Risk evaluation of a spare part stock by stochastic simulation

P.-E. Hagmark & H. Pernu
Tampere University of Technology, Tampere, Finland

ABSTRACT: Spare part inventories are necessities in keeping equipment in operating condition. The design of a spare part inventory is risk management by nature. It is a multi-phase task to meet the associated economical and technical requirements. A typical target is to optimize the size of the stock by balancing the costs against the stock-out risk. This paper is a brief presentation of an effort toward a broad-based methodology. Our model can be characterized as a simulation-calculation scheme that connects organically inventory design, process redundancy, maintenance policy, spare part reliability & demand, and shortage & cost accounting. The multitude of variables and concepts in our model enables a vast number of interpretations. Since the sub-models imitate the reality in time order, the model is easy to extend when new knowledge or new features need to be taken into account. The corresponding and still developing software extends continuously the applicability. At present, a tool for the optimization of selected cost combinations has been implemented.

KEYWORDS: Inventory design and control, Risk management, Spare parts, Discrete-event simulation.

1 INTRODUCTION

Spare part inventory control and management has been intensively researched for many decades. Extensive literature and research reports have been published. Kennedy et al. (2002) have provided a comprehensive literature overview. A large part of the reported methods seem to be case specific and restrictive in scope. The model in question consists often of a few analytic-numeric formulas and a small number of variables.

Simulation-based and more versatile models exist, but a more comprising methodology would be desirable. Our model contributes to a generic approach. The kernel, a time-ordered simulation-calculation scheme (Fig. 1), is readily open for extensions and new details, and not so exposed to distorting and restricting assumptions as analytic-numeric methods. The inventory policy is based on continuous review. The actors are one stock, one critical part (or group of parts), one or more part suppliers, and one or more part consumers (customers).

We begin with a superficial description following Figure 1. The first module, *PartRel*, offers several methods for the construction of probability distributions describing the failure tendency of the part at different stress levels.

The calculation of spare part demand is then based on a thorough model of the technical system. *PrepA* introduces a 'cluster structure' with a multiplicity of parameters for the definition of failure logic, stress, redundancies, operation and maintenance strategies, etc. This leads to each customer's consumption of spare parts and the total distribution of predictable and unpredictable requests to the stock during a design horizon 0…T (often typical for budgeting).

In *StockA*, the stock balance is simulated event by event under the guidance of control parameters (order point, order quantity, and the interval for advance order) and suppliers' delivery data. The resulting detailed event log is the base for calculation of decision-guiding results. A variety of figures, tables, probabilities and distributions describing the properties of the stock follow: Order interval, consumption, turnover, etc. Several results on probability and duration for different types of part shortage are applicable in subsequent risk evaluation.

After cost inputs on the cluster level, the loss associated with parts shortage and its consequences are obtained in *PrepB* on the customer level. General economical inputs in *StockB* lead then to the final cost and penalty accounting. Means, deviations, quantiles and distributions follow for different areas of cost and their combinations.

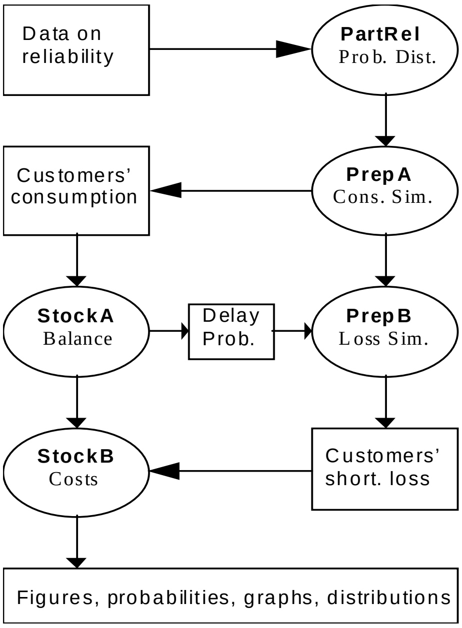

Figure 1. Main structure of the model

2 PART CONSUMPTION AND DEMAND ON STOCK

The very short description of our model given above will now be deepened. Spare part demand, process redundancy, and maintenance policy are main subjects of this section. The following samples from the literature will guide and illuminate the reader in his/her comparisons with our method.

Willemain et al. (2004) and Regattieri et al. (2005) deal with forecasting methods for intermittent demand. Smidt-Destombes et al. (2005) provide a model for redundant systems with identical, reparable components under condition based maintenance policy. Vaughan (2005) brings out a model where demand for the spare parts arises due to regularly scheduled preventive maintenance and random failure of units in service with constant failure rate. Akcali et al. (2001) use the Gamma-distribution for the demand of slow moving parts during lead-time. (Authors employ rarely other distributions than the exponential, normal or Poisson distribution.)

