must be boundedly supported on the left, i.e., it cannot have a left tail, in order to exclude logically inconsistent negative reorder point values). Evidence from the application of the more reliable and logically consistent methods discussed in next sections clearly shows that S is not normally distributed, and that the actual PDF may be remarkably different; see also Fig. 6 for a visual comparison.

2.4 Output Probabilities: Monte Carlo

We are interested in adopting (2) instead of (8) in order to estimate the quantity to be stocked. Consequently, ϕ^S is required. One way is the well-known *Monte Carlo* (MC) method. MC basically amounts to repeatedly evaluating the deterministic model (1), with input values D, L, R randomly extracted from relevant PDFs. The correspondingly many random output S values are then used as samples to produce ϕ^S. On the one hand, MC method is conceptually simple, easy to implement and non-intrusive, in the sense that the deterministic model (1) is used as a black-box. An additional benefit, by far the most important, is that the convergence rate does not depend on the number of stochastic variables (which does *not* mean that the number of necessary runs does not depend on the number—and type—of stochastic variables).

On the other hand, the convergence rate is extremely slow, so that the number of deterministic model runs is usually very high, thus requiring long computational times, unless the deterministic model is extremely fast. Additionally, the successive moments of the sought for PDF converge progressively slower, let alone particular cases enjoying special properties, such as odd/even parity, and the like. As a consequence, the correct shape of ϕ^S may be very slow to converge.

2.5 Output Probabilities: Polynomial Chaos

An alternative way to obtain ϕ^S is based on the so-called (non-intrusive) *Polynomial Chaos* (PC) method, for which we need a digression into *orthogonal polynomials*. It is well-known that a general probability density distribution $\phi^x(\xi)$ associated with a

stochastic variable x satisfies the hypotheses of a measure. Thus it naturally induces the inner product $(\cdot,\cdot)_x : L^2(\mathbb{R}) \times L^2(\mathbb{R}) \to \mathbb{R}$ such that

$$(u(\xi), v(\xi))_x := \int_{\mathbb{R}} u(\xi)v(\xi)\phi^x(\xi)\,\mathrm{d}\xi. \tag{9}$$

Scalar product (9) naturally induces the norm $\|\cdot\|_x : L^2(\mathbb{R}) \to \mathbb{R}$, such that $\|u(\xi)\|_x^2 = (u(\xi), u(\xi))_x$. Let $L^2(\mathbb{R};\phi^x)$ be the set of bounded functions with reference to norm $\|\cdot\|_x$. One can use scalar product (9) in the frame of Gram-Schmidt process applied to the sequence $\{\xi^k\}_{k=0}^{+\infty} = \{1, \xi, \xi^2, \ldots, \xi^k, \ldots\}$. The outcome is an *orthogonal polynomial sequence* (OPS), that is, a sequence $\{\psi_k^x(\xi)\}_{k=0}^{+\infty}$ of polynomials which are mutually orthogonal with reference to (9). Therefore, by construction, $(\psi_j^x, \psi_k^x)_x = \|\psi_k^x\|_x^2\delta_{j,k}$, $\forall j,k \in \mathbb{Z}_0$, where $\delta_{j,k}$ is Kronecker symbol. If desired, the OPS could be made ortho*normal*, to simplify notation and without modifying the essence of the problem. The above mentioned polynomials are readily seen to be linearly independent by a standard degree argument, leading to the conclusion that they form a basis for $L^2(\mathbb{R};\phi^x)$; see, e.g., [3] for the easy details.

Superscript x in ψ_k^x explicitly shows the descendence of the OPS from stochastic variable x, to which it is linked and specific. It can be shown that some of the most common PDFs, continuous or discrete, induce celebrated OPSs; see, e.g., [10, p. 37]. For instance, uniform PDF ϕ^R induces correspondingly an OPS which coincides with the family of Legendre polynomials. Considering reorder point theory, three stochastic input variables are involved, each leading to its own OPS. Another OPS is obtained from the three above mentioned OPS by tensorization, as follows. A vector $\boldsymbol{\xi} := (\xi^D, \xi^L, \xi^R)$ is defined to collect 3-ples of values from the three input variables. A multi-index $\mathbf{k} := (k^D, k^L, k^R)$ is introduced to collect the degrees of polynomials from the three OPS. Then, the OPS $\{\psi_{\mathbf{k}}(\boldsymbol{\xi})\}_{|\mathbf{k}|=0}^{+\infty}$ is defined such that

$$\psi_{\mathbf{k}}(\boldsymbol{\xi}) := \psi_{k^D}^D(\xi^D)\psi_{k^L}^L(\xi^L)\psi_{k^R}^R(\xi^R) = \prod_x \psi_{k^x}^x(\xi^x), \tag{10}$$

where $x \in \{D, L, R\}$ in the products here and below and where $|\mathbf{k}| := k^D + k^L + k^R$ is the degree of each polynomial resulting from the tensorization. A new inner product $(\cdot,\cdot) : L^2(\mathbb{R}^3) \times L^2(\mathbb{R}^3) \to \mathbb{R}$ is naturally induced which, in turn, naturally induces the norm $\|\cdot\| : L^2(\mathbb{R}^3) \to \mathbb{R}$, such that $\|u(\boldsymbol{\xi})\|^2 = (u(\boldsymbol{\xi}), u(\boldsymbol{\xi}))$. Orthogonality is inherited from orthogonality of the three OPS, so that

$$(\psi_{\mathbf{j}}, \psi_{\mathbf{k}}) = \int_{\mathbb{R}^3} \prod_x \left[\psi_{j^x}^x(\xi^x)\psi_{k^x}^x(\xi^x)\phi^x(\xi^x)d\xi^x\right] = \|\psi_{\mathbf{k}}\|^2\delta_{\mathbf{j},\mathbf{k}}, \tag{11}$$

where $\delta_{\mathbf{j},\mathbf{k}}$ generalizes Kronecker symbol to the present context, differing from 0 iff multi-indices $\mathbf{j}$ and $\mathbf{k}$ are equal in all of their portions. In (11) and in similar expressions, $\prod_x \phi^x(\xi^x) =: \phi(\boldsymbol{\xi})$ represents the joint probability density distribution of the three stochastic variables, in the hypothesis of independence. The extension

to the general, dependent case follows on the formal side by simply replacing the multiplicative distribution with the true joint PDF. Needless to say, practical difficulties would arise in both the theoretical and the empirical determination of the latter distribution.

