International Journal of Advances in Computer Science & Its Applications– IJCSIA
Volume 5: Issue 1 [ISSN : 2250-3765]
Publication Date : 30 April, 2015

Analyzing Supply Chain Nodes in Heterogeneous Environments Based on Transaction Data with Respect to Independent Item Behavior

Sebastian Lauck, Simon Boxnick, and Lukas Kopecki

***Abstract*— Sizing and redesign of new and existing supply chain nodes, like warehouses or logistic hubs are highly data-dependent tasks, incorporating lots of transaction information which are mostly available through ERP systems. Therefore, item-based forecasting and lot sizing models, approximations or both are used widely when planning new supply chain nodes. This paper introduces a new approach of data-aggregation based on probability functions of each item (e.g. stock keeping units (SKU)) incorporating compensating behavior and time-dependent aspects.**

***Keywords*— data mining, big data aggregation, supply chain analysis, material flow analysis, warehouse planning**

I. Introduction

Analyzing material flows and stock levels is a main task when trying to optimize single supply chain items, as well as considering flows along the supply chain, but much less emphasis has been placed on the material flow and requirements of single supply chain nodes [1]. Problems for this task are the processing of huge amounts of data collected and the calculation of reliable overall key indicators on different aggregation levels. Especially when planning new supply chain nodes, like warehouses it can be a tough job to retrieve good expectation-values for the requirements — especially capacity and turnover requirements [2] of these elements.

When having large sets of transaction data [3,4] (assumed considering only inbound and outbound transactions for one node), there are two main trends in retrieving overall information: First, the analysis for single SKUs or load carriers with time-series models or probability functions. This results in precise results for the single item (e.g. a high stock level) but neglects interdependencies between different items, like adding high stock level predictions of winter-items to the results of summer parts, because the association between data and time is not kept. Second the usage of aggregated high-level information like realized turnover or stock level rates from the past — which ignore specific item details. Assuming a new warehouse C shall be built, with fractions of the stock from some warehouse A and fractions of the stock from warehouse B — the main question arises what size the new warehouse should have and what turnover rates will arise from the given SKUs to justify customer needs.

Sebastian Lauck, Simon Boxnick and Lukas Kopecki
Heinz Nixdorf Institute, University of Paderborn,
Germany

A lot of work about the stock level behavior aggregating different warehouses has been done under the term risk pooling [5], but the effects on turnover rates are not well studied. Therefore, a structural approach on incorporating all information in sets of probability functions describing the article behavior and reducing the amount of necessary data is introduced here. Furthermore, the data is reduced much more by defining a hierarchical structure where each node represents the combined information set by storing the overall probabilities over all children.

One benefit of this structure is the quick alteration of node to node associations and, therefore, on-the-fly retrieval of predicted node requirements without the need of recalculation or repeated simulation of the whole dataset. The presented approach furthermore analyzes the dataset for peak usages on daily, weekly and monthly base, so high turnover phases like deliveries in the morning or seasonal peaks are recognized but not accumulated in an imprecise fashion.

II. Literature overview

Sodhi gives an overview on the integration of strategic planning for supply chains [6], but presents primary a conceptual sketch on what to do and not a mathematical model which shows how to gain the values.

Much research points to optimal sizing [7,8] and siting [9,10] or both [11] warehouses based on cost per SKU or cost per most of these papers use the amount of necessary stock capacity and turnover rates, but do not formally describe how to retrieve this information. Hindi and Toczylowski present a multistage model of production items and calculate item lot-sizes based on absolute item demands [12]. Arnold and Furman are determining stock levels in stochastic environments based on probability distributions [13], but their procedure supports neither turnover rates nor the aggregation of information over different levels. While the sizing of supply chains nodes — mainly warehouses — is studied mostly under cost-minimization aspects, neglecting heterogeneous SKU behavior, nearly no effort has been done to develop general models to analyze the turnover needs for material flow systems inside the supply chain nodes.