2.1 *Stress levels and reliability of parts (PartRel)*

The part locations in customers' assemblies are the real spare part consumers. The locations can be associated with different stress/*uu*, where *uu* is a natural unit for measuring usage of the part. The designer defines a number of *stress levels* ($\phi = 1, 2, \ldots$) and assesses each location to an appropriate level. The stress levels are common for all customers. Note that it is not necessary to define the concept of stress exactly.

For each stress level ϕ, the designer prepares in PartRel (Fig. 1) a Gamma model for the probability $R_\phi(x)$ that the part is non-failed at x (*uu*). The average μ_ϕ (*uu*) and the deviation σ_ϕ (*uu*) are input data to PrepA. This distribution can be designed as a weighted mixture of several versions arising from different types of initial data: means, deviations, quantiles, censored/non-censored data, etc. (Fig. 2).

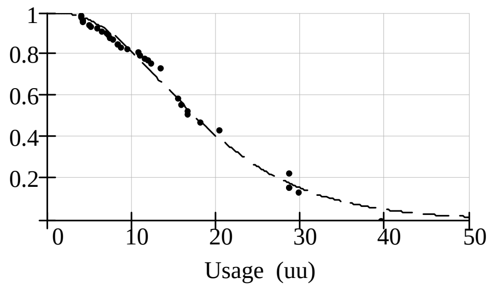

Figure 2. Reliability from censored data

2.2 *Fault logic and maintenance policy (PrepA)*

For the modeling of a customer's spare part consumption, our model provides the concept of a *k/n-cluster*. This is a subset of part locations defined by the following parameters:

n	number of parts in the cluster, $n \geq 1$
k	the cluster is *failed* when k parts have failed, $1 \leq k \leq n$
n_ϕ	number of parts belonging to stress level $\phi = 1,2...$ ($n = n_1+n_2+...$)
K	usage (*uu*) during the design interval $0\ldots T$ (*tu*), *tu* = time unit
H	scheduled exchange interval (*uu*), $0 \leq H \leq \infty$
ΔH	interval (*uu*) to be described below, $0 \leq \Delta H \leq H$
dH	interval (*uu*) to be described below, $0 \leq dH \leq \Delta H$
δ	probability that a part fails immediately after exchange
ε	probability that a scheduled exchange is left unperformed

A customer's part locations can form clusters of different type. Single parts are of course 1/1-clusters. Some additional descriptions can be informative.

If a *k/n*-cluster does not fail (less than k parts failed) during a scheduled interval H, the whole cluster is replaced (*n-exchange*). If a cluster failure instant is not nearer than ΔH from the next scheduled n-exchange, then the failed (k) parts only will be replaced (*k-exchange*); otherwise, the whole cluster is replaced (*kn-exchange*).

After a k-exchange, the planned schedule is still valid, but after a *kn*-exchange, there are two alternatives: If the *kn*-exchange was nearer than dH to the next scheduled n-exchange, this will only be ignored; otherwise the schedule restarts from this *kn*-exchange.

The simulation principle as such corresponds to hot redundancy. However, the usage parameter K and the possibility that the parts of a cluster can work at different stress levels enable also modeling of cold redundancy.

2.3 *Customer's part consumption and demand on stock (PrepA)*

Every installed spare part is assumed to have the same initial condition. Thus, its failing is determined by the stress level of its location (μ_ϕ, σ_ϕ, Sect. 2.1). Nonzero part age (*uu*) at the start of simulation can also be taken into account. The customer's cluster structure is then simulated, leading to the number of *n*-, *k*- and *kn*-exchanges during $0\ldots T$, and the probability for different quantities wanted.

A customer's request to the stock (instant and number of parts) will be called *predictable*, if the stock knows at least Δt (*tu*) in advance. All other requests will be called *unpredictable*. The model assumes that all *n*-exchanges and a fraction $0 \le p \le 1$ of *k*- and *kn*-exchanges lead to predictable requests. For example, the presence of failure diagnostics can be modeled with a nonzero parameter *p*. A customer's demand on stock during the interval $0\ldots T$ consists of the following information:

Data 1. Customer's demand.

Unpredictable requests:
- Average # of requests (k & kn, fraction 1-*p*)
- Corresponding deviation
- Probabilities for required # of parts

Predictable requests:
- Instants of requests caused by n-exchanges
- Corresponding # of required parts
- # of parts of k- or kn-exchanges, fraction *p*

This data can also be entered directly, without employing PartRel and PrepA. For example, the 'waste' (obsolescence, theft, etc.) is a customer of this type.