A basis for $L^2(\mathbb{R};\phi^D) \otimes L^2(\mathbb{R};\phi^L) \otimes L^2(\mathbb{R};\phi^R)$ is now available. We look for the closest approximation in this function space to function (1) expressing the dependance of stochastic variable S from stochastic variables D, L, R. Then, one may expand S in a (generalized) Fourier series according to the basis of polynomials (10) as the *polynomial chaos expansion* (PCE)

$$S(\boldsymbol{\xi}) = \sum_{|\mathbf{k}|=0}^{+\infty} s_{\mathbf{k}}\, \psi_{\mathbf{k}}(\boldsymbol{\xi}) \simeq \sum_{|\mathbf{k}|=0}^{p} s_{\mathbf{k}}\, \psi_{\mathbf{k}}(\boldsymbol{\xi}), \tag{12}$$

where $p \geq 0$ determines a practically unavoidable truncation. The choice of p may be deduced a priori in simple cases like the one at hand (viz., $p = 2$ to get the exact expansion) or induced experimentally in more difficult ones, for instance by progressively increasing p until the marginal accuracy in the final results becomes negligible. Care must be taken, because some important modeling feature may require some high order term after a possibly long sequence of null lower order terms. A sound theoretical analysis of the kind of expected functional dependence is very important in order to deal with complex cases, and comparison with MC method in a simple and inexpensive test case may help solve the question.

After the expansion, the information content expressed by S resides in the sequence of coefficients $s_{\mathbf{k}}$, which, owing to orthogonality (11) of the basis, may be individually computed by orthogonal projection of S over each basis element. Precisely, scalarly multiplying both sides of (12) by $\psi_{\mathbf{k}}$ and recalling (11), one gets

$$s_{\mathbf{k}} = \frac{(S, \psi_{\mathbf{k}})}{\|\psi_{\mathbf{k}}\|^2}. \tag{13}$$

The scalar product in (13) requires numerically evaluating a multiple integral, a task that can be accomplished by means of a suitably defined Gauss quadrature formula; see [16]. One gets

$$(S, \psi_{\mathbf{k}}) = \int_{\mathbb{R}^3} S(\boldsymbol{\xi})\psi_{\mathbf{k}}(\boldsymbol{\xi})\phi(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} = \sum_{j=1}^{N} w_j S(\boldsymbol{\xi}_j)\psi_{\mathbf{k}}(\boldsymbol{\xi}_j)\phi(\boldsymbol{\xi}_j), \tag{14}$$

where N is the number of Gauss points while w_j and $\boldsymbol{\xi}_j$ are, resp., the jth Gauss point and weight, $j \in \{1, \ldots, N\}$. In simple cases like the one at hand, the necessary number of Gauss points allowing exact integration may be found a priori. In more difficult cases one may try heuristically, experimenting by progressively increasing N until the marginal accuracy increment becomes negligible. Each Gauss point requires a function evaluation

$$S(\boldsymbol{\xi}_j) = S(\xi_j^D, \xi_j^L, \xi_j^R) = \xi_j^D(\xi_j^L + \xi_j^R \Delta t). \tag{15}$$

Owing to the tensor nature of the approach, the number of model evaluations grows rapidly with the number of input stochastic variables. This problem prevents the method from being applied in cases with large numbers of input stochastic variables. Reduced schemes could be pursued, alternative to the full factorial scheme discussed above, that help mitigating the computational burden without excessively sacrificing accuracy. Still, the issue remains for large enough number of variables.

As soon as coefficients $s_{\mathbf{k}}$ have been computed, a basic, standard change of variable and orthogonality (11) easily lead (see, e.g., [10, p. 39]) to the analytical deduction of the expected value (i.e., the mean)

$$E[S] = \int_{\mathbb{R}} \xi^S \phi^S(\xi^S)\, \mathrm{d}\xi^S = s_{\mathbf{0}} \|\psi_{\mathbf{0}}\|^2 \tag{16}$$

and of the variance

$$\sigma_S^2 = \int_{\mathbb{R}} (\xi^S - E[S])^2 \phi^S(\xi^S)\, \mathrm{d}\xi^S = \sum_{|\mathbf{k}|=1}^{+\infty} s_{\mathbf{k}} \|\psi_{\mathbf{k}}\|^2 \simeq \sum_{|\mathbf{k}|=1}^{p} s_{\mathbf{k}} \|\psi_{\mathbf{k}}\|^2. \tag{17}$$

Notice that the actual expression of $\phi^S(\xi^S)$ is not required in order to compute mean and variance: (16) and (17) immediately allow removing the linearization drawback from traditional approaches to ROP; see Sect. 2.3. More elaborated expressions may be deduced for higher order moments. If, like in the case at hand, $\phi^S(\xi^S)$ becomes necessary for the following developments, it can be computed by, e.g., "fast" MC simulation adopting the last term of (12) as a surrogate model for the real $S(\boldsymbol{\xi})$ functional dependence (1). The CDF (2) is finally computed in a straightforward manner. See Fig. 4 (right) for the illustration of a sample product code in later discussed ABB inventory test case.

Clearly, the basic reorder point model (1) at hand, of natively polynomial kind, is so simple that, on the one hand, well suits the explanatory goal but, on the other hand, PCE (eventually requiring a "fast" MC) is not advantageous over a direct MC. Nonetheless, as soon as additional features are introduced into the model, thus complicating its analytical expression and possibly requiring the solution of ODE, or PDE, or even possibly resulting into a very complicated black box, like in the case of DES or other dynamics based inventory models, then PCE may outscore direct MC in terms of required model runs, with a remarkable computational gain.

3 Local Data Regularization via QP

The CDF (2) is potentially affected by artifacts deriving from the empirical sampling of the input data that lead to its deduction. Typically, spurious spikes in the originating PDFs may appear, due to rare events, i.e., with a recurrence period greater than the long time scale, but that nonetheless happened to manifest therein

by mere chance. If not suitably dealt with, that is, if data are kept as they are after sampling (resp., if they are removed), then rare event recurrence frequency would result artificially and misleadingly increased (resp., decreased). Owing to a reasonable underlying regularity assumption, the presence of such anomalies is easily spotted in PDFs by inspection.

Since spikes in PDFs result into concavity changes in CDFs, a mathematical regularization technique based on concavity healing is here proposed to remove or reduce the above mentioned artifacts, thus improving the quality of (2). We shall see, when dealing with the final global optimization problem, that the proposed regularization is also beneficial in reducing the computational effort; see Sect. 4. For this reason we privilege this approach compared to other possible ways of filtering. See Fig. 5 for an applicative example based on later discussed ABB inventory analysis.