SEEK DIGITAL LIBRARY

III. Solution concept

The process of resolving relevant key-values is described by the term “dimensioning" in the following description and donates for the given formal models expected inbound and outbound flows for the analyzed node, as well as stock levels. These values can be used for further analysis like planning of new nodes or structure changes in existing ones like altering load carriers. For given service-level the amount of necessary storage in the form of load carrier-units, as well as the turnover rates shall be calculated independent of the count of underlying transactions.

First, we will describe the calculation of derived stock level and turnover rates for single SKUs. Based on this model a structural extension for the calculation of combined levels like comparable SKUs, load carriers and whole supply chain nodes is introduced.

A. *Stock level calculation*

Arnold and Furmans core idea of using probability distributions and the convolution of them to gain evident information about overall stock levels are the foundation of the following model, but their model is extended in several ways making it usable in a broader context.

We constitute that a set of historical data is given for each SKU representing information about inbound and outbound flow, further called transaction data (Considering only demands, common lot-sizing models can be used to calculate optimal inbound strategies and vice versa.). One transaction T, is defined as a tuple $T = (T_A, T_t, T_m)$ and contains following information:

- The SKU A, on which the transaction is performed (T_A),
- The period t (a specific time-range) when the transaction occurred (T_t) and
- The amount of the SKU which has been moved in- or outbound (T_m), while Tm < 0 represents an outbound transaction.

For one SKU, the associated transaction data is labelled with $A^{\uparrow}$ (outbound transactions) and $A^{\downarrow}$(inbound transactions). The total amount of the outbound flow that A has been transferred in one given period t is called $A^{\uparrow}(t)$ respectively $A^{\downarrow}$(t) for the inbound flow. Because multiple transactions for one SKU can occur in one period $A^{\downarrow}(t)$ is calculated by using equation (1). Same holds for $A^{\uparrow}(t)$ which is defined by equation (2). For the example period-length, initially one hour is considered.

$$A^{\downarrow}(t) = \sum T_m, \forall (T, T_A = A \wedge T_t = t \wedge T_m > 0) \quad (1)$$

$$A^{\uparrow}(t) = \sum T_m, \forall (T, T_A = A \wedge T_t = t \wedge T_m < 0) \quad (2)$$

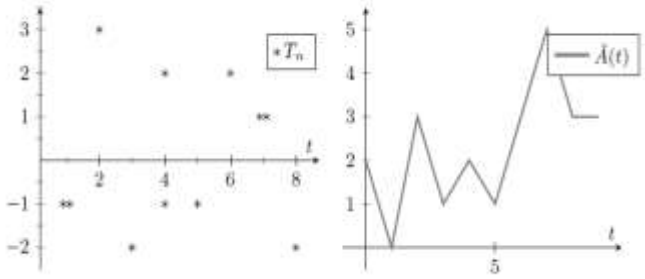

Figure 1. Transaction to stock level mapping for one article

The resulting stock level for SKU A at the end of period t ($\check{A}(t)$) can be determined recursively using equation (3). The initial stock level $\check{A}(t = 0)$ must be given as cancellation condition, representing the initial stock level. Fig. 1 shows the connection between transaction data and resulting stock level.

$$\check{A}(t) = \check{A}(t-1) + A^{\downarrow}(t) - A^{\uparrow}(t) \quad (3)$$

Some authors argue that it is sufficient to consider the sum of average stock levels to determine the overall stock level. These models seem to be based on models defined in less computerized concepts. Distinct SKUs can exhibit large variations in demand and, therefore, more detailed analysis is suggested, by referring to the complete distribution representing the probability for each stock level for every SKU. The usage of distributions leads to two advances: First, a much better data-basis than an average value but also this gives the possibility to reduce the transaction set to one distribution, which can be processed faster.