2.4 *Distribution of requests (StockA)*

The total mean μ and deviation σ of the number N of unpredictable requests during $0\ldots T$ are determined from customers' demand data (Data 1), assuming customers are independent of each other. Since often $\mu \ne \sigma^2$, the Poisson model is too restrictive. Our count data model possesses full dispersion ability: $(\mu - \lfloor\mu\rfloor)(\lfloor\mu\rfloor + 1 - \mu) \le \sigma^2 \le (\mu - a)(b - \mu)$, $a \le N \le b$, where $\lfloor x \rfloor$ denotes the greatest integer $\le x$.

During simulation, a random variate N is generated. A point process spread N request instants on $0\ldots T$. At each instant, the customer and the number of parts wanted are then generated using probabilities derived from demand data. Periodicity and unevenness in the spread can also be modeled (Fig. 3).

Predictable requests caused by *n*-exchange are of course directly placed on $0\ldots T$, and the predictable fraction *p* of *k*- or *kn*-exchange can be placed 'here and there' (Fig. 4). The 'real' instants (simulated in section 2.3) can differ from this, e.g. since the schedule may change ex tempore (Sect. 2.2). However, the reorder principle below indicates the difference can be ignored (Sect. 3.1).

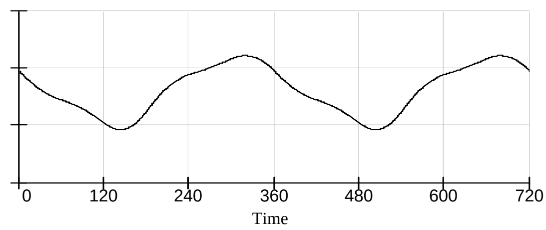

Figure 3. Spread of unpredictable requests (T=720)

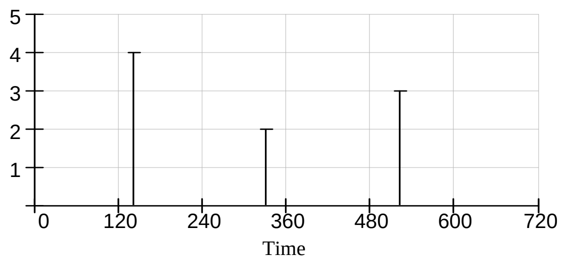

Figure 4. Predictable requests (T=720)

3 MANAGING INVENTORY BALANCE

The objective of inventory control is to achieve, in some sense, an optimal policy. Kennedy et al. (2001) encapsulated some central issues to be resolved: when to place an order, how many units to order, and the choice of objective, e.g. to reduce costs or to increase availability.

Kabir & Farrash (1996) have implemented an inventory model by using the SLAM simulation language for determining a jointly optimal age replacement and spare part provisioning policy. The policy is formulated by combining age replacement policy with an inventory policy of the continuous review type. The optimal values of the decision variables are obtained by minimizing the total cost of replacement and inventory.

Ni et al. (2004) present a model where the information of demand and replenishment is abstracted in terms of asset arrivals, part scraps, and service leadtimes. They present an objective function for a recyclable inventory to achieve on time maintenance with low inventory costs, where tardiness penalties for the maintenance delay are converted to inventory backorder costs.

We will now continue the description of our method. Again, we hope the references mentioned above can serve the reader as comparisons.

3.1 *Part supply and control variables (StockA)*

Two groups of part suppliers can be modeled, one is for predictable and the other for unpredictable requests. A supplier is characterized by a Gamma-modeled lead-time (minimum, mean, deviation), the probability to be chosen, and the prices for delivery (per order and per item). Thus, the choice of supplier is both dynamic and random.

The lead-time is of course a sum of several phases. For example, if the repair shop is a supplier, the repair time is part of lead-time. (Some versions of our model support this decomposition.).

The model treats predictable requests in the following way: When the stock knows a customer or customers will request for m parts totally at a certain time instant t (*tu*), an *advance order* for exactly m parts is set at the time instant $t\text{-}\Delta t$. Note that the same Δt defines predictability (Sect. 2.3)!

Unpredictable requests again are governed as follows: At the beginning of the simulation, the stock balance equals the *reorder point*, r parts, and the *order quantity* ('lot size') is q parts. Thereafter, always when q parts have been unpredictably requested, the stock orders q parts. This happens even if previously ordered goods have not yet arrived. Otherwise, a balance would perhaps not be attainable.