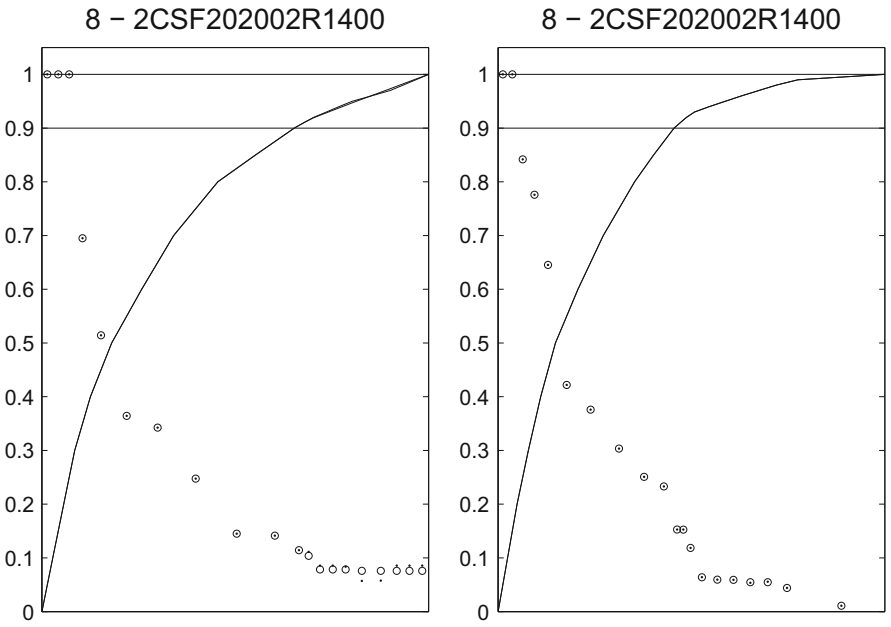

Fig. 5 Effects of QP regularization procedure over one product code in ABB inventory test case (downsampled curves are shown, for clarity). *Left* If PDF ϕ^S and CDF Φ^S are generated from a 10^2 runs MC analysis (extremely poor and inadequate) some differences can be noted between original (*dots*) and regularized (*circles*) PDFs, while the effect over CDFs (*solid*) is barely noticeable. *Right* With 10^3 runs MC originated curves, differences become barely noticeable also in PDFs. CDFs are already concave in the region of interest with 10^4 runs MC or higher, so that no regularization is needed. In ABB inventory test case, 10^5 runs MC are used

3.1 Concave CDFs

A general real function $y = f(x)$ of real variable x is convex over a (possibly unbounded) connex interval $U \subseteq D$, where D is its domain, iff $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$, $\forall x_1, x_2 \in U$, $\forall t \in [0,1]$, i.e., if the graph of function f over the interval between any two points x_1 and x_2 in U does not lie "above" the line segment from $(x_1, f(x_1))$ to $(x_2, f(x_2))$. A function f is concave over U if $-f$ is convex over U, i.e., if the graph does not lie "below" such segment. Linear and affine functions are immediately seen to be both convex and concave. A straightforward link with calculus shows that, in case of piece-wise twice differentiable functions, concavity coincides with non-positive second derivative. In other words, the first derivative needs being monotonically non-increasing over U in order f to be concave over U. See, e.g., [15] for in-depth discussion. As a consequence, when $f = \Phi^x$ is a CDF, concavity over U means that the relevant PDF $\phi^x = (\Phi^x)'$ be non-increasing over U.

In the context of ROP estimation, for a given stockable item the relevant PDF $\phi^S(\xi^S)$ is typically single tailed on its "right" portion, witnessing the increasing unlikeliness of extreme events. Along the tail, ϕ^S is non-increasing, by the definition of tail, so that Φ^S needs being concave over a suitable set U contained in the tail. Nonetheless, spurious and unphysical spikes may propagate through the computations above from empirical PDFs ϕ^D and ϕ^L to computed PDF ϕ^S. Spikes alter the natural non-increasing nature of ϕ^S, thus spoiling the concavity of Φ^S in its rightmost portion. In most cases, one could always envisage rare enough events that require a correspondingly longer historical data series than what experimentally available. As a consequence, spikes should be deemed a negative noise affecting data, to be removed for the sake of accuracy, and data regularization must be a due diligence before attacking the problem.

3.2 Concavity Enforcement

In the general case of a PDF $\phi(\xi)$ which is expected to be right-tailed, an a posteriori method is proposed to remove possible spikes by finding the right-tailed, spikeless, non-increasing PDF $\phi'(\xi)$ preserving mean value and minimizing a natural distance from the original one. After this found, the relevant CDF is immediately found by integration as $\Phi'(\xi) = \int_{-\infty}^{\xi} \phi'(\zeta)\,\mathrm{d}\zeta$. Application to ϕ^S and Φ^S will constitute a special case. According to the preceding developments, we may assume without loss of generality to deal with a general, compactly supported, *piece-wise constant* (PWC) function reading

$$\phi(\xi) = \sum_{j=1}^{n-1} \phi_j \chi_{[\xi_j, \xi_{j+1}]}(\xi), \tag{18}$$

where the convex support U has been partitioned into $n-1$ intervals $[\xi_j, \xi_{j+1}]$, for $j \in \{1, \ldots, n-1\}$, so that $\chi_{[\xi_j,\xi_{j+1}]}(\xi)$ is the characteristic function of $[\xi_j, \xi_{j+1}]$, equal to 1 if ξ belongs to such interval and null otherwise, while ϕ_j is the value constantly taken by $\phi(\xi)$ over such interval. An analogous expression holds for $\phi'(\xi)$.

As a result of integration, the CDF $\Phi(\xi) = \int_{-\infty}^{\xi} \phi(\zeta)\,\mathrm{d}\zeta$ is continuous and *piece-wise linear* (PWL). As such, it is characterized by the sequence of n vertices $\{(\xi_j, \Phi_j)\}_{j=1}^{n}$, where $\Phi_j = \Phi(\xi_j)$, so that $\phi_j = (\Phi_{j+1} - \Phi_j)/(\xi_{j+1} - \xi_j) \geq 0$, the latter sign constraint holding because PDFs are non-negative by definition. Let us collect known values ϕ_j, for $j \in \{1, \ldots, n-1\}$, into vector $\boldsymbol{\phi} \in \mathbb{R}_+^{n-1}$, where $\mathbb{R}_+$ denotes the non-negative real half line. Similarly, let us collect the sought for values ϕ'_j into vector $\boldsymbol{\phi}' \in \mathbb{R}_+^{n-1}$.

For the application at hand, only the final tail of the PDF is potentially affected by spikes. Therefore, an initial CDF part, up to a given number of points $m \leq n$, may be preserved as is. For any given m, the unknown vector $\boldsymbol{\phi}'$ is found by solving the *quadratic programming* (QP) problem

$$QP_m : \begin{cases} \min\limits_{\boldsymbol{\phi}' \in \mathbb{R}_+^{n-1}} g(\boldsymbol{\phi}') = \sum\limits_{j=1}^{n-1} \phi_j'^2 - 2 \sum\limits_{j=1}^{n-1} \phi_j \phi'_j \\ \text{subject to} \\ \sum\limits_{j=1}^{n-1} \phi'_j (\xi_{j+1} - \xi_j) = 1 \\ \sum\limits_{j=1}^{n-1} \phi'_j (\xi_{j+1}^2 - \xi_j^2) = \sum\limits_{j=1}^{n-1} \phi_j (\xi_{j+1}^2 - \xi_j^2) \\ \phi'_j = \phi_j, \qquad \forall j \in \{1, \ldots, m-1\} \\ \phi'_j \leq \phi'_{j-1}, \qquad \forall j \in \{m, \ldots, n-1\}, \end{cases} \tag{19}$$

now described in detail.