To determine the stock level probabilities of one SKU the time-date series $\check{A}(t)$ is mapped to the empirical probability mass function $p(\check{A})$calculating the total occurrences of each stock level in the interval [0, max($\check{A}$)] and setting them in relation to all other values.

Claiming a service grade for sufficient space $\overline{SG}$ for SKU A the necessary stock-capacity $\check{E}_A$ can be determined using the empirical cumulative distribution function of A’s stock levels $p^*(\check{A})$ (the value where the relative cumulative frequency exceeds $\overline{SG}$). Calling the $Q(\overline{SG})$ quantile of the cumulated distribution $p^*(\check{A})$ short $p^*_{\overline{SG}}(\check{A})$ like defined in equation (4).

$$\check{E}_A = p^*_{\overline{SG}}(\check{A}) \quad (4)$$

For complete supply chain nodes, not the stock level of one article is relevant, but the aggregated stock over all articles. Fundamental concepts of risk-pooling can be transferred here: It can be formally shown that the aggregated standard deviation over many SKUS is always smaller than the sum the single standard deviation for each SKU [5]. Therefore, the concept of analyzing of the aggregated distribution over all SKUs outreaches single SKU stock levels:

SEEK DIGITAL LIBRARY

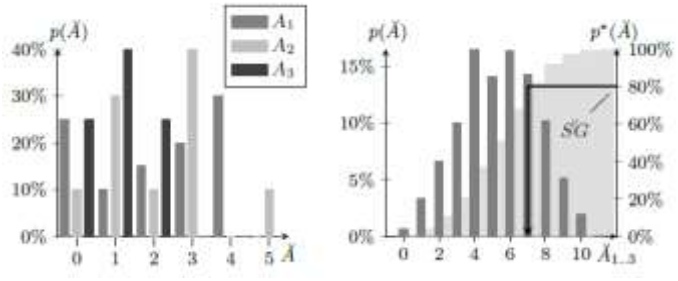

Figure 2. Convolution of three SKU stock level distributions and retrieval of stock- size for reliability $\widetilde{SG}$ = 80%

$$\sigma = \sum_{i=1}^{n} \sigma_i \geq \sigma_\alpha = \sqrt{\sum_{i=1}^{n} (\sigma_i)^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i \sigma_j p_{ij}} \quad (5)$$

Taking n (stochastically independent) SKUs ($A_1 .. A_n$) of the same size, the aggregated stock level distribution can be obtained by calculation of all possible linear combinations for the SKU based stock level distribution p($\check{A}_j$) written as j=1 $\sum_{j=1}^{n} p(\check{A}_j)$ [13] (compare Fig. 2). The convolution over the stock level distributions for n SKUs is written short as $p(\check{A}_{1..n})$ (equation (6)). The necessary reserved space to stock n SKUs is calculated with equation (7):

$$p(\check{A}_{1..n}) = \sum_{j=1}^{n} p(\check{A}_j) = p(\check{A}_1) * p(\check{A}_2) * \ldots * p(\check{A}_n) \quad (6)$$

$$\check{E}_{A_{1..n}} = p^*_{\widetilde{SG}}(\check{A}_{1..n}) \quad (7)$$

B. *Turnover calculation*

Besides the analysis for stock levels, the calculation of turnover rates supply chain nodes is important. For example in highly BTO (build to order) driven supply chain nodes stock levels may be constantly low, but huge counts of transactions occur being neglected by solely looking at stock levels. Turnover rates are often measured as items processed per hour — therefore we use this duration as well. One transaction is defined as the movement of tm units of SKU A in one direction. Therefore, one transaction can force more than one handling operations in the supply chain node [14]: Given a load carrier with capacity four, two handling operations have to be executed to operate one transaction with $T_m = 5$. Opposite to this it cannot be assumed that two or more transactions can be aggregated to one operation because the exact time of occurrence could differ (in the same period). To consider this turnover property, the inbound handled units per period $A^{\downarrow}(t)$are stored separately to the amount of handlings per period $A^{|\downarrow|}(t)$. When associating the SKU A to a load carrier L ($\chi_{L,A} = 1$), with a capacity relationship of A^L (A^L units of A fit in L) the amount handled in the inbound flow for A in one period can be calculated with equation (8).