3.2 *Stock balance and service capacity (StockA)*

Stochastic simulation under the control r, q, Δt (Sect. 3.1) produces a 'logbook' describing events and causes in chronological order for a large number of intervals 0…T. The first indicators of design success or non-success are the random samples of the *stock balance* (Fig. 5).

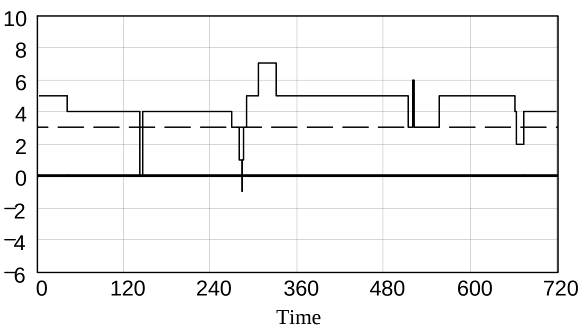

Figure 5. Balance sample (r=3, q=2, Δt=40, T=720)

Good general measures on how the inventory design succeeds are the *service grades*, which describe, from different viewpoints, how often stock-out periods occur, and the *stock-out period*, i.e. the length of single intervals with negative balance (Fig. 5):

- P_1 probability that the customer's request is fully satisfied
- P_2 probability that the required part is available in the stock
- P_3 probability of no lack of parts during the T-interval
- *mp* average stock-out period (*tu*)
- *sp* corresponding deviation (*tu*)

The stock-out period is typically a heavy-tailed random variable. Our model employs a modified Beta-distribution. Several more differentiated results on non-service are also obtained (partial shortages, types of request, etc).

Many common stock output variables are of course computed: part consumption, order interval, deliveries, stock turn, usage volume, the time a part spends in stock, etc. (means, deviations, quantiles and probability distributions).

3.3 *Customer's shortage costs (PrepB)*

Let the triple P_2, *mp*, *sp* stand for service capacity (Sect. 3.2). The stock-out period (*mp*, *sp*) is an upper estimate for the delay experienced by customers. For example, a customer may arrive during a shortage period, or a delivery may satisfy some of the waiting customers but not all. The less such (undesirable) situations occur, the more accurate is the estimate. Further, we assume the same stock shortage probability 1-P_2 is valid for each requested part. Thus, the number of parts obtained by a customer comes from a truncated geometric distribution with the parameters P_2 and the requested quantity.

According to preceding sections, all customers' are factors behind the service capacity. Consider a customer, who took part in the stock simulation above and whose consumption was assessed in PrepA (Sect. 2.2-3). The shortage costs of this customer can now be computed in PrepB. First, the shortage costs (prices) are given for each cluster in Data 2 format. Then, the raw simulation data obtained in PrepA for failures and service actions leads to the shortage cost of each cluster and each T-interval. Summing with respect to clusters yields the customer's cost distribution/T, and summing with respect to T-intervals yields e.g. cost sharing among clusters.

The final crystallization step produces the shortage costs for all clusters of this customer (again in Data 2 format). If a customer's consumption was not modeled using PrepA, this cost data must be assessed directly.

Data 2. Customer's loss caused by shortage.

Loss caused by 1 and 2 lacking parts:

- Cost per shortage (euro)
- Time-dependent cost in 0…t_0 (euro/tu)
- Time-dependent cost in t_0… (euro/tu)
- Time instant of change, t_0 (tu)

The loss for more than two lacking parts is determined by the difference between one and two parts

3.4 *Cost accounting (StockB)*

For the final cost calculation, the following inputs already exist: Supply prices (Sect. 3.1), the logbook (Sect. 3.2), and customers' shortage prices (Sect. 3.3). General economic figures are now given: item purchase price, item holding cost, annual percentage rate, value decline, and value of lost item. There can also be a simple *penalty* contract between a customer and the stock keeper: the stock pays a fraction ($0 \leq ps \leq 1$) of the customer's shortage costs.

Results follow for cost areas like facilities, upkeep, delivery, interests, waste, value decline, and customers' shortage (fraction *ps* and 1-*ps*). Since the model keeps these cost areas separated, and arbitrary selected combinations can be summed, the costs can be shared between responsible parties.

4 SPARE VALVE INVENTORY (A CASE STUDY)

We applied our model in a case study concerning the control valve of the air turbine start of a Fighter Aircraft F-18 (Hagmark et al. 2004). The failure data delivered by the Finnish Air Forces comprised about 7 years, 142 valves with age range 0…1511 operations, totally 94097 operations and 150 failures.