One obvious goal is not to excessively perturbate the basic PDF structure, that is, to aim at minimizing some reasonably defined distance between the original and regularized PDFs. A straightforward goal function is thus (the square of) the Euclidean distance $f(\boldsymbol{\phi}') := \|\boldsymbol{\phi} - \boldsymbol{\phi}'\|_2^2 = \sum_{j=1}^{n-1} (\phi_j - \phi'_j)^2$, to be minimized. The minimization problem is equivalently set with reference to goal function

$$g(\boldsymbol{\phi}') = f(\boldsymbol{\phi}') - \sum_{j=1}^{n-1} \phi_j^2 = \sum_{j=1}^{n-1} \phi_j'^2 - 2 \sum_{j=1}^{n-1} \phi_j \phi'_j = \boldsymbol{\phi}'^T \mathbf{I} \boldsymbol{\phi}' - 2 \boldsymbol{\phi}^T \boldsymbol{\phi}', \tag{20}$$

differing from $f(\boldsymbol{\phi}')$ by an inessential additive constant and being homogeneous in $\boldsymbol{\phi}'$. The quadratic form (20) is convex because the identity matrix $\mathbf{I}$ is positive definite. As a PDF mandatory requirement, ϕ'_j must be non-negative, $j \in \{1, \ldots, n-1\}$. This is explicitly enforced by searching for a minimizer $\boldsymbol{\phi}'$ in $\mathbb{R}_+^{n-1}$, i.e., the space obtained by the tensorization of $n-1$ copies of the non-negative half line.

The 0th moment (i.e., the area) of the sought for PWC PDF ϕ' reads

$$\int_{\mathbb{R}} \sum_{j=1}^{n-1} \phi'_j \chi_{[\xi_j, \xi_{j+1}]}(\xi)\, \mathrm{d}\xi = \sum_{j=1}^{n-1} \phi'_j \int_{\xi_j}^{\xi_{j+1}} d\xi = \sum_{j=1}^{n-1} \phi'_j (\xi_{j+1} - \xi_j). \tag{21}$$

Therefore, the first constraint in QP_m (19) traduces a linear (and thus convex) PDF normalization condition.

The first moment (i.e., the mean value) of the sought for PWC PDF ϕ' reads

$$\int_{\mathbb{R}} \xi \sum_{j=1}^{n-1} \phi'_j \chi_{[\xi_j, \xi_{j+1}]}(\xi)\, \mathrm{d}\xi = \sum_{j=1}^{n-1} \phi'_j \int_{\xi_j}^{\xi_{j+1}} \xi\, \mathrm{d}\xi = \sum_{j=1}^{n-1} \phi'_j \frac{\xi_{j+1}^2 - \xi_j^2}{2}. \tag{22}$$

Doing similarly with known PWC PDF ϕ and neglecting an inessential factor $1/2$ in both terms, the second constraint in QP_m (19) is obtained, traducing a linear (and thus convex) mean preserving condition.

Initial PDF portion preservation is easily enforced into problem QP_m as $\phi'_j = \phi_j$, $j \in \{1, \ldots, m-1\}$, meaning that $m-1$ variables are immediately resolved. If desired, such equalities may be used to a priori resolve up to $m-1$ variables and later deal with a reduced problem. The last part of the sought for PDF ϕ' is required to be monotonically non-increasing, which implies that CDF Φ' be concave over the range of interest. The requirement is immediately enforced in problem QP_m (19) as the linear inequality constraints $\phi'_j \leq \phi'_{j-1}$, $\forall j \in \{m, \ldots, n-1\}$. Clearly, other ranges than $j \in \{m, \ldots, n-1\}$ could be considered in different contexts, depending on the specific nature of the application at hand, without requiring conceptual modifications to the proposed technique. This completes the deduction of (19).

The constrained optimization problem QP_m (19) may be attacked as follows. As noted above, the goal function and all constraints are convex. It is well-known that a convex minimization problem either has a unique minimum (not necessarily a unique minimizer), or it is infeasible (empty admissible region). An initial guess $m = m_0$ value is chosen small enough so to cover a correspondingly large enough interval U. Even though, for a given m, there is no mathematical guarantee that the admissible region be non-empty, nonetheless it is clear that a non-increasing PDF with suitable mean value exists if $m = 0$, i.e., if any ϕ'_j is free to take a different value from the corresponding ϕ_j. Therefore, starting from $m = m_0$, if m is progressively decreased, so that a progressively larger number of degrees of freedom is introduced into the problem, then a feasible QP_m is eventually attained. The solution to problem QP_m corresponding to the first (i.e., largest) m for which this happens is retained as the final solution to the problem. In the unlikely event that, in order to find a feasible

QP_m, m needs being reduced excessively (say, below a given and reasonably defined threshold m'), then the decision may be drawn to abandon the iterative procedure earlier and to keep a non completely concave CDF. Under this latter respect, large possibilities exist for heuristic criteria.

A pseudo-code description of the proposed, iterative method reads

```
m=m0; while (QPm is not feasible AND m>m'), m=m-1; end
if (QPm is feasible) then
    solve QPm; substitute original PDF and CDF;
else
    keep original PDF and CDF;
end
```

where both feasibility detection and QP solution are addressed by standard, state of the art Mathematical Programming techniques; see [1, 4–6].

4 Global Warehouse Optimal Sizing via MP

Given a collection of curves expressing the relationship between ROP and confidence level not to stock-out, obtained according to the above developments, the final step consists in assigning suitable and reasonably different ROPs to all involved goods in stock (typically hundreds of goods, considering single product families individually), and consequently choosing suitable service levels for all items, so that the global (and suitably weighted) service level be maximized and under the constraint that the global cost comply with a global budget; see Sect. 1, problem (B).

4.1 Pareto Fronts

Let us consider the ith stockable item. A *key performance indicator* (KPI), or *service level*, is naturally expressed by the probability not to stock-out

$$P_i(\xi_i^S) = \Phi_i^S(\xi_i^S) = \int_{-\infty}^{\xi^S} \phi_i^S(\zeta)\,\mathrm{d}\zeta. \tag{23}$$

The *cost* of stock may be estimated based on the average stock level over time; see Fig. 3. Within the assumptions of the model adopted, as soon as the reordered quantity $O_i \geq E[D_i](E[L_i]+E[R]\Delta t)$ enters the stock, a level $S_i - D_i(L_i + R\Delta t) + O_i$ is attained, eventually decreasing linearly down to a level $S_i - D_i(L_i + R\Delta t)$ just before receiving the next reordered quantity. Such saw-tooth behavior is repeated indefinitely. Adopting average values for other variables than the reorder point S_i

and recalling that $E[R] = 1/2$, the average stock level[1] for the ith good is related to ξ_i^S as

$$\overline{S}_i = \xi_i^S + \frac{O_i}{2} - E[D_i]\left(E[L_i] + \frac{\Delta t}{2}\right). \tag{24}$$

Stock costs can then be expressed as

$$C_i(\xi_i^S) = C_i^a + C_i^u \overline{S}_i = C_i^f + C_i^u \xi_i^S, \tag{25}$$

where $C_i^a \geq 0$ is an *activation cost*, i.e., independent of the stocked quantity, while $C_i^u > 0$ is a *unit cost*, so that $C_i^u \overline{S}_i$ is a cost proportional to the average stock level $\overline{S}_i$ and, finally, $C_i^f = C_i^a + C_i^u(O_i/2 - E[D_i](E[L_i] + \Delta t/2))$ is a *fixed cost* with reference to variable ξ_i^S. The unit cost C_i^u is intrinsically non-null, for otherwise the stock problem for the relevant good would be trivially solved by assigning an arbitrarily large reorder point, a meaningless case in the applicative context. Indeed, $C_i^u \neq 0$ implies that map (25) be bijective and thus invertible. Moreover, the inverse map C_i^{-1} is still affine and such that $\xi_i^S = (C_i - C_i^f)/C_i^u$.