$$E_A^{\downarrow}(t) = max\left(\frac{A^{\downarrow}(t)}{A^L}, A^{|\downarrow|}(t)\right) \quad (8)$$

To retrieve the amount of operations in one period, the maximum of transactions per hour and inbound handlings per hour is chosen (equation (9), where $p_L(A^{\downarrow})$, donates the distribution of amounts of load carriers (L) to transfer SKU A inbound.

$$E_A^{\downarrow} = max\left(p^*_{SG^{\updownarrow}}(a^{\downarrow}), p^*_{SG^{\updownarrow}}(a^{|\downarrow|})\right), with \quad (9)$$
$$a^{\downarrow} = p_{L_m}(A^{\downarrow}),$$
$$a^{|\downarrow|} = p(A^{|\downarrow|})$$

When designing new material flow systems for supply chain nodes, the amount of transferred articles is not the first relevant measure, but the amount of transferred load carriers. The distribution of inbound transferred load carriers $p(L^{\downarrow})$ can be calculated by convolving the load-carrier corrected distributions of every associated SKU ($p_L(A^{\downarrow})$), respecting the connection between the amount transferred and the amount of transactions:

$$E_{L_m}^{\downarrow} = max\left(p^*_{SG^{\updownarrow}}(l_m^{\downarrow}), p^*_{SG^{\updownarrow}}(l_m^{|\downarrow|})\right), with \quad (10)$$
$$l_m^{\downarrow} = \sum_{n=1}^{N}\left(\chi_{m,n} * p_{L_m}(A_n^{\downarrow})\right) * \sum_{o=1}^{O}\left(\chi_{m,o} * p_{L_m}(L_o^{\downarrow})\right)$$
$$l_m^{|\downarrow|} = \sum_{n=1}^{N}\left(\chi_{m,n} * p(A_n^{|\downarrow|})\right) * \sum_{o=1}^{O}\left(\chi_{m,o} * p(L_o^{|\downarrow|})\right)$$

This kind of data-manipulation results in much fewer items to process when calculating logistic measures than analyzing the whole dataset of transactions, but it leads to complete loss of data-to-time relationships. This is a huge problem especially when looking at turnover data, because turnover frequencies tend to vary strongly on daily, weekly or monthly basis.

To gather the real peaks in turnover behavior, periods are extended with an additional identifier in the context to some given superior period description called period-system. For example, part-periods for the third hour of a day or the second day of a week. The count of part-periods is signed with *I*. The distribution view on all SKUs and load carriers is extended now with distinct distributions for each part-period. This process extends the amount of data to store and compute (*I* times the amount of measures, 96 for transaction-data on an hour metric), but is still much quicker to analyze than respecting all distinct transaction items.

The turnover distributions are calculated by indexing all periods (t_j). Then the data is grouped to sets with same part-period type (t_j mod I = t_i), and distributions are built as described above for each set.

The relative amount of items for each turnover for every part-period is calculated with equation (11), whereby $H(A^{\downarrow}, t_i)$ represents the absolute amount of ingoing turnovers in $A^{\downarrow}$ in all equal-typed part-periods (e. g. “3 times 5 items of SKU A have been dispatched between 8 and 9 o’clock”). n_i donates the total amount of periods with the same index-type (the amount of days analyzed and, therefore, the amount of the *i*-th hour).