When a valve does not operate correctly at a gas turbine start, a repaired spare valve is taken from the stock, and is installed. When 6 valves have failed (one by one), these valves are sent to service for repair (or renewing). In other words: There is one customer, who requests for one part unpredictably (Sect. 2.3). The order quantity is $q = 6$, and an order is placed when the stock balance have reached the order point $r = M$-6, where M is the so far unknown number of spare valves (Sect. 3.1).

The maintenance time for 6 valves is 18 days and three weekends waste 6 days extra, so the total time that 6 valves spent in service is 24 days. The trip to service and back takes 16 days. In other words: We have one part supplier with 40 days lead-time on the average (Sect. 3.1). The deviation was directly assessed to be 15 days.

Data analysis yields the unpredictable consumption of spare valves: On the average 33.1 valves per year, with variance 33.6 and slight season dependence (Sect. 2.3-4). Now, how large must M be to avoid a shortage possibility of more than (say) 1%? Stock simulation yields $P_2 = 0.9919$ if $M = 17$, and $P_2 = 0.9875$ if $M = 16$ (Sect. 3.2). Thus, an acceptable service capacity is guaranteed by keeping 17 spare valves.

5 CONCLUSIONS

Experimentation with the inputs of our model can solve a rich variety of problems concerning inventory design and related topics. The multitude of variables and concepts enables a vast number of practical interpretations. There is much to do in mapping out possibilities. For example, suitable interpretations of the pair 'supplier/customer' cover lateral shipment of parts (Wong et al. 2006), the steps in a multi-echelon system (Rustenburg 2001), the repair shop as a supplier (Ni et al. 2004), etc.

The corresponding and continuously developing software enlarges the applicability of the model. Automatic optimization of any selected cost combination can be performed, and the set of partaking variables $\{r, q, \Delta t\}$ can gradually be extended (Sect. 3.1 & 3.4). On the other hand, optimization of e.g. complex multi-echelon systems would require very effective algorithms (cp. genetic algorithms).

Constraint checking during optimization has also been implemented to some extent. For example, the (stock keeper's) costs can be balanced against the (customer's) risk caused by spare part shortage. Other constraints from budget, capacity, etc. will provide new challenges (Rustenburg 2000).

At present, our model does not treat multi-item systems (Rustenburg, 2000).

REFERENCES

Akcali, E., Davis, M., Hamlin, R. D., McCullough, T., Teyner, T., Uzsoy, R. A. 2001. Decision support system for spare parts management in a wafer fabrication facility. *IEEE transactions on semiconductor manufacturing, 14(1), February. 2001*.

Hagmark, P-E. & Yli-Kovero, T. & Pernu, H. 2004. Valve Inventory Sizing based on Data Analysis of Failure Tendency. *26th ESReDa Seminar, 11-12 May 2004.* European Safety, Reliability & Data Association.

Ni, M. & Luh, P. B. & Xiong, B. & Chang, S-C. 2004. An Inventory Control Policy for Maintenance Networks. *Proceedings of 2004 IEEE/RSI International Conference on Intelligent Robots and Systems, September 28 - October 2*: 1238-1244. Sendai, Japan.

Kabir, A.B.M.Z. & Farrash, S. H. A. 1996. Simulation of an integrated age replacement and spare provisioning policy using SLAM. *Reliability Engineering and System Safety* 52: 129-138.

Kennedy, W. J. & Patterson, J. W. & Fredendall, L. D. 2002. An overview of recent literature on spare parts inventories. *Int. J. Production Economics* 76: 201-215.

Regattieri, A., Gamberi, M., Gamberini, R., Manzini, R. 2005. Managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management* 11: 426–431.

Rustenburg, W., 2000. A System Approach to Budget-Constrained Spare Parts Management. PhD thesis. Febodruk B.V., Enschede, Netherlands.

Smidt-Destombes, K. S. & de Heijden, M. C. & van der Harten, A. 2005. On the interaction between maintenance, spare part inventories and repair capacity for a k-out-of-n system with wear-out. *European J. Operational Research.* IN PRESS.

Vaughan, T. S. 2005. Failure replacement and preventive maintenance spare parts ordering policy. *European J. Operational Research* 161: 183–190.

Willemain, T. R. & Smart, C. N. & Schwarz, H. F. 2004. A new approach to forecasting intermittent demand for service parts inventories. *Int J. Forecasting* 20: 375– 387.

Wong, H. & van Houtum, G.J. & Cattrysse, D. & van Oudheusden, D. 2006. Multi-item spare parts systems with lateral transshipments and waiting time constraints. *European J. Operational Research* 171: 1071–1093.