According to the commutative diagram

$$\begin{array}{ccccc} C_i \in & \mathbb{R} & \xrightarrow{C_i^{-1}} & \mathbb{R} & \ni \xi_i^S \\ & & \searrow^{F_i := (C_i^{-1})^*(\Phi_i^S) = \Phi_i^S \circ C_i^{-1}} & \downarrow{\scriptstyle \Phi_i^S} & \\ & & & [0,1] & \ni P_i \end{array}$$

it is possible to translate (23) and (25) into a parametric description of the $P_i = F_i(C_i)$ curve (*performance vs. cost curve*; see, e.g., Fig. 6), where $F_i := (C_i^{-1})^*(\Phi_i^S) = \Phi_i^S \circ C_i^{-1}$ is the *pullback* of Φ_i^S. Notice that concavity (as well as convexity) is invariant under composition with affine maps, that is, if $g = f \circ l$, where f is concave and l is affine, then

$$\begin{aligned} (f \circ l)(tx_1 + (1-t)x_2) &= f(l(tx_1 + (1-t)x_2)) \\ &= f(tl(x_1) + (1-t)l(x_2)) \\ &\geq tf(l(x_1)) + (1-t)f(l(x_2)) \\ &= t(f \circ l)(x_1) + (1-t)(f \circ l)(x_2), \end{aligned}$$

so that also g is concave. Therefore, since both C_i and its inverse C_i^{-1} are affine, concavity of F_i is equivalent to concavity of Φ_i^S. Concavity in case of performance vs. cost curves is reasonably justified empirically: Saturation frequently occurs in

[1] In case of models (3)–(5), similar considerations lead to $\overline{S}_i = \xi_i^S - E[D_i](E[L_i] + \Delta t/2 + T/2)$, $\overline{S}_i = \xi_i^S + O_i/2$ and $\overline{S}_i = \xi_i^S$, resp.

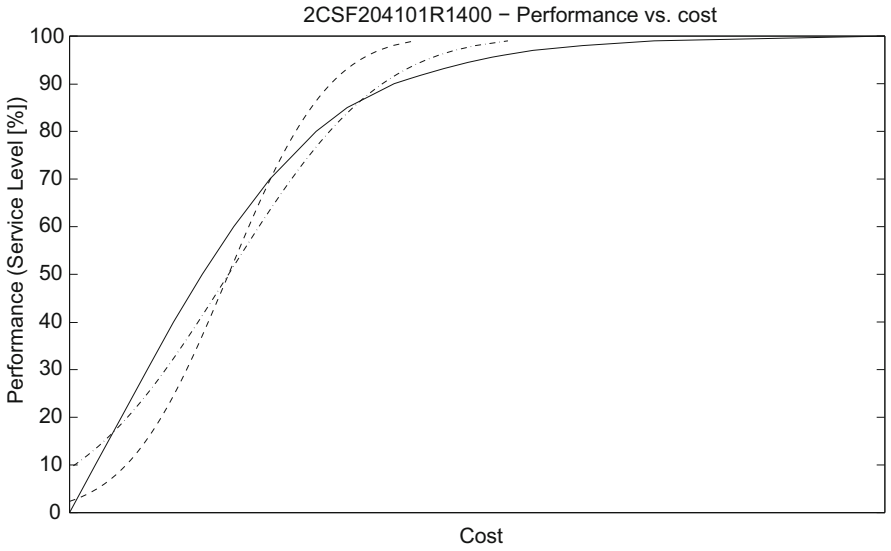

Fig. 6 Performance vs. cost curve for a sample product code, computed according to the proposed method, i.e., by means of a nonlinear model and without any normality assumption (*solid*). Also shown are the curves computed according to the traditional normality assumption and with standard deviation computed by means of variance propagation, i.e., with model linearization, (*dash–dot*) or with Hadley-Whitin formula (*dashed*); non negligible deviations from the correct behavior are observed in the latter two curves

many economical systems, so that the same marginal cost increment results into progressively smaller marginal performance increments as the performance level increases. The performance vs. cost curve for a sample product code is shown in Fig. 6, where the discrepancy between proposed and traditional approach can be appreciated.

Performance vs. cost curves satisfy *Pareto front* requirements. Precisely, any given point (C_i, P_i) on the curve is optimal in the sense that with a given cost C_i it is not possible to obtain a better performance than $P_i = F_i(C_i)$, whereas lower performances than P_i are unjustified, since the system has the capability to behave better. Dually, to attain a given performance P_i it is not possible to spend less than C_i, whereas to spend more is economically unjustified. In conclusion, points above the front are infeasible and those below are uneconomical; see Fig. 8, where the same considerations are illustrated for the whole inventory Pareto front, to be later deduced.

We consider now the problem of choosing a suitable positioning somewhere along the front. This amounts to solving a global optimization problem over a collection $\{(G_i, w_i)\}_{i=1}^{N}$ of N stockable goods G_i, each one with a relevant performance vs. cost curve and additionally provided with a global weight $w_i \in [0, 1]$, characterized

in that $\sum_{i=1}^{N} w_i = 1$ and expressing the good's relevance in the frame of the global weighted average performance

$$\overline{P} = \sum_{i=1}^{N} w_i P_i. \tag{26}$$

Weights are possibly chosen equal to $1/N$ for all goods, whenever no distinction has to be made.

4.2 Problem Formulation: LP

Let us first assume that all $P_i = F_i(C_i)$ curves are concave, $i \in \{1, \ldots, N\}$. Local data regularization via QP may be used to force non concave cases to the assumption at hand; see Sect. 3. Let us collect costs C_i and performances P_i into vectors $\mathbf{c}$ and $\mathbf{p}$, resp. Then, the unknown vectors $\mathbf{c}$ and $\mathbf{p}$ are found by solving the *linear programming* (LP) problem

$$LP: \begin{cases} \max\limits_{\mathbf{c},\mathbf{p}\in\mathbb{R}^N_+} \quad \sum\limits_{i=1}^{N} w_i P_i \\ \text{subject to} \\ \sum\limits_{i=1}^{N} w_i P_i \geq P^l \\ \sum\limits_{i=1}^{N} C_i \leq B \\ a^C_{i,j} C_i + a^P_{i,j} P_i \leq b_{i,j}, \quad \forall i \in \{1, \ldots, N\}, \forall j \in \{1, \ldots, n-1\} \\ C^l_i \leq C_i \leq C^u_i, \qquad \forall i \in \{1, \ldots, N\} \\ P^l_i \leq P_i \leq P^u_i, \qquad \forall i \in \{1, \ldots, N\}, \end{cases} \tag{27}$$

now described in details.