SEEK DIGITAL LIBRARY

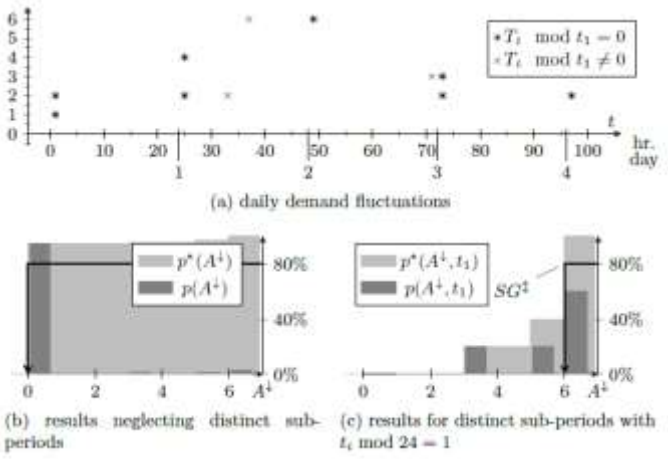

Figure 3. Influence of stochastic demands (a) on the distribution function of ingoing transactions without (b) and with (c) dedicated sub-periods t_i

$$p\left(A^{\downarrow}, t_i\right) = \frac{H\left(A^{\downarrow}, t_i\right)}{n_i} \tag{11}$$

Fig. 3 shows the described model with an example: Analyzing one SKU on an undifferentiated hour-scale for a given service-grade an underestimate of turnovers can occur (zero items / hr.) Distinguishing the hours for one day, much better values (6 items / hr. in the worst case) can be achieved. The necessary capacity to cover a turnover service grade $SG_{\updownarrow}$ can be identified with the cumulated distribution functions $p^*\left(A^{\downarrow}, t_i\right)$ and $p^*\left(A^{|\downarrow|}, t_i\right)$.The overall representative turnover peak is defined as the maximum of the sum from inbound and out- bound turnover-values for a given service level in the worst part-period. Calling the $SG^{\updownarrow}$-fulfilling value $p^*_{SG_{\updownarrow}}$, the part-period with the highest expected turnover for one SKU $t^{\updownarrow}_{A,i_{max}}$ max is computed with equation (12). To retain readability of the equations short variables $a^{\downarrow}_{t_i}$, $a^{\uparrow}_{t_i}$ are used for the load-carrier corrected distributions (see equation (12c) and (12f)). Same holds for (12d) and (12g)), but here no load-carrier correction occurs, because the total amount of handlings counts. Equations (12b) and (12e)) revise equation (8) for the multi-period model.

$$t^{\updownarrow}_{A,i_{max}} = \underset{0 \le i \le I}{\operatorname{argmax}}\left(E^{\downarrow}_{A,t_i} + E^{\uparrow}_{A,t_i}\right), with \tag{12a}$$

$$E^{\downarrow}_{A,t_i} = \max\left(p^*_{SG^{\updownarrow}}\left(a^{\downarrow}_{t_i}\right), p^*_{SG^{\updownarrow}}\left(a^{|\downarrow|}_{t_i}\right)\right) \tag{12b}$$

$$a^{\downarrow}_{t_i} = p_{L_m}\left(A^{\downarrow}, t_i\right) \tag{12c}$$

$$a^{|\downarrow|}_{t_i} = p\left(A^{|\downarrow|}, t_i\right) \tag{12d}$$

$$E^{\uparrow}_{A,t_i} = \max\left(p^*_{SG^{\updownarrow}}\left(a^{\uparrow}_{t_i}\right), p^*_{SG^{\updownarrow}}\left(a^{|\uparrow|}_{t_i}\right)\right) \tag{12e}$$

$$a^{\uparrow}_{t_i} = p_{L_m}\left(A^{\uparrow}, t_i\right) \tag{12f}$$

$$a^{|\uparrow|}_{t_i} = p\left(A^{|\uparrow|}, t_i\right) \tag{12g}$$

The resulting values for inbound ($E^{\downarrow}_{SG\updownarrow}$) and outbound ($E^{\uparrow}_{SG\updownarrow}$) turnovers can be obtained easily using the obtained worst-case part-periods:

$$E^{\downarrow}_{A} = E^{\downarrow}_{A,t^{\updownarrow}_{imax}} \tag{13a}$$

$$E^{\uparrow}_{A} = E^{\uparrow}_{A,t^{\updownarrow}_{imax}} \tag{13b}$$

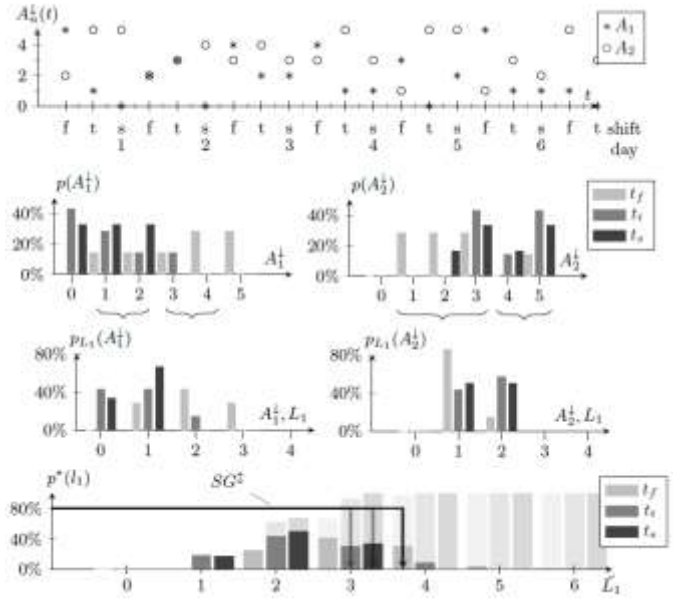

Figure 4. Determination of turnover peak points for an load carrier associated to SKUs for given service-level $SG_{\updownarrow}$ = 80%

Looking at the peak levels on load carrier-level, the amounts can be retrieved again by folding all relevant distribution of associated articles. For each part-period t_i the distributions $l^{\downarrow}_{m,t_i}, l^{\uparrow}_{m,t_i}, l^{|\downarrow|}_{m,t_i}$ and $l^{|\uparrow|}_{m,t_i}$ have to be computed. The worst-case part period in load carrier turnover $t^{\updownarrow}_{L_m,i_{max}}$ can be obtained with equation (13). The service grade driven turnover key indicators $E^{\downarrow}_{L_m}$ and $E^{\uparrow}_{L_m}$ are the results.

$$t^{\updownarrow}_{L_m,i_{max}} = \underset{0 \le i \le I}{\operatorname{argmax}}\left(E^{\downarrow}_{L_m,t_i} + E^{\uparrow}_{L_m,t_i}\right), with \tag{14}$$

$$E^{\downarrow}_{L_m} = E^{\downarrow}_{L_m,t^{\updownarrow}_{i_{max}}}$$

$$E^{\uparrow}_{L_m} = E^{\uparrow}_{L_m,t^{\updownarrow}_{i_{max}}}$$

The process of retrieving ingoing turnovers with distributions (for a different period-system) is given in Fig. 4: First for each SKU the distributions for an early shift, day shift and late shift are obtained. Based on them the load-carrier capacity corrected distributions are used in the convolution stage. Last, the convolved distributions for each shift are cumulated to get the $SG^{\updownarrow} = 80\%$- secure turnover values. The turnover to cover is, therefore, four in the given example, occurring in the early-shift.