The linear goal function is the weighted average performance (26), to be maximized. A (possibly null) minimal admissible global performance P^l is enforced by the first constraint in problem *LP* (27). A global budget constraint is expressed by the second constraint in problem *LP* (27), where B is a given global cost budget.

Notice that, owing to the joint constraining action of the two inequalities above, the problem is either infeasible or a performance non worse than P^l is obtained. In the former case, a problem revision is needed in order to make the problem feasible, and either the minimal performance P^l is reduced or the maximal budget B is increased.

The actual shapes of performance vs. cost curves have to be enforced. Since we deal with PWL curves, the information relevant to the ith such curve is expressed by the collection of n vertices $\{(C_{i,j}, P_{i,j})\}_{j=1}^{n}$ the Pareto front consists of (notice that index i runs over goods while index j runs over the curve vertices for the ith good). The equation of the line through any two consecutive vertices $(C_{i,j}, P_{i,j})$ and $(C_{i,j+1}, P_{i,j+1})$ along the ith front is of the kind

$$a_{i,j}^{C} C_i + a_{i,j}^{P} P_i = b_{i,j}, \quad j \in \{1, \ldots, n-1\}, \tag{28}$$

where $a_{i,j}^{C} := P_{i,j} - P_{i,j+1}$, $a_{i,j}^{P} := C_{i,j+1} - C_{i,j}$ and $b_{i,j} := C_{i,j+1} P_{i,j} - C_{i,j} P_{i,j+1}$. Since the ith front is concave by assumption, the plane portion "below" the front is the locus of points (C_i, P_i) "below" the plurality of all lines (28), as expressed by the third collection of inequality constraints found in problem *LP* (27); see Fig. 7. The mathematical characterization of the Pareto front, as discussed in Sect. 4.1, will force (C_i, P_i) to adhere to the front for, otherwise, a more economical solution and with same performance could be found (or, dually, a better performing solution and with the same cost). Finally, lower bounds C_i^l (resp., P_i^l) and upper bounds C_i^u (resp.,

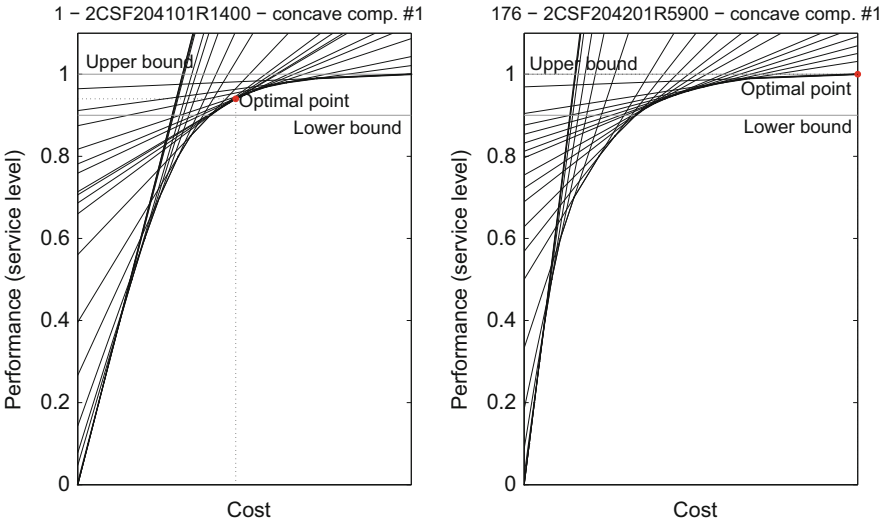

Fig. 7 PWL performance vs. cost curves (pareto fronts) for two sample stockable goods (first and last moving in the inventory), as rendered by means of a system of linear inequalities in LP problem (27); minimally and maximally acceptable service levels are also shown as *horizontal lines* defining the front portion of interest (*bold*) along which the optimal point is restrained to reside

P_i^u) could be imposed onto C_i (resp., P_i), as expressed by the remaining constraints in problem *LP* (27).

Different inventory management policies may modify the restraint set, retaining only part of the above. The LP problem is solved with standard, state of the art Mathematical Programming techniques, such as the well-known and efficient *simplex method*; see, e.g., [2].

4.3 Problem Formulation: MILP

As an extension of Sect. 4.2, let us now address the general case in which some or all performance vs. cost curves are non concave. Unfortunately, this case is mathematically way harder than the concave case. To understand why, one must recall that each pair (C_i, P_i) of variables in LP problem (27) belongs to the region below the corresponding concave performance vs. cost curve, including the latter curve as part of the boundary. (In the optimal solution, the variables do belong, pairwise, to such boundary curves.) Now, since the performance vs. cost curve is concave, the feasible region for each pair of variables is convex. Additionally, as detailed above, the feasible region is described by a collection of linear inequalities.

Both of these advantages are lost in the general, non-concave case. Since the intersection of convex regions is convex and since half planes delimited by lines are convex regions, if one should now try to describe the region below a non-concave performance vs. cost curve by means of linear inequalities, a convex region would be obtained, which is not the case by assumption. Nonetheless, the attack strategy here proposed consists of partitioning each generally non-convex feasible region into convex parts. Then, the optimal solution must reside in one, and only one, of such parts. This fact may be handled by introducing a potential pair of variables for each part, along with Boolean, disjunctive variables. The latter are used to express the constraint that one, and only one, of such potential pairs of variables be active, all the others being "phantoms". The contributions of phantoms must be suitably removed, since such fictitious points are only introduced for the sake of convenience, but they are not "real". It is easily understood that the disjunctive nature of the variables and constraint make the problem combinatorial. The mathematical tool of integer programming is then required, with a remarkable increase of both the theoretical complexity and, above all, of the computational burden.

We now show the proposed approach in details. Due to its PWL nature, it is always possible, without loss of generality, to assume that the generic ith performance vs. cost curve consist of K_i concave portions. If such curve is originally concave, then $K_i = 1$. As an extreme case, $K_i = n-1$ and concave portions coincide with the linear spans of the PWL function. With reference to the ith performance vs. cost curve, let us introduce a collection $\{(C_{ik}, P_{ik})\}_{k=1}^{K_i}$ of the sought for points, the kth of which belonging to the kth concave portion. Vectors $\mathbf{c}$ and $\mathbf{p}$ need being suitably enlarged so to host all of the so introduced unknown variables for all curves, for a total of $K = \sum_{i=1}^{N} K_i$ entries. Let us also assume that the kth portion of the ith

curve extend over the $[C_{ik}^l, C_{ik}^u)$ interval, so that the "right"-most extreme coincide with the "left"-most extreme of the next interval.