C. *Hierarchical information aggregation*

The main benefit of the described distribution model is the opportunity to rep- resent the complete turnover- and stock level values in a hierarchical fashion like shown in Fig. 5. Nodes in the graph represent the set of all relevant

SEEK DIGITAL LIBRARY

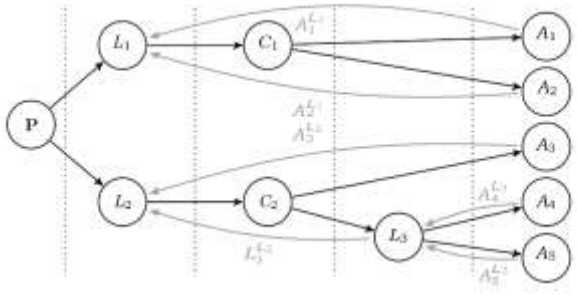

Figure 5. Overall logistics structure combining all load carriers L, SKUs A, nodes for similar behavior clusters C and root node for the project P

distributions for each SKU and load carrier. The distributions for each parent can be obtained by using the shown capacity levels (A^L and distributions of the children. This leads to an easy alterable structure. Additional relations like load carrier in load carrier (e.g. boxes on pallets), or clusters of similar items (like high turnover, or high-stock SKUs) can be integrated. Alteration of load-carrier structures or capacities can be investigated using a much smaller set of data because only the relevant distributions of the children have to be stored and computed.

IV. Conclusion

In this paper, we have described a model to aggregate transaction data in probability distributions without losing relevant time relationship. Using these distributions leads to a much quicker retrieval of service grade driven planning results than raw data sets. The model splits the data source by defined periodical structures (period systems) into representing probability sets to keep a distinction between representative phases like hours of a day or month of a year. Periods are defined solely on a formal level to keep flexibility to custom problems. Furthermore, the hierarchical structure presented leads to additional performance gains because each node represents the aggregated probability function over all children.

Alternative hierarchies like altering the used load carriers, or different article clustering can be calculated with little effort by convolving the probability functions of all direct children.

References

[1] Higginson, J., Bookbinder, J.: Distribution centres in supply chain operations Langevin, Riopel (Ed.) 2005 Logistics Systems, 67–91 (2005)

[2] McGinnis, L., Mulaik, S.: Your Data and How to Analyze It Proceedings of the Industrial Engineering Solutions 2000 Conference (2000)

[3] DHL: Big Data in Logistics. A DHL perspective on how to move beyond the hype DHL Customer Solutions & Innovation (2013)

[4]. Data, data everywhere The economist, Feb. 27th, 2010 (http://www.economist.com/node/15557443 retrieved: 07/14/14) (2010)

[5] Oeser, G.: A Framework for Risk Pooling in Business Logistics Supply Management Research Aktuelle Forschungsergebnisse 2012, 153–299 (2012)

[6] Sodhi, M. S.: How to do strategic supply-chain planning MIT Sloan Management Review, 69–76 (2003)

[7] Cormier, G., Gunn, E. A. Simple models and insights for warehouse sizing. Journal of the Operational Research Society, 47, 690-696 (1996)

[8] White, J. A., Francis, R. L.: Normative models for some warehouse sizing problems. AIIE Transactions, 3, 185-190 (1971).

[9] Brimberg, J., Mehrez, A.; Wesolowsky, G. O: Allocation of queueing facilities using a minimax criterion. Location Science, 5, 89-101 (1997)

[10] Huang, S., Batta, R., Nagi R.: Simultaneous siting and sizing of distribution centers on a plane Annals of Operations Research March 2009, Volume 167, Issue 1, 157–170 (2009)

[11] Francis R. L., White J. A.: Facility layout and location Prentice-Hall (1974)

[12] Hindi K. S., Toczylowski E. Aggregation and Disaggregation of End Items in a Class of Multistage Production Systems The International Journal of Advanced Manufacturing Technology February 1988, Volume 3, Issue 1, 45–54 (1988)

[13] Arnold, D., Furmans, K. Lagern und Komissionieren Materialfluss in Logistiksys- temen, Springer-Verlag 174–195 (2009)

[14] Schaab, W. Technisch-wirtschaftliche Studie ueber die optimalen Abmessungen automatischer Hochregallager, VDI-Verlag Duesseldorf (1968)

SEEK DIGITAL LIBRARY