Additionally, Boolean decision variables are introduced, so that $z_{ik} \in \mathbb{Z}_2 \cong \{0, 1\}$ be associated with the kth concave portion of ith stockable item performance vs. cost curve. The z_{ik}'s are collected into vector $\mathbf{z} \in \mathbb{Z}_2^K$, where the latter space $\mathbb{Z}_2^K$ is the lattice of multidimensional points obtained by tensorization of K copies of $\mathbb{Z}_2$. The basic idea is that, if $z_{ik} = 1$, then the point on the ith Pareto front lies on its kth concave portion (the *active portion*), and $z_{ik} = 0$ for all other k's. For the sake of convenience, the points (C_{ik}, P_{ik}) relevant to unused concave portions are still formally present, and they are conventionally located at the "left"-most point of such portion, i.e., $(C_{ik}^l, F_i(C_{ik}^l))$. Suitable terms will consequently arise in the mathematical description, so to compensate the presence of such "phantom" points.

The unknown vectors $\mathbf{c}$, $\mathbf{p}$ and $\mathbf{z}$ are found by solving the *mixed integer linear programming* (MILP) problem

$$
MP : \begin{cases}
\max\limits_{\mathbf{c},\mathbf{p}\in\mathbb{R}_+^K,\mathbf{z}\in\mathbb{Z}_2^K} \quad \sum\limits_{i=1}^{N}\sum\limits_{k=1}^{K_i} w_i P_{ik} - \sum\limits_{i=1}^{N}\sum\limits_{k=1}^{K_i} w_i F_i(C_{ik}^l)(1 - z_{ik}) \\
\text{subject to} \\
C_{ik} - (C_{ik}^u - C_{ik}^l) z_{ik} \le C_{ik}^l, \\
\qquad \forall i \in \{1, \ldots, N\}, \forall k \in \{1, \ldots, K_i\} \\
\sum\limits_{i=1}^{N}\sum\limits_{k=1}^{K_i} w_i P_{ik} - \sum\limits_{i=1}^{N}\sum\limits_{k=1}^{K_i} w_i F_i(C_{ik}^l)(1 - z_{ik}) \ge P^l \\
\sum\limits_{i=1}^{N}\sum\limits_{k=1}^{K_i} C_{ik} - \sum\limits_{i=1}^{N}\sum\limits_{k=1}^{K_i} (1 - z_{ik}) C_{ik}^l \le B \\
a_{ik,j}^C C_{ik} + a_{ik,j}^P P_{ik} \le b_{ik,j}, \\
\qquad \forall i \in \{1, \ldots, N\}, \forall j \in \{1, \ldots, n-1\}, \forall k \in \{1, \ldots, K_i\} \\
\sum_{k=1}^{K_i} z_{ik} = 1, \qquad \forall i \in \mathscr{I}_{mh} \\
\sum_{k=1}^{K_i} z_{ik} \le 1, \qquad \forall i \notin \mathscr{I}_{mh} \\
C_{ik}^l \le C_{ik} \le C_{ik}^u, \qquad \forall i \in \{1, \ldots, N\}, \forall k \in \{1, \ldots, K_i\} \\
P_i^l \le P_{ik} \le P_i^u, \qquad \forall i \in \{1, \ldots, N\}, \forall k \in \{1, \ldots, K_i\},
\end{cases} \tag{29}
$$

now described in details.

The basic structure of problem *MP* (29) is clearly inherited from problem *LP* (27). Notice that the maximizer is sought for in a multidimensional space, obtained by tensorization, where $\mathbf{c}$ and $\mathbf{p}$ are both in $\mathbb{R}_+^K$, and thus the relevant entries must be non-negative, while $\mathbf{z}$ is in $\mathbb{Z}_2^K$, and thus the relevant entries are only allowed to take either value 1 or 0.

The penultimate collection of constraints forces C_{ik} to belong to the relevant concave portion. Let us consider the first collection of constraints. In case the kth is the ith good active portion, then $z_{ik} = 1$ and the constraint reduces to $C_{ik} \leq C_{ik}^u$, that is, a redundancy. Otherwise, $z_{ik} = 0$ and the constraint reduces to $C_{ik} \leq C_{ik}^l$. The only possibility is thus $C_{ik} = C_{ik}^l$, so that phantom points are forced to coincide with the "left"-most point of the relevant concave portion. The goal function is still the weighted average global performance (26), with the additional summation compensating phantom points. As a matter of fact, the active portions do not contribute to such additional summation ($1 - z_{ik} = 0$), whereas all other potions remove a $w_i F_i(C_{ik}^l)$ contribution from the global performance ($P_{ik} = F_i(C_{ik}^l)$ for phantom points).

The same reasoning applies to the minimal performance P^l constraint and to the maximal budget B constraint ($C_{ik} = C_{ik}^l$ for phantom points). As a matter of fact, also in this problem a (possibly null) minimal admissible global performance P^l is enforced by the second constraint in problem *MP* (29), and a global budget constraint is expressed by the third constraint in problem *MP* (29), where B is a given global cost budget. Notice that, exactly like in problem *LP* (27) and owing to the joint constraining action of the two inequalities above, the problem is either infeasible or a performance non worse than P^l is obtained. In the former case, a problem revision is needed in order to make the problem feasible, and either the minimal performance P^l is reduced or the maximal budget B is increased.

Since discrete decision variables have to be introduced in order to handle the non-concave case, it is worth taking advantage of their modeling power and introduce the possibility to decide whether to stock or not some goods, based on global optimality considerations. Precisely, let us introduce an index set $\mathscr{I}_{mh}$ of *must-have* goods, necessarily to be stocked for any relevant strategic reason. Obviously, the subset $\mathscr{I}_{mh}$ can possibly coincide with the universe of stockable goods, if desired. All other goods are subject to possibly being excluded from the MTS inventory management policy, depending on the solution of problem *MP* (29). Must-have goods ($i \in \mathscr{I}_{mh}$) are characterized in that *exactly one* concave portion is active (i.e., exactly one $z_{ik} = 1$, all others being null), while all other goods ($i \notin \mathscr{I}_{mh}$) are characterized in that *at most one* concave portion is active (i.e., possibly one $z_{ik} = 1$ and not more, all others—possibly all—being null). Such conditions are mathematically enforced by means of the fourth last and third last constraints, resp.

The resulting problem, especially after the data regularization procedure discussed in Sect. 3, is usually lean enough to be readily handled with standard branch-and-bound techniques on standard computers; see, e.g., [12].

5 Conclusions, Extensions and Future Research

The proposed approach to stochastic optimal sizing of inventories is the result of activities that the present author has undertaken in order to support the rationalization of ABB Low Voltage Products Division warehouse located in Vercelli, Italy, and serving the local national market from ABB factories located nationwide and abroad. Global optimization according to Sect. 4 leads to a global Pareto front, at inventory level, obtained pointwise by varying the global cost budget B in (27) or (29); see Fig. 8.

As for problem (A), i.e., reorder point based performance vs. cost curve determination, the proposed approach goes beyond ABB's state of the art (i.e., Hadley-Whitin formula) in that the stochastic problem is not forcedly linearized and in that the assumption that involved probability density functions be Gaussian is relaxed to handling generic, experimental distributions, deduced on a sampling

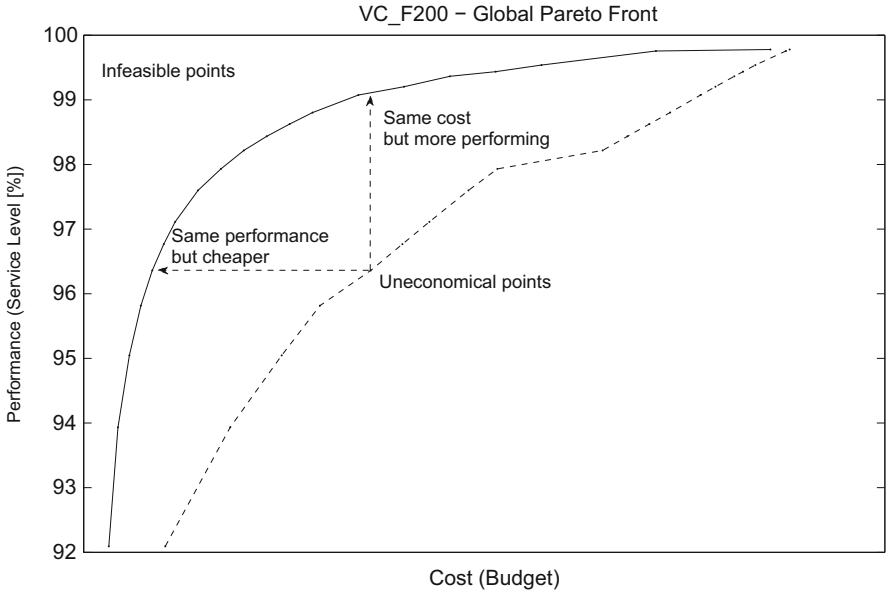

Fig. 8 Global Pareto front (*solid*) for a portion of ABB Italy—Low Voltage Products Division warehouse, relating global costs (sum over 176 different products) and overall weighted service level; *points above the front* are infeasible, while *points below* are uneconomical, because others may be found on the front with same performance and cheaper or, dually, with same cost and better performing. In case no global optimization is carried out but, rather, the same prescribed service level is assigned to all goods, then, by varying the prescribed service level, a remarkably lower curve is obtained (*dashed*), inside the uneconomical region. As expected, the optimal and non-optimal curve converge at the lowest admissible (90%) and highest possible (100%) service level, where exactly the same quantities are stocked for all goods, while inoptimality is greater elsewhere

basis. A posteriori, it is found that involved distributions are actually very far away from being Gaussian.

As for the global optimization problem (B) of locating working points for all involved goods along relevant performance vs. cost curves, a very fast approach is here proposed, based on forced regularization of non concave curves and LP, along with an alternative, more refined, but also potentially more cumbersome approach based on MILP. ABB priorly followed a rule-based approach following an ABC categorization of goods, intrinsically "local" and with no guarantee for global optimality.

The whole mathematical machinery has been condensed into a Matlab package currently available to ABB professionals operating in Logistics. This application reads suitably formatted data extracted from the Company's ERP (basically, the last 12 months movements, on a rolling basis), processes them, and finally delivers suggested reorder points, along with graphical indication of performance vs. cost curves, on a good-by-good basis.

The proposed methodology for the global optimization of systems consisting of a collection of performance vs. cost curves (see Sect. 4) could be extended to other systems than just inventories, by abstracting from what "performance" and "cost" actually represent. Examples in the world of industrial operations include finding the optimal strategy for running a multi-commodity productive system, where the performance may be the volumes of each commodity produced, and the cost may be some suitable measure of allocated resources, such as, e.g., manpower, worked hours, externally outsourced work, and the like, each one being usually restrained to a globally available budget. Many other interesting applicative fields could be easily spotted, with different interpretation of involved quantities but with the same or similar mathematical structure.

Performance vs. cost curves may also have a different origin than the one discussed above for inventories, including empirical curves. Concavity driven data regularization (see Sect. 3), may still be adopted in order to simplify the global optimization problem.

Applications may be envisaged requiring large and thus challenging instances of the MILP problem to be solved. Therefore, devising highly efficient and robust approaches to problem MP (29) could find sound motivation. Future research may go beyond general purpose branch-and-bound and seek for an attacking strategy possibly exploiting the specific mathematical structure of the problem. Cumulating a plurality of additional constraints than those examined here may also be an interesting and motivated future research effort. Last, but not least, a deeper statistical analysis and handling of input data could be highly beneficial. The method hereby proposed could easily be adapted in order to be compatible with other than PWC PDFs.

References

1. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming – Theory and Algorithms, 3rd edn. Wiley, New York (2006)
2. Bazaraa, M.S., Jarvis, J.J., Sherali, H.D.: Linear Programming and Network Flows, 4th edn. Wiley, New York (2010)
3. Chihara, T.S.: An Introduction to Orthogonal Polynomials. Dover, New York (2011)
4. Coleman, T.F., Li, Y.: A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. SIAM J. Optim. **6**, 1040–1058 (1996)
5. Gill, P.E., Murray, W., Wright, M.H.: Practical Optimization. Academic, London (1981)
6. Gould, N., Toint, P.L.: Preprocessing for quadratic programming. Math. Program. Ser. B **100**, 95–132 (2004)
7. Hadley, G., Whitin, T.M.: Analysis of Inventory Systems. Prentice-Hall, Englewood Cliffs, NJ (1963)
8. Harris, F.H.: How many parts to make at once. Fact. Mag. Manag. **10**, 135–136 (1913)
9. Kimura, O., Terada, H.: Design and analysis of pull system, a method of multistage production control. Int. J. Prod. Res. **19**, 241–253 (1981)
10. Le Maître, O.P., Knio, O.M.: Spectral Methods for Uncertainty Quantification. Springer, Berlin (2010)
11. Murphy, E.: A comparison of inventory optimization and discrete-event simulation for supply chain analysis. Simulation Conference, 2007 Winter, IEEE, New York (2007)
12. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. Wiley, New York (1999)
13. Orlicky, J.: Material Requirements Planning – The New Way of Life in Production and Inventory Management. McGraw-Hill, New York (1975)
14. Ptak, C., Smith, C.: Orlicky's Material Requirements Planning. McGraw-Hill, New York (2011)
15. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ (1970)
16. Szegö, G.: Orthogonal Polynomials. American Mathematical Society Colloquium Publications, vol. XXIII, 4th edn. American Mathematical Society, Providence, RI (1975)
17. Wright, O.: Manufacturing Resource Planning – MRP II, Revised edn. Wiley, New York (1996)
18. Zipkin, P.: Foundations of Inventory Management. McGraw-Hill, New York (2